INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Rationalization in Signaling Games: Theory and Applications

Pierpaolo Battigalli

Working Paper n. 275

December 2004

IGIER – Università Bocconi, Via Salasco 5, 20136 Milano –Italy
http://www.igier.uni-bocconi.it

Rationalization in Signaling Games: Theory and Applications

Pierpaolo Battigalli
Bocconi University and IGIER
pierpaolo.battigalli@unibocconi.it

Revised: September 6, 2004

Abstract

Focusing on signaling games, I illustrate the relevance of the rationalizability approach for the analysis multistage games with incomplete information. I define a class of iterative solution procedures, featuring a notion of "forward induction": the Receiver tries to explain the Sender's message in a way which is consistent with the Sender's strategic sophistication and certain given restrictions on beliefs. The approach is applied to some numerical examples and economic models. In a standard model with verifiable messages a full disclosure result is obtained. In a model of job market signaling the best separating equilibrium emerges as the unique rationalizable outcome only when the high and low types are sufficiently different. Otherwise, rationalizability only puts bounds on the education choices of different types.

Keywords: incomplete information, signaling, rationalization.
Subject Classification: C72, D82.

1 Introduction

Rationalizability is a solution concept that captures the implications of rationality and common belief in rationality. It has been argued that rationalizability is relevant and important in the analysis of incomplete information games (Battigalli and Siniscalchi 2003a, Battigalli 2003, Dekel *et al.* 2003, Ely and Peski 2004). In the context of dynamic games, a strong version of rationalizability also involves the forward-induction assumption that players try to rationalize the past actions of their opponents. In this paper I apply this strong rationalizability approach to the analysis of signaling games. For the sake of completeness in this Introduction I first provide a general discussion of incomplete information games and then I consider signaling games.

In a typical game of incomplete information the relationship between actions and payoffs is not commonly known. Different players have different pieces of information about this relationship. I call such pieces of information "payoff-relevant types", or – more simply, *payoff-types*,[1] and I call a complete specification of the payoff-relevant parameters of the game *state of nature.* A basic description of an incomplete information game would simply specify the rules of interaction (feasible sequences of actions, information about previous actions, etc.), a set of conceivable states of nature, and a set of conceivable payoff-types for each player. However, according to Harsanyi's (1967-68) approach, this description of the strategic situation is insufficient. In order to apply game-theoretic (equilibrium) analysis and derive implications about behavior this basic structure has to be augmented with a *type space*, that is, a mathematical structure which provides an implicit description of the possible configurations of interactive believes: each player's beliefs about his opponents' payoff-types (first-order beliefs) and each player's beliefs about his opponents' beliefs (higher-order beliefs). A *Harsanyi-type* encodes both a player's payoff-types and his first-order and higher-order beliefs concerning payoff-types. A type-space-augmented incomplete information game – also called *Bayesian game* – is structurally similar to a standard game where the players have asymmetric information about an initial chance move affecting their payoffs. Hence an appropriate analog of the Nash equilibrium concept, the Bayesian-Nash equilibrium, can be used to analyze strategic interaction under incomplete information. A *Bayesian-Nash equilibrium* specifies a choice[2] for every Harsanyi-type of every player so that the choice of each Harsanyi-type is a best response to its beliefs given a correct conjecture about the choice that each possible type of the opponents would make.

There is an important conceptual difference between games with incomplete information and games with asymmetric information about an initial chance move. In the latter there is an *ex ante* stage in which the players are equally uninformed, after which the players *learn* their private information. Games with asymmetric information about a chance move (such as Poker) have been studied since the very infancy of game theory, well before Harsanyi's seminal contribution on incomplete information games. On the other hand, in games with incomplete information there is *no ex ante* stage. It is quite simply the case that some players happen to know some relevant facts unknown to other players.[3] For example, an economic agent typically knows more than other agents about his own preferences and innate abilities. In what follows I argue that this difference has been too often overlooked in applications, and I propose an alternative method of analysis of incomplete information games that does not rely on type spaces *à la* Harsanyi and Bayesian-Nash equilibrium.

Harsanyi's analysis of games with incomplete information is in principle quite flexible. Indeed, it has been argued that without specific assumptions

[1] This terminology has become rather standard in the recent literature on incomplete information games and type spaces: see, e.g., Battigalli (2003), Battigalli and Siniscalchi (2003a) and Bergemann and Morris (2004). Harsanyi (1967-68) used the term "attribute vector".

[2] In a dynamic game, the choice concerns a plan of action.

[3] One may consider the case of incomplete, and yet completely symmetric information, but it is less interesting. Also, one may consider games featuring both asymmetric information about an initial chance move, and incomplete information. The methodology I propose can be easily extended to cover these case.

about players' interactive beliefs, the Bayesian-Nash equilibrium concept has very weak behavioral implications. Generalizing an observation due to Brandenburger and Dekel (1987), Battigalli and Siniscalchi (2003a) show that any behavior consistent with common certainty of rationality is also consistent with some Bayesian-Nash equilibrium as long as we consider a sufficiently rich type space.[4] In other words, the equilibrium assumption that players have correct conjectures about the choice that each Harsanyi-type would make does not have any behavioral implication beyond common certainty of rationality, if it is not coupled with specific assumptions about interactive beliefs. As in games of complete information, the behavioral consequences of common certainty of rationality can be weak or strong, depending on the details of the model at hand. But it is worth noting that in order to derive such implications it is not necessary to refer to type spaces *à la* Harsanyi. One only has to perform a solution procedure similar to iterated dominance: at each round of the procedure one eliminates, for each *payoff*-type of each player, the choices that are not best responses to any belief about the combinations of payoff-types and choices for the other players that survived the previous rounds. It is also possible to adapt the procedure to take into account that some features of the players' first-order beliefs may be common certainty (for example, there may be common certainty of the fact that every opponent of player i assigns at least a 50% probability to a particular payoff-type θ_i^*).

To sum up, in order to derive behavioral implications going beyond the consequences of common certainty of rationality, one has to make specific assumptions about the players' interactive beliefs concerning payoff-types. But, in my opinion, Harsanyi's emphasis on the structural similarity between Bayesian games and standard games with asymmetric information, has led many applied economists to acritically use assumptions about interactive beliefs that are often implausible and not well-understood. For example, in economic models with incomplete information on one side, such as signaling games, it is almost always assumed that the beliefs of the uninformed players about the payoff-type of the informed player are commonly known. This amounts to assuming a small type space where there is a one-to-one correspondence among payoff-types and Harsanyi-types.[5]

This widespread modeling strategy raises several problems. It is clear that small type spaces are used for tractability reasons: calculating the Bayesian-Nash equilibria of a game-theoretic model with a large and complex type space may be very difficult. But, except for tractability, economists often do not have other compelling reasons for using small type spaces. Furthermore, it is not clear, *a priori*, how these assumptions about interactive beliefs affect the set of

[4] Related results can be found in in Dekel *et al.* (2003) and Ely and Peski (2004).

[5] Bergemann and Morris (2004) call "naive" a type space where beliefs are derived from a common prior on the set of *states of nature*. This property of interactive beliefs is much stronger than the existence of a common prior on the set of *states of the world*. In two-person games with incomplete information on one side, we have a naive type space if and only if the (first-order) belief of the uninformed player is common knowledge.

On the conceptual interpretation of the common prior assumption see, for example, Morris (1995), Gul (1998), Aumann (1998), Bonanno and Nehring (1999), and Feinberg (2000).

equilibrium outcomes in particular cases.

Another shortcoming of the standard approach *à la* Harsanyi is that it does not provide an adequate framework to formulate and evaluate assumptions about how the players would revise their beliefs if they observed unexpected moves by their opponents. A plethora of refinements of the Bayesian-Nash equilibrium concept have been proposed, much to the confusion of applied economists. Many of these refinements are supposed to capture the "forward-induction" assumption that players try to rationalize the observed behavior of their opponents in order to make inferences about their private information and/or strategic intent. This kind of strategic considerations have been extensively studied within a class of simple dynamic games of incomplete information: signaling games, i.e. leader-follower games where (only) the leader knows the state of nature.

In this paper I illustrate a different approach to the analysis of signaling games. I consider a class of iterative solution procedures that take as given some restrictions on players' beliefs about the payoff-types and the strategies of their opponents. Such solution procedures are akin to extensive-form rationalizability (Pearce (1984)).

As in the complete information case, there are several possible definitions of the rationalizability solution concept for dynamic games, corresponding to different assumptions about how players would update their beliefs if they observed unexpected behavior. Here I consider a class of rationalizability procedures capturing different notions of forward induction. Each procedure corresponds to a parametrically given pair of subsets of first-order beliefs (about the opponent's payoff-types and strategies) of the Sender (Player 1) and Receiver (Player 2). These procedures, on top of common belief in rationality, also capture the assumption that the Receiver always tries to "rationalize" the observed choice of the Sender, that is, he ascribes to the Sender the highest degree of "strategic sophistication" consistent with the Sender's message, given common knowledge of the explicit restrictions of first-order beliefs.[6] (Of course, the case of no explicit restriction on first-order beliefs is also consistent with our general analysis.)

---

[6] I will be more explicit and precise in Subsection 2.4.

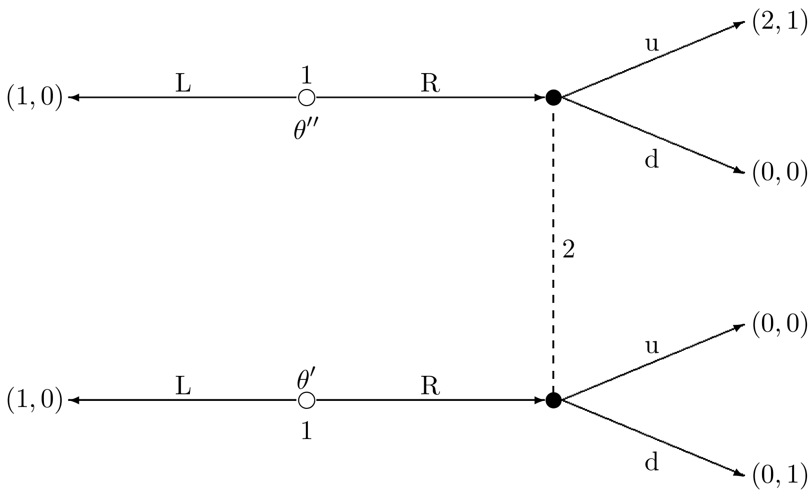

Figure 1

To illustrate, consider the example depicted in Figure 1. In a signaling game the states of nature correspond to the payoff-types of the Sender. There are two conceivable states of nature, θ' and θ''. The Sender can go left (L) or right (R). Action (message) L terminates the game. The Receiver can respond to right with up (u) or down (d). The first number at each terminal node is the Sender's payoff, the second number is the Receiver's payoff. Note that the figure does not represent an initial chance move selecting the type, nor prior probabilities attached to θ' and θ''. In this particular example, I do not consider any explicit restriction on first-order beliefs. Action R is dominated when the state of nature is θ'. But, if the state is θ'', R is a best response to the conjecture that the Receiver would play up with probability at least 1/2. Thus, the only way to rationalize R is to believe that the state is θ'', and the best response to R given this belief is up. The Sender anticipates this response. Therefore the rationalizable solution is that the Sender chooses left if the state is θ', and right if the state is θ'', and the Receiver plays the strategy "up if right".[7]

I apply this approach to some examples and economic models. In some cases it is possible to obtain the same qualitative results as in the more standard equilibrium analysis based on Bayesian games with small type spaces. In other cases weaker results obtain.

I first consider a model due to Sanford Grossman[8] whereby the Sender can make statements about his type (e.g., "quality") and would like to convince the Receiver that his type is as high as possible. Such statements are certifiable (hence truthful) but may be only partially revealing. The standard result

[7] This is the result we would obtain in any Bayesian game based on Figure 1 by looking at the Bayesian-Nash equilibria satisfying the test of dominated messages.

[8] See Grossman (1981) and also Grossman and Hart (1980).

obtained in the literature is that in every (perfect Bayesian) equilibrium the Sender fully discloses his true type. Under a very weak restriction on beliefs, I obtain the same result with the rationalizability solution procedure.

Next I analyze a version of Spence's job market signaling model whereby education complements ability in enhancing productivity. As is well known, for any prior distribution on abilities, this model has a continuum of pooling and separating (perfect Bayesian) equilibrium outcomes, but only the most efficient separating equilibrium outcome passes the Intuitive Criterion, a forward-induction refinement. Under very weak restrictions on beliefs, I show that the same outcome is selected by the rationalizability solution procedure, provided that high ability workers are sufficiently different from low ability workers. If this condition does not hold, rationalizability only yields a lower and upper bound on education.

Notions of rationalizability incorporating explicit restrictions on beliefs are discussed in Rabin (1994) and used in the analysis of specific economic models by Watson (1993, 1996, 1998), Cho (1994, 2003, 2004), Battigalli and Watson (1997), Perry and Reny (1999), Battigalli (2001), Battigalli and Siniscalchi (2003b) and Dekel and Wolinsky (2003). Notions of rationalizability incorporating forward induction assumptions in the context of dynamic games of incomplete information are discussed and analyzed in Sobel *et al.* (1990), Battigalli and Siniscalchi (2002, 2003a) and Battigalli (2003). Focusing on finite games, Battigalli and Siniscalchi (2002) provide a complete epistemic characterization of such solution procedures, and Battigalli and Siniscalchi (2003a) relate them to the Bayesian equilibrium and self-confirming equilibrium concepts. Battigalli (2003) provides existence and characterization results for infinite dynamic games. Hu (2004) provides epistemic characterizations and robustness results for infinite games with simultaneous moves. Bergemann and Morris (2003, 2004) study implementation and mechanism design with large type spaces. Dekel *et al.* (2003) and Ely and Peski (2004) analyze *interim* (correlated) rationalizability in Bayesian games. Restricting attention to games with simultaneous moves, the main difference between interim rationalizability and the class of solution procedures put forward by Battigalli and Siniscalchi (2003a) is that the former takes as given a specific type space *à la Harsanyi* while the latter refer only to payoff-types. But (not surprisingly) interim rationalizability is equivalent to the Battigalli-Siniscalchi procedure that takes as given the restrictions on beliefs implied by the type space.[9]

The remainder of the paper is organized as follows. Section 2 introduces the general definition of rationalizability in signaling games for given restrictions on beliefs. This solution concept is applied to a simple model of disclosure in Section 3 and to a model of job market signaling in Section 4. Section 5 offers some concluding remarks. The Appendix contains a more general specification of the disclosure model and the most tedious proofs.

[9] More precisely, the equivalence holds whenever the restrictions on beliefs implied by the given type space only reflect common knowledge restrictions on first-order beliefs.

2 Rationalization in Signaling Games

2.1 Signaling Games

A signaling game is a two-stage game with incomplete information on one side where the informed party (Player 1, or *Sender*) chooses a "message" m from some set M and the uninformed party (Player 2, or *Receiver*) responds with an action a from some set A. Here I assume, without substantial loss of generality, that the set of feasible messages of the Sender does not depend on his private information and that the set of feasible responses for the Receiver does not depend on the message sent by the Sender. Therefore a signaling game can be represented as a mathematical structure

$$\Gamma = \langle \Theta, M, A, u, v \rangle$$

with the following interpretation:

- Θ is the (nonempty) set of conceivable *payoff-types* and coincides with the set of *states of nature*, an element $\theta \in \Theta$ represents what the Sender might know about how messages and responses are associated to payoffs;
- M is the (nonempty) set *feasible messages* by the Sender;
- A is the (nonempty) set of *feasible responses* by the Receiver;
- $u : \Theta \times M \times A \to \mathbb{R}$ is the Sender's payoff function, $v : \Theta \times M \times A \to \mathbb{R}$ is the Receiver's payoff function.

The set of *strategies* of the Receiver is $S_2 = A^M$.[10]

To illustrate, in the example depicted in Figure 1 one has $\Theta = \{\theta', \theta''\}$, $M = \{\mathrm{L}, \mathrm{R}\}$, $A = \{\mathrm{u}, \mathrm{d}\}$. Payoff function u is given by the first number at each terminal node, and payoff function v by the second number. In particular, to represent that fact that the Receiver is inactive after message L and that payoffs after this message are as in Figure 1, I let $u(\theta,\mathrm{L}, a) = 1$ and $v(\theta,\mathrm{L}, a) = 0$ for all $\theta \in \Theta$ and $a \in A$. Therefore, the action of the Receiver after message L can be omitted from the graphical representation.

In the applications in Sections 3 and 4 I analyze infinite signaling games, but to avoid technicalities in the abstract analysis, I assume in this section that Θ, M, and A are *finite* sets.[11]

Note that I am *not* including in Γ any representation of the Receiver's beliefs about the Sender's payoff-type. Therefore Γ is not a game is the usual technical sense. In order to obtain a Bayesian (extensive-form) game from Γ, one has to append to Γ a type space based on Θ (for more on this, see Subsection 2.6).

[10] For given sets X and Y, Y^X denotes the set of functions from X to Y.

[11] See Battigalli (2003) for an analysis of rationalizability in infinite dynamic games of incomplete information.

2.2 First-order beliefs and best responses

The first-order beliefs of the players describe their (probabilistic) conjectures about the payoff-type and behavior of the opponent as the play unfolds. Higher-order beliefs will not be explicitly represented in the formalism of this paper. Therefore I will often omit the "first-order" qualification.

Since the Receiver has no private information and the Sender moves only once, at the beginning of the game, I can simply represent the Sender's beliefs about the Receiver as a probability measure $\mu^1 \in \Delta(S_2)$. As is well known, the Sender's beliefs can be equivalently represented by a vector of conditional probability measures $\pi \in [\Delta(A)]^M$ (formally, π corresponds to a behavioral strategy of the Receiver). I let $\pi(a|m;\mu^1)$ denote the conditional probability of action a given m derived from belief μ^1:

$$\pi(a|m;\mu^1) := \mu^1\left(\{s_2 \in S_2 : s_2(m) = a\}\right).$$

The Receiver has beliefs about the Sender's payoff-type and behavior. Furthermore, the Receiver updates his initial beliefs after receiving a message, using Bayes rule whenever he has initially assigned a strictly positive probability to the message actually received. Therefore the Receiver's beliefs are represented by some system of conditional probabilities

$$\mu^2 = \left(\mu^2(\cdot|\phi), \left(\mu^2(\cdot|m)\right)_{m\in M}\right) \in \Delta(\Theta \times M) \times [\Delta(\Theta)]^M$$

(where ϕ is the "empty history" and $\mu^2(\cdot|\phi)$ is the initial belief of the Receiver) such that, for all $m \in M$ and all $\theta \in \Theta$,

(i) $\mu^2(\Theta(m)|m) = 1$ (the Receiver believes what he observes)

(ii) if $\mu^2(\Theta \times \{m\}|\phi) > 0$, then

$$\mu^2(\theta|m) = \frac{\mu^2((\theta, m)|\phi)}{\mu^2(\Theta \times \{m\}|\phi)}.$$

The set of conditional probability systems satisfying (i) and (ii) is denoted $\Delta^*(\Theta, M)$.

The best response correspondence for the Sender is $BR_1 : \Theta \times \Delta(S_2) \twoheadrightarrow M$, where

$$\forall\theta \in \Theta,\ \forall\mu^1 \in \Delta(S_2),\ BR_1(\theta, \mu^1) := \arg\max_m \left\{\sum_a u(\theta, m, a)\pi(a|m, \mu^1)\right\}.$$

The best response correspondence for the Receiver is $BR_2 : M \times \Delta(\Theta) \twoheadrightarrow A$, where

$$\forall m \in M,\ \forall p \in \Delta(\Theta),\ BR_2(m, p) := \arg\max_a \left\{\sum_\theta v(\theta, m, a)p(\theta)\right\}.$$

Thus, a (sequentially) rational Receiver with a system of conditional beliefs $\mu^2 \in \Delta^*(\Theta, M)$ follows a strategy s_2 such that $s_2(m) \in BR_2(m, \mu^2(\cdot|m))$ for all $m \in M$.

2.3 Explicit assumptions on beliefs

Roughly speaking, I call "explicit" those assumptions about first-order beliefs that are not derived from iterated mutual belief in rationality, and I represent them with *given* restricted sets of beliefs $\Delta = (\Delta^1, \Delta^2)$, $\Delta^1 \subseteq \Delta(S_2)$, $\Delta^2 \subseteq \Delta^*(\Theta, M)$. For example, if it is assumed that (a) *the Receiver initially believes that* θ' *is as least as likely as* θ'', then attention is restricted to the set

$$\Delta^2 = \left\{ \mu^2 \in \Delta^*(\Theta, M) : \mu^2(\{\theta'\} \times M|\phi) \geq \mu^2(\{\theta''\} \times M|\phi) \right\}.$$

If the Sender believes (a) and he also believes that the Receiver is rational, then the Sender's belief μ^1 must satisfy

$$\mu^1 \left(\left\{ s_2 : \exists \mu^2 \in \Delta^2, \forall m \in M,\ s_2(m) \in BR_2(m, \mu^2(\cdot|m)) \right\} \right) = 1.$$

I do not call this restriction on μ^1 "explicit" because it is not assumed at the outset, but rather derived by the standard assumption that what the modeler assumes about the Receiver (in this case (a) and the Receiver's rationality) is also believed by the Sender.[12]

2.4 Rationalization

Fix some explicit restrictions about first-order beliefs represented by the pair of subsets $\Delta = (\Delta^1, \Delta^2)$. As explained in the Introduction, I would like to define an iterative solution procedure that captures a form of forward-induction reasoning based on the "rationalization" of the messages of the Sender given the restrictions Δ. More specifically the procedure reflects the following assumptions about behavior and interactive beliefs:

(A1.S) The Sender is rational and has beliefs in Δ^1
(A1.R) The Receiver is rational and has beliefs in Δ^2

(A2.S) The Sender believes (A1.R)
(A2.R) The Receiver believes (A1.S) *whenever possible* (that is, he initially believes (A1.S) and continues to do so after each message m consistent with (A1.S))

....

(Ak+1.S) The Sender believes (A1.R),...,(Ak.R)
(Ak+1.R) The Receiver believes (A1.S),...,(Ak.S) *whenever possible*

....

Assumptions (A2.R),..., (Ak.R) capture a notion of *forward induction*: even if he is "surprised" by a message, the Receiver tries to *rationalize* the observed message in a way which is consistent with the Sender being strategically sophisticated. The higher the index k, the higher the degree of strategic sophistication ascribed to the Sender.

[12] One could consider type-dependent explicit restrictions, that is, $\Delta^1 = (\Delta^1_\theta)_{\theta \in \Theta}$, with $\Delta^1_\theta \subseteq \Delta(S_2)$ for all θ (cf. Battigalli and Siniscalchi (2003a)). I ignore this generalization here because it is not relevant in the examples and applications of this paper.

Note that assumption (Ak+1.R) does not imply assumption (Ak.R). The reason is that there may be some message m which is consistent with (A1.S),...,(Ak-1.S), but inconsistent with (A1.S),...,(A.k-1.S),(Ak.S). In this case (Ak.R) implies that, upon observing m, the Receiver believes (A1.S),...,(Ak-1.S), while (Ak+1.R) does not have any implication about the Receiver's beliefs after m. For further discussion of this point see Battigalli and Siniscalchi (2002).

Battigalli and Siniscalchi (2002) formally express these assumptions (by means of "complete extensive-form type spaces") and show that the pairs of payoff-types and messages of the Sender, and the strategies of the Receiver consistent with assumptions (A1)-(Ak) are those and only those which belong, respectively, to the subsets $\Sigma_1(k, \Delta)$ and $S_2(k, \Delta)$ defined by the following procedure that iteratively deletes type-message pairs for the Sender and strategies for the Receiver:

- Let $\Sigma_1(0, \Delta) = \Theta \times M$ and $S_2(0, \Delta) = S_2$.
- For $k = 1, 2, ...$, let $\Theta(m, k-1, \Delta) := \{\theta : (\theta, m) \in \Sigma_1(k-1, \Delta)\}$ (set of types consistent with step $k-1$ and message m), then

$$\Sigma_1(k, \Delta) =$$

$$\left\{(\theta, m) \in \Sigma_1(k-1, \Delta) : \exists \mu^1 \in \Delta^1,\, m \in BR_1(\theta, \mu^1) \text{ and } \mu^1(S_2(k-1, \Delta) = 1\right\},$$

$$S_2(k, \Delta) = \{s_2 \in S_2(k-1, \Delta) : \exists \mu^2 \in \Delta^2,\, \forall m, s_2(m) \in BR_2(m, \mu^2(\cdot|m)),$$

$$\Theta(m, k-1, \Delta) \neq \emptyset \;\Rightarrow\; \mu^2(\Theta(m, k-1, \Delta)|m) = 1\}.$$

The last equation corresponds to the assumption that the Receiver rationalizes, if possible, the observed message.

Definition 1 *Fix a pair of subsets of beliefs $\Delta = (\Delta^1, \Delta^2)$, where $\emptyset \neq \Delta^1 \subseteq \Delta(S_2)$ and $\emptyset \neq \Delta^2 \subseteq \Delta^*(\Theta, M)$. A message m is* (k, Δ)-rationalizable *for θ if $(\theta, m) \in \Sigma_1(k, \Delta)$; strategy s_2 is (k, Δ)-rationalizable if $s_2 \in S_2(k, \Delta)$. Message m is* Δ-rationalizable *for θ if $(\theta, m) \in \Sigma_1(\infty, \Delta) := \bigcap_k \Sigma_1(k, \Delta)$; strategy s_2 is Δ-rationalizable if $s_2 \in S_2(\infty, \Delta) := \bigcap_k S_2(k, \Delta)$.*[13]

Remark 2 *By finiteness of Σ_1 and S_2, there is some index K such that $\Sigma_1(K, \Delta) = \Sigma_1(\infty, \Delta)$ and $S_2(K, \Delta) = S_2(\infty, \Delta)$.*

[13] In Battigalli (2003) I use the phrase *strong Δ-rationalizability* and I compare this solution concept to a weaker one which does not capture forward-induction reasoning. The weak rationalizability concept is not very interesting in signaling games.

The simplest illustration of this solution procedure is given by the forward induction solution of the game depicted in Figure 1, which I informally discussed in the Introduction. In this example, there are no explicit restrictions on beliefs (i.e., $\Delta = (\Delta(S_2), \Delta^*(\Theta, M)))$, therefore I omit Δ from the notation. It can be easily checked that $\Sigma_1(1) = \{(\theta', \mathrm{L}), (\theta'', \mathrm{L}), (\theta'', \mathrm{R})\}$. Since there is a pair $(\theta, m) \in \Sigma_1(1)$ such that $m = \mathrm{R}$ (i.e., $\Theta(\mathrm{R}, 1) \neq \emptyset$), then the Receiver rationalizes message R, and the only possible rationalization is that the Sender's type must be θ''. Thus $S_2(\infty) = S_2(2) = \{\mathrm{u}\}$ and $\Sigma_1(\infty) = \Sigma_1(3) = \{(\theta', \mathrm{L}), (\theta'', \mathrm{R})\}$.

The set of Δ-rationalizable profiles may be empty because the explicit restrictions on beliefs represented by Δ may conflict with iterated mutual belief in rationality. I here report two existence results proved elsewhere (Battigalli 2003, Battigalli and Siniscalchi 2003a).

First, it can be proved by standard methods that if Δ only restricts the initial beliefs of the Receiver about the state of nature, then the Δ-rationalizable solution is non-empty. This also holds in infinite signaling games if some regularity assumptions are satisfied.

The second result concerns the special case where the restrictions Δ state that players' beliefs "agree" with a particular distribution on the terminal nodes of the arborescence representing the signaling game, say $\zeta \in \Delta(\Theta \times M \times A)$. This may be the case when Sender and Receiver are repeatedly drawn at random from large heterogenous populations and joint statistics about payoff-types and actions are made public, so that beliefs reflect these statistics. Then, it can be shown that the Δ-rationalizable solution is not empty if and only if distribution ζ is a self-confirming equilibrium outcome satisfying the Intuitive Criterion of Cho and Kreps (1987).[14]

2.5 A "Beer-Quiche" example

The game depicted in Figure 2 corresponds to the well-known Beer-Quiche example used by Cho and Kreps (1987) to discuss equilibrium refinements in signaling games. Of course, Cho and Kreps analyze a standard extensive-form game with a common prior on the set of payoff-types. In their example the **surly** type $\theta(\sigma)$ has prior probability $\frac{9}{10}$. They show that only the equilibrium whereby each type chooses B satisfies their Intuitive Criterion.

I mentioned above how Δ-rationalizability can be used to characterize the Intuitive Criterion. But the analysis of this subsection is not related to this result. Here I illustrate the Δ-rationalizability procedure showing that the same result obtained by Cho and Kreps for their Quiche-Beer example can be obtained with very weak restrictions on beliefs. I assume that (it is common belief that) the prior probability assigned by player 2 to $\theta(\sigma)$ is more than $\frac{1}{2}$. Furthermore, I also assume that (it is common belief that) player 2's posterior probability of the surly type $\theta(\sigma)$ is higher after observing B (**beer**) than after observing Q

[14] Cho and Kreps defined a refinement of sequential equilibrium; but their criterion can be applied to any self-confirming equilibrium distribution.

(**quiche**). Thus the restricted set of beliefs for player 2 is

$$\Delta^2 = \left\{\mu^2 : \mu^2(\theta(\sigma)|\phi) > 1/2, \mu^2(\theta(\sigma)|Q) < \mu^2(\theta(\sigma)|B)\right\}$$

(I use obvious abbreviations for marginal probabilities). There are no restrictions on player 1's beliefs.

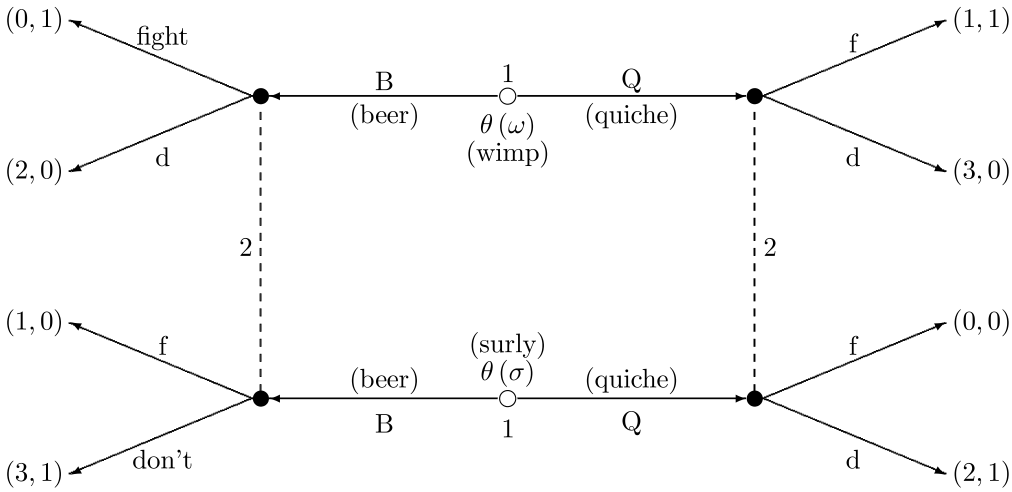

Figure 2

First step: It is easy to see that both Q and B are $(1, \Delta)$-rationalizable for both types.[15] But the second restriction on beliefs implies that, if the Receiver fights after B, he also fights after Q. Thus, strategy $[f \text{ if } B,\ d \text{ if } Q]$ (fight after beer, don't fight after quiche) is deleted, and

$$S_2(1, \Delta) = \{[f \text{ if } B,\ f \text{ if } Q], [d \text{ if } B,\ f \text{ if } Q], [d \text{ if } B,\ d \text{ if } Q]\}.$$

Second step: This in turn implies that a fight after B is less likely than a fight after Q. Formally,

$$\begin{aligned} \mu^1(S_2(1,\Delta)) &= \mu^1\left(\{[f \text{ if } B,\ f \text{ if } Q], [d \text{ if } B,\ f \text{ if } Q], [d \text{ if } B,\ d \text{ if } Q]\}\right) = 1 \ \Rightarrow \\ \pi(f|B;\mu^1) &= \mu^1\left([f \text{ if } B,\ f \text{ if } Q]\right) \leq \mu^1\left(\{[f \text{ if } B,\ f \text{ if } Q], [d \text{ if } B,\ f \text{ if } Q]\}\right) = \pi(f|Q;\mu^1) \end{aligned}$$

Since the only reason for a surly type to have quiche for breakfast is to decrease the probability of a fight, the only $(\Delta, 2)$-rationalizable choice for the surly type $\theta(\sigma)$ is B. On the other hand, it makes sense for type $\theta(\omega)$ (**wimp**) to forgo his preferred breakfast (quiche) hoping to avoid a fight. Thus,

$$\Sigma_1(2, \Delta) = \{(\theta(\sigma), B), (\theta(\omega), B), (\theta(\omega), Q)\}.$$

[15] Note that if $\Sigma_1(1, \Delta) = \Sigma_1$, then $S_2(k+1, \Delta) = S_2(k, \Delta)$ for k odd, and $\Sigma_1(k+1, \Delta) = \Sigma_1(k, \Delta)$ for k even. Therefore we may consider only one player at each step.

Third step (forward induction): Since Q is a $(2,\Delta)$-rationalizable choice for $\theta(\omega)$, but not for $\theta(\sigma)$, Q is sure evidence that player 1 is a wimp ($\theta=\theta(\omega)$). Formally $\Theta(Q,2,\Delta)=\{\theta(\omega)\}\neq\emptyset$ implies $\mu^2(\theta(\omega)|Q)=1$.

Furthermore, since the only $(2,\Delta)$-rationalizable choice for $\theta(\sigma)$ is B, observing B cannot decrease the probability of $\theta(\sigma)$.[16] Therefore $\mu^2(\theta(\sigma)|B)\geq\mu^2(\theta(\sigma)|\phi)>1/2$, where the latter inequality is an explicit restriction of beliefs. This implies that the unique Δ-rationalizable strategy for Player 2 is "fight after quiche, don't fight after beer." To summarize:

$$\bigcap_k S_2(k,\Delta)=S_2(3,\Delta)=\{[d\,\text{if}\,B,\ f\,\text{if}\,Q]\}.$$

Fourth step. Given this, the only Δ-rationalizable choice for type $\theta(\omega)$ is B:

$$\bigcap_k \Sigma_1(k,\Delta)=\Sigma_1(4,\Delta)=\{(\theta(\omega),B),(\theta(\sigma),B)\}.$$

[Note that Q is *not* rationalizable for either type, but the best Δ-rationalization of message Q is that the state of nature must be $\theta(\omega)$ and that Player 1, not having fully rationalizable beliefs, chooses his *coeteris paribus* preferred breakfast.]

2.6 Comparison with Bayesian games

The simplest and most common way to obtain a Bayesian game from the signaling game $\Gamma=\langle\Theta,M,A,u,v\rangle$ is to add a probability measure on the states of nature, $\rho\in\Delta(\Theta)$. The interpretation would be that it is common belief at the beginning of the game that the Receiver's beliefs about the Sender's payoff-type are given by ρ. Therefore payoff-types are in one-to-one correspondence with Harsanyi-types.

More generally, one may obtain a Bayesian game by appending to Γ a larger type space *à la* Harsanyi based on Θ, that is, a structure $\langle\Theta,B_1,B_2,\tau_1,\tau_2\rangle$, where B_i is a set of "purely epistemic parameters" for Player i, $\tau_1:\Theta\times B_1\to\Delta(B_2)$ and $\tau_2:B_2\to\Delta(\Theta\times B_1)$ are functions specifying the players' beliefs for each epistemic state. Pairs $(\theta,b_1)\in\Theta\times B_1$ and parameters $b_2\in B_2$ are "Harsanyi-types". The first-order beliefs of the Receiver, if his Harsanyi-type is b_2, are given by $\tau_2^1(b_2):=\text{marg}_\Theta\tau_2(b_2)\in\Delta(\Theta)$. The second-order beliefs of Harsanyi-type (θ,b_1) of the Sender about the first-order beliefs of the Receiver are obtained from the probability measure $\tau_1(\theta,b_1)\in\Delta(B_2)$ and the Receiver's first-order belief function $\tau_2^1(\cdot):B_2\to\Delta(\Theta)$. For example, the probability

[16] By Bayes rule,

$$\begin{aligned}\mu^2(\theta(\sigma)|B) &= \frac{\mu^2(B|\theta(\sigma))\mu^2(\theta(\sigma)|\phi)}{\mu^2(B|\theta(\sigma))\mu^2(\theta(\sigma)|\phi)+\mu^2(B|\theta(\omega))\mu^2(\theta(\omega)|\phi)}\\ &= \frac{\mu^2(\theta(\sigma)|\phi)}{\mu^2(\theta(\sigma)|\phi)+\mu^2(B|\theta(\omega))\mu^2(\theta(\omega)|\phi)}\geq\mu^2(\theta(\sigma)|\phi).\end{aligned}$$

that Harsanyi-type $t_1 \in \Theta \times B_1$ would assign to the event *"the Receiver assigns probability at least $\frac{1}{2}$ to payoff-type θ"* is $\tau_1(t_1)\left(\left\{b_2 \in B_2 : \tau_2^1(b_2)(\theta) \geq \frac{1}{2}\right\}\right)$. Higher and higher order beliefs can be derived in a similar fashion.

The reason why I do *not* append a type space to Γ is twofold.

(i) On the one hand, I want to be able to consider assumptions about beliefs, such as

"At the beginning of the game, the Receiver assigns probability at least $\frac{1}{2}$ to payoff-type θ, and there is common certainty of this fact."

If this is the only assumption about initial (interactive) beliefs one is willing to make, then one has to consider a type space so large and complex that it is possible to represent it and analyze the equilibria of the corresponding Bayesian game only through indirect methods. These indirect methods amount to a kind of iterative deletion procedure somewhat similar to the one put forward in subsection 2.4. This iterative deletion procedure is called *weak rationalizability* in Battigalli (2003) because, unlike the rationalizability procedure defined in this paper, it does not capture any kind of forward induction reasoning.

This is to be contrasted with the standard applications of Harsanyi's theory, which consider extremely simple type spaces corresponding to implausibly strong assumptions on interactive beliefs, directly compute the (relatively few) equilibria, and maybe proceed to apply some refinement to get rid of the "implausible" ones.

(ii) On the other hand, one might want to take as given some assumptions about beliefs concerning the opponents' behavior and/or assumptions about how players update their beliefs. Consider for example the following assumptions:
(a) *"The conditional probability the Receiver would assign to payoff-type θ after message m^2 is higher than the conditional probability he would assign to θ after message m^1"*, or
(b) *"The Receiver would believe that the Sender is rational whenever this is consistent with the Sender's message"*,
as well as further assumptions concerning interactive beliefs about (a) and (b).

These assumptions involve conditional beliefs about the opponent's payoff-type *and* behavior. Thus they cannot be represented using a type space of the form $\langle \Theta, B_1, B_2, \tau_1, \tau_2 \rangle$ (i.e. a type space *à la* Harsanyi based on Θ), which only describes possible beliefs about the Sender's payoff-type, possible beliefs about such beliefs, etc. In order to represent such assumptions one would have to work with the more complex "dynamic" type spaces first put forward by Ben Porath (1997) and fully analyzed in Battigalli and Siniscalchi (1999). This means going beyond Harsanyi's methodology because such type spaces include (1) beliefs about behavior (which Harsanyi opposed on the ground that all such beliefs must be endogenously derived through equilibrium analysis) and (2) the conditional beliefs that the Receiver would hold after each message. Note that including (1) is a necessary condition for including (2): without (1) it is impossible to relate conditional and unconditional beliefs *via* Bayes rule. Dynamic type spaces of this kind can be used to formally express the above mentioned assumptions, but - again - in order to find the behavioral consequences of (rationality and) these assumptions about beliefs, one must use characterization results saying something like "the only outcomes consistent with (a), (b), rationality (r), and assumptions concerning interactive believes about (a), (b) and (r) are those found with solution procedure $\mathcal{S}$" (see Battigalli and Siniscalchi (2002)).

Here I rely on these characterization results, which allow me to use the solution procedure with no direct reference to type spaces of any sort.

3 Application (i): Disclosure[17]

Consider a two-person signaling game where the Sender, Player 1, provides *certifiable information* about the state of nature (his payoff-type). The Receiver, Player 2, observes the Sender's message and then takes an action affecting the Sender's payoff as well as his own. For concreteness, the Sender may be thought of as a seller, the Receiver as a buyer. The state of nature $\theta \in \Theta$ can be thought of as the quality of the product and the Receiver's action, $a \in A$, as the quantity purchased or the total price paid. In this section I analyze a simplified model. The Appendix contains a much more general analysis.

The set of states of nature is the finite set of integers $\Theta = \{1, 2, ..., K\}$. Player 1 can send messages of the form "The state of nature is *at least* k". I denote such a message with the symbol $[\theta \geq k]$. If the Sender does not tell the truth, this is verified and he pays a very large fine. Thus, rationality implies that he always tells the truth. However, rationality *per se* does not rule out "understatements", i.e. Player 1 could send message $[\theta \geq k]$ even if the true state is $k^* > k$. Player 2 responds with an action $a \in A = [0, +\infty)$. The Sender's payoff is a strictly *increasing* function of a. The Receiver preferences are given by a loss function $L(\theta, a) = -(\theta - a)^2$, thus, the Receiver always wants to choose his (conditional) estimate of state of nature $[BR_2(p) = \mathrm{E}^p(\theta)]$, and a rational Sender who anticipates a rational response would like to induce with his message the highest possible estimate by the Receiver.

There is no exogenous restriction on the Sender's beliefs (i.e., $\Delta^1 = \Delta(S_2)$). As for the Receiver, I consider a very weak restriction:

(*Mild skepticism*) When Player 2 receives message $m = [\theta \geq k]$ he assigns positive probability to k, the lowest state consistent with m (under the assumption that m is true). That is

$$\Delta^2 = \{\mu \in \Delta^*(\Theta, M) : \forall k, \mu(k|[\theta \geq k]) > 0\}.$$

I show that Δ-rationalizability yields *full disclosure*, that is, Player 1 never makes "understatements" and, for each k, Player 2 responds to message $[\theta \geq k]$ with action (estimate) $a = k$.

To see this, first note that every message is consistent with the Sender's rationality, and a rational Sender always tells the truth to avoid punishments. Therefore, by forward induction, the Receiver always believes that the observed message is true. In particular, this means that if he observes message $[\theta \geq K]$

[17] The model of information transmission of this section builds on Grossman and Hart (1980) and Grossman (1981). See also Okuno-Fujiwara *et al.* (1990), Bolton and Dewatripont (1997, Chapter 5) and the references therein.

his estimate of the state is $a = K$, but if he observes any other message $[\theta \geq k]$ (with $k < K$) – by mild skepticism – his estimate is strictly below the highest state K. Anticipating this, the Sender chooses message $[\theta \geq K]$ when the state is K because this induces the highest estimate.

Now suppose that the Receiver observes message $[\theta \geq K - 1]$. What is the "best rationalization" of this message? The Receiver reasons that if the state were K the Sender would try to induce the highest estimate by choosing message $[\theta \geq K]$ and therefore he would not make the "understatement" $[\theta \geq K - 1]$, and since a rational Sender always tells the truth, it must be the case that the state is indeed $K-1$. Anticipating this, if the state is $K-1$ the Sender chooses message $[\theta \geq K-1]$ because it induces the highest possible estimate among the truthful messages.

Assume by way of induction that Δ-rationalizability implies that for some integer ℓ and each state $k \geq K - \ell$ the Sender chooses message $[\theta \geq k]$, and that the Receiver's estimate conditional on each message $[\theta \geq k']$ such that $k' \geq K - \ell$ is $a = k'$. Then a similar "best rationalization" argument shows that if the Receiver observes message $[\theta \geq K - \ell - 1]$ his estimate is precisely $a = K - \ell - 1$. Therefore

Proposition 3 *For each $k \in \Theta$, the unique Δ-rationalizable message for k is $[\theta \geq k]$ and the unique Δ-rationalizable response to message $[\theta \geq k]$ is $a = k$.*

The argument above is similar to an *intuitive* "unraveling" argument[18] used to show why a perfect Bayesian equilibrium that passes the test of dominated messages must satisfy full disclosure (since sending false messages is dominated, the test of dominated messages guarantees that the Receiver would believe the literal meaning of every message, including those off the equilibrium path).[19] The compellingness of such "unraveling" arguments is due to their inductive structure. But a *rigorous* proof of the equilibrium result, one way or the other, has to proceed by contradiction.

The key step is that the equilibrium must be separating. Let m be a message on the equilibrium path, and let $\Theta^*(m)$ be the set of types sending message m with positive probability in equilibrium; if $\Theta^*(m)$ is not a singleton, then m is an "understatement" for the payoff type $\theta^*_m := \max \Theta^*(m)$ and the equilibrium estimate conditional on m is below θ^*_m, therefore payoff-type θ^*_m would be strictly better off sending message $[\theta \geq \theta^*_m]$, which contradicts the equilibrium assumption. Thus $\Theta^*(m)$ must be a singleton for each message on the equilibrium path, that is, the equilibrium must be separating. It follows that the equilibrium must satisfy full disclosure.

As with other applications of equilibrium analysis, the mathematical argument is simple enough, but it does not show why strategic reasoning should make the players hold equilibrium beliefs in the first place.

Note also that more general Bayesian extensions of the given economic model, whereby belief functions are consistent with a common prior on the

[18] See, e.g., Chapter 5 of Bolton and Dewatripont (2005).

[19] Instead of applying the test of dominated messages, most disclosure models directly assume that the Sender is *constrained* to tell the truth.

set of states of the world but Harsanyi-types and payoff-types do not coincide, may have perfect Bayesian equilibria which satisfy mild skepticism and pass the test of dominated messages, and yet do *not* satisfy full disclosure *off* the equilibrium path (an example is provided in the Appendix).

4 Application (ii): Job Market Signaling

Consider a standard game-theoretic version of Spence's model of job market signaling with two types of workers (see e.g. Cho and Kreps (1987)).

Player 1, a worker of ability θ' or θ'', with $0 < \theta' < \theta''$, chooses an education level $e \in [0, +\infty)$ and has payoff function $u(\theta, e, w) = w - c(\theta, e)$, where $c : \mathbb{R}^2_+ \to \mathbb{R}_+$ is a smooth cost function that satisfies the standard assumption that the marginal cost of education is positive, increasing in e and decreasing in ability ($\frac{\partial c(\theta,e)}{\partial e} > 0$, $\frac{\partial^2 c(\theta,e)}{\partial e^2} > 0$, $\frac{\partial^2 c(\theta,e)}{\partial e \partial \theta} > 0$).

Player 2, a "representative firm," observes e and chooses the wage $w \in [0, +\infty)$. Player 2's payoff is $v(\theta, e, w) = -(e\theta - w)^2$ and thus he "rationally" sets the wage equal to the subjectively expected value of $e\theta$ conditional on e.

The restricted set of beliefs for Player 1, Δ^1, is the set of probability measures $\mu^1 \in \Delta(S_2)$ with *countable support*. As for player 2, I assume that Δ^2 is the set of *monotonic* conditional probability systems, that is, the set of μ^2 such that $\mu^2(\theta''|e)$, the conditional probability assigned to the high-ability type, is non-decreasing in e. Countability of supports is merely a technical assumption that simplifies the analysis. Monotonicity is similar to the "plausibility" property postulated by Kreps and Wilson (1982) in their analysis of reputation and entry deterrence.[20]

Player 2's strategies can be represented by functions $\vartheta(e)$ fixing the wage per unit of education. Function $\vartheta(\cdot)$ is a bset response to the system of conditional beliefs μ^2 if and only if $\vartheta(e) = \theta'[1 - \mu^2(\theta''|e)] + \theta''\mu^2(\theta''|e)$. Therefore best responses to beliefs in Δ^2 are in one-to-one correspondence with the set of non-decreasing expectation functions $\vartheta(e)$ with range $[\theta', \theta'']$. Let

$$\Omega(1, \Delta) = \{\vartheta(\cdot) \in [\theta', \theta'']^{\mathbb{R}_+} : e'' > e' \Rightarrow \vartheta(e'') \geq \vartheta(e')\}.$$

$\Omega(1, \Delta)$ is the set of Player 2's $(1, \Delta)$-rationalizable strategies represented as contingent choices of wage per unit of education.

Player 1's $(1, \Delta)$-rationalizable beliefs are summarized by his expectation of Player 2's expectation of θ conditional on the chosen education level e. Let this second-order expectation (which coincides with the expected wage per unit of education) be denoted by $\widehat{\vartheta}(e)$. Assuming that Player 2 is a maximizer (expected-loss minimizer), Player 1 expects to get wage $e\widehat{\vartheta}(e)$, with $\widehat{\vartheta}(\cdot) \in \Omega(1, \Delta)$.[21] At a subjectively optimal choice of education for payoff-

[20] See also the analysis of rationalizable bidding in auctions with interdependent values due to Battigalli and Siniscalchi (2003b) and Cho (2003, 2004).

[21] Fix belief $\mu \in \Delta(S_2)$ with countable support $\{s_2^1(\cdot), ..., s_2^k(\cdot), ...\}$ and corresponding wages per unit of education $\{\vartheta^1(\cdot), ..., \vartheta^k(\cdot), ...\}$. Player 1's expected wage conditional on e is $e\widehat{\vartheta}(e) =$

type θ, say e^*, $\widehat{\vartheta}(\cdot)$ must be continuous from the right and the marginal rate of substitution $MRS(\theta, e^*) = \frac{\partial c(\theta,e)}{\partial e}$ must satisfy the first-order condition

$$MRS(\theta, e^*) \geq \widehat{\vartheta}(e^*) + e^* \cdot \frac{d\widehat{\vartheta}(e^*+)}{de}. \tag{1}$$

where $\frac{d\widehat{\vartheta}(e^*+)}{de}$ is the right-derivative of $\widehat{\vartheta}(\cdot)$ at e^*.[22]

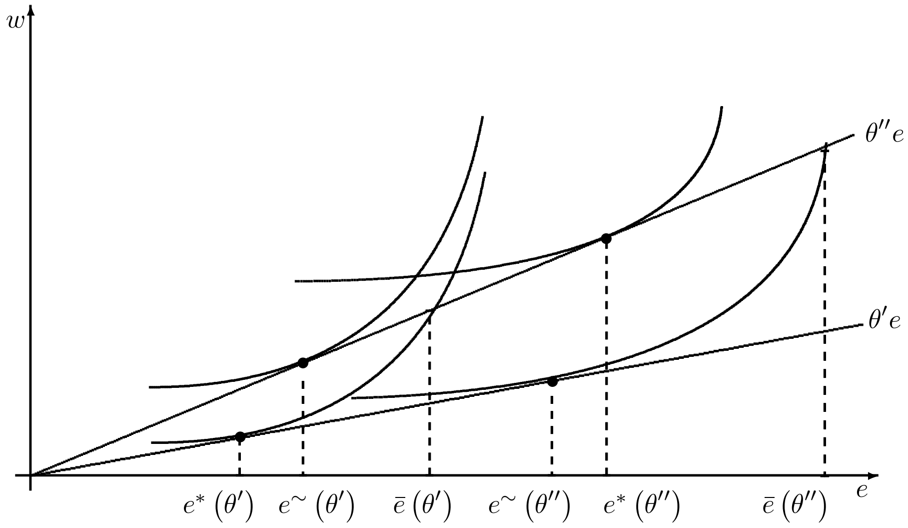

Figure 3

$e \sum_k \mu(s_2^k)\vartheta^k(e)$. Since for each k, $\vartheta^k(\cdot)$ is non decreasing with range in $[\theta', \theta'']$, $\widehat{\vartheta}(\cdot)$ must have the same properties.

[22] More generally, it is the right-limsup of the incremental ratio of $\widehat{\vartheta}(\cdot)$ at e^*.

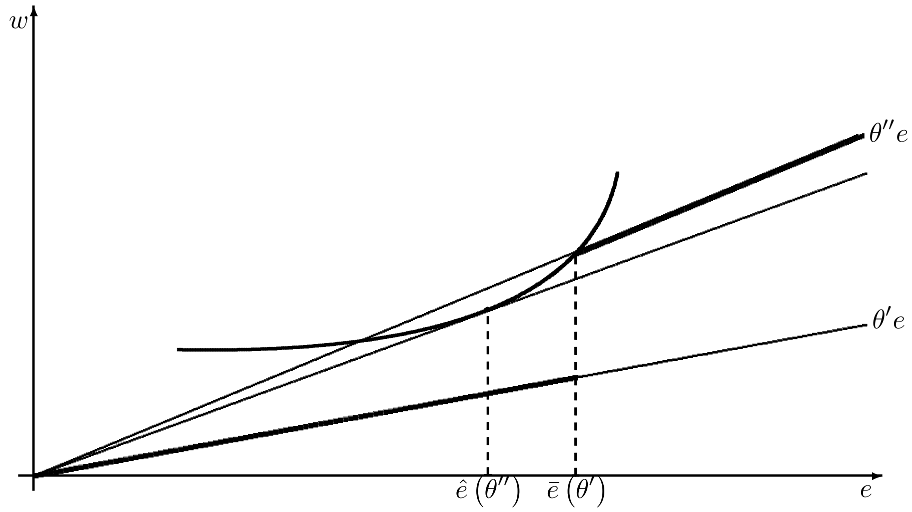

Figure 4

It turns out that the set of Δ-rationalizable choices depends on how close θ' and θ'' are to each other. In particular, it depends on the relation between the following numbers (see Figures 3 and 4):

- $e^*(\theta) = \arg\max_{e\geq 0} u(\theta, e, \theta e)$, $\theta = \theta', \theta''$ (complete information choice),
- $e^{\sim}(\theta) = \arg\max_{e\geq 0} u(\theta, e, \hat{\theta} e)$, $\theta \neq \widehat{\theta}$ ($e^{\sim}(\theta')$ is the choice that payoff-type θ' would make in the "best case scenario" where Player 2 has the unshakable certainty that the true type is θ'', similarly $e^{\sim}(\theta'')$ is the best choice of θ'' in the "worst case scenario"),
- $\overline{e}(\theta')$ solves $u(\theta', e, \theta'' e) = u(\theta', e^*(\theta'), \theta' e^*(\theta'))$ and $\overline{e}(\theta'')$ solves $u(\theta'', e, \theta'' e) = u(\theta'', e^{\sim}(\theta''), \theta' e^{\sim}(\theta''))$,
- $\hat{e}(\theta'')$ solves $u(\theta'', e, MRS(\theta'', e) \cdot e) = u(\theta'', \overline{e}(\theta'), \theta'' \overline{e}(\theta'))$.

If θ' and θ'' are not too close to each other, then $\overline{e}(\theta') \leq e^*(\theta'')$. Note that strict monotonicity, strict convexity of the cost of education and the single-crossing property imply

$$e^*(\theta') < e^{\sim}(\theta') < \overline{e}(\theta') < \overline{e}(\theta''),$$

$$e^*(\theta') < e^{\sim}(\theta'') < \hat{e}(\theta'') \leq e^*(\theta'') < \overline{e}(\theta'').$$

The following result shows that, if θ' and θ'' are not too close to each other, Δ-rationalizability yields the same result as in the most efficient separating equilibria (i.e. those that satisfy the Intuitive Criterion), otherwise Δ-rationalizability only yields bounds on the possible education choices for each ability level.

Proposition 4 *If (a) $\bar{e}(\theta') < e^{\tilde{}}(\theta'')$ or (b) $e^{\tilde{}}(\theta'') \leq \bar{e}(\theta') \leq e^*(\theta'')$, then the unique Δ-rationalizable choice of payoff type $\theta \in \{\theta', \theta''\}$ is the same level of education as in the complete information model, that is, $e^*(\theta)$.*
If (c) $\bar{e}(\theta') > e^*(\theta'')$, then each choice $e \in [\hat{e}(\theta''), \bar{e}(\theta')]$ is Δ-rationalizable for both types and $e^*(\theta')$ is also rationalizable for type θ'.

Proof. Any education level can be justified as a best reply to some belief. Thus $\Sigma_1(1, \Delta) = \Sigma_1$. This implies that $S_2(k+1, \Delta) = S_2(k, \Delta)$, for k odd, and $\Sigma_1(k+1, \Delta) = \Sigma_1(k, \Delta)$ for k even.

Let $\Omega(k, \Delta)$ denote Player 2's (k, Δ)-rationalizable choices of wage per unit of education. In general, (k, Δ)-rationalizable beliefs for Player 1 can be summarized by some function $\widehat{\vartheta}(\cdot) \in \Omega(k, \Delta)$ giving the expected wage per unit of education and having the same properties of Player 2's $(k-1, \Delta)$-rationalizable expectation functions. Let $M(\theta, k, \Delta)$ denote the set of (k, Δ)-rationalizable messages for θ. Then

$$M(\theta', 2, \Delta) = [e^*(\theta'), \bar{e}(\theta')], \ M(\theta'', 2, \Delta) = [e^{\tilde{}}(\theta''), \bar{e}(\theta'')].$$

To see this, first note that for any conjecture $\widehat{\vartheta}(\cdot) \in \Omega(1, \Delta)$ about Player 2, the first order condition (1) for type θ' is necessarily violated for every $e^* < e^*(\theta')$ because strict convexity of the disutility of education, monotonicity of $\widehat{\vartheta}(\cdot)$ and $\widehat{\vartheta}(e) \geq \theta'$ imply

$$MRS(\theta', e^*) < MRS(\theta', e^*(\theta')) = \theta' \leq \widehat{\vartheta}(e^*) + e^* \cdot \frac{d\widehat{\vartheta}(e^*+)}{de}.$$

No education level $e > \bar{e}(\theta')$ can be justified for θ' because, since $\widehat{\vartheta}(e) \leq \theta''$ for all e, type θ' would get a higher expected utility by choosing $e^*(\theta')$. Every $e^* \in [e^*(\theta'), e^{\tilde{}}(\theta')]$ is a best response to the $(1, \Delta)$-rationalizable constant conjecture $\widehat{\vartheta}(e) \equiv MRS(\theta', e^*) \in [\theta', \theta'']$. Every $e^* \in [e^{\tilde{}}(\theta'), \bar{e}(\theta')]$ is a best reply to the $(1, \Delta)$-rationalizable conjecture

$$\widehat{\vartheta}(e) = \begin{cases} \theta' & if \quad e < e^*, \\ \theta'' & if \quad e \geq e^*. \end{cases} \tag{2}$$

$M(\theta'', 2, \Delta)$ is obtained in a similar way. Using forward induction and monotonicity, the $(2, \Delta)$-rationalizable beliefs of the firm are (monotonic and) such that

$$\mu(\theta'' \mid e) = \begin{cases} 0 & if \quad e < e^{\tilde{}}(\theta''), \quad e \leq \bar{e}(\theta'), \\ 1 & if \quad e \geq e^{\tilde{}}(\theta''), \quad e > \bar{e}(\theta'). \end{cases}$$

Thus one obtains

$$\Omega(3, \Delta) = \left\{ \vartheta(\cdot) \in \Omega(1, \Delta) : \vartheta(e) = \begin{cases} \theta' & if \quad e < e^{\tilde{}}(\theta''), \quad e \leq \bar{e}(\theta'), \\ \theta'' & if \quad e \geq e^{\tilde{}}(\theta''), \quad e > \bar{e}(\theta') \end{cases} \right\}.$$

At this point the analysis must proceed on a case by case basis. Here I consider only case (a). The other cases are analyzed in the Appendix.

Case (a): $e^{\tilde{}}(\theta'') > \bar{e}(\theta')$. In this case $e^*(\theta)$ ($\theta = \theta', \theta''$) is the unique best reply for type θ to every right-continuous conjecture $\widehat{\vartheta}(\cdot) \in \Omega(3, \Delta)$. Non-right-continuous conjectures in $\Omega(3, \Delta)$ either have no best reply at all or have $e^*(\theta)$ as the unique best reply. Thus the unique $(4, \Delta)$-rationalizable action for type θ is $e^*(\theta)$, $\theta = \theta'$, θ''. The Δ-rationalizable strategies for Player 2 are represented by functions in the set

$$\Omega(\infty, \Delta) = \Omega(5, \Delta) = \left\{ \vartheta(\cdot) \in \Omega(3, \Delta) : \vartheta(e) = \left\{ \begin{array}{lll} \theta' & if & e \leq e^*(\theta') \\ \theta'' & if & e \geq e^*(\theta'') \end{array} \right\} \right. .$$

■

5 Conclusions

In this paper I analyzed and applied a class rationalizability solution procedures for incomplete information games, focusing on signaling games. These procedures are parametrized by given explicit restrictions on players' beliefs about payoff-types and behavior, and also capture the forward induction principle that a player tries to rationalize the past moves of his opponent. The solutions procedures are given a transparent interpretation in terms of interactive beliefs. To illustrate the methodology I analyzed some numerical examples, a model of disclosure and a model of job market signaling. In some cases I obtain the same results as with standard equilibrium analysis complemented by forward induction selection criteria. In other cases (some parameterization of the job market signaling model) I only obtain bounds on behavior, whereas the forward induction equilibrium is unique.

Battigalli and Siniscalchi (2003a) show that the proposed methodology is consistent with Harsanyi's (1967-68) analysis of incomplete information games in its most general form (i.e. without Harsanyi's consistency assumption). Indeed, it can be regarded as a way to characterize specific subsets of Bayesian equilibrium outcomes. Yet, it differs from the typical applications of Harsanyi's approach, which assume "small" type spaces, e.g. by postulating a one-to-one correspondence between payoff-types and Harsanyi-types. I refer to such applications as the *standard* methodology.

I see the following advantages of my approach over the standard methodology. First, unlike Bayesian equilibrium, the iterative solutions proposed here can be computed without specifying an epistemic type space; information partitions on the set of states of nature are sufficient. Second, the assumptions about first-order beliefs (the explicit restrictions) are typically weaker and more intuitive than in the standard theory, and the assumptions about higher order beliefs are more transparent. Third, my approach can be used to test the robustness of the results obtained by standard methods with respect to the equilibrium assumption and the specification of the space of interactive beliefs (the type space *à la* Harsanyi). Fourth, the applications show that looking at the step-by-step procedures which yield rationalizable outcomes may clarify some aspects of strategic thinking that are overlooked by standard equilibrium analysis.

Acknowledgment

I thank Giuseppe Cappelletti and two anonimous referees for useful comments.

Appendix

Rationalizability in a general model of disclosure

The following model is a generalization of the model of Section 3:

- There is a finite ordered set of Sender's types, $\Theta = \{\theta^1, ..., \theta^K\}$, $\theta^1 > \theta^2 > ... > \theta^K$, and a continuum of Receiver's actions $A = [0, +\infty)$.
- The set of messages is $M = 2^{\Theta} \setminus \{\emptyset\}$, where $m \subseteq \Theta$ has the literal interpretation "My type belongs to m."[23]
- The payoff of the Sender is increasing in the action of the Receiver, and if he sends a false message he has to pay a fine:

$$u(\theta, m, a) = \begin{cases} f(a), & if \quad \theta \in m \\ f(a) - P, & if \quad \theta \notin m \end{cases}$$

  where f is a positive strictly increasing bounded function and $P > \sup f$.
- The Receiver's payoff $v : \Theta \times A \to \mathbb{R}$ is independent of the message and is such that there is a well-defined best reply *function* $BR_2 : \Delta(\Theta) \to A$ satisfying the following *weak monotonicity* property: $\forall q', q'' \in \Delta(\Theta)$,

$$[q' \neq q'' \ \wedge \ \max Supp(q') \leq \min Supp(q'')] \implies BR_2(q') < BR_2(q'') \quad (3)$$

  (standard conditions such as supermodularity of function v imply the weak monotonicity property (3)). When $Supp(q) = \{\theta\}$ we write $BR_2(q) = BR_2(\theta)$.

Fix any simple Bayesian game obtained from this model by assuming a strictly positive common prior on the set of states of nature Θ. It can be shown that any sequential (or perfect Bayesian) equilibrium that passes the test of dominated messages[24] must satisfy *full disclosure*, that is, for all messages m,

[23] More generally, it is suffices to assume that $M \subseteq 2^{\Theta}$ is *rich*, i.e. for each θ there is some $m \in M$ such that $\theta = \min m$.

[24] That is, each Bayesian perfect equilibrium with a system of beliefs μ such that $Supp(\mu(\cdot|m)) \subseteq m$ for each $m \in M$.

the Receiver's chooses $a = BR_2(\min m)$ and no type θ sends a message m with $\min m < \theta$. This means that vague messages like "my quality is at least θ" are implicitly understood as revealing that the quality is indeed θ.

Full disclosure is also implied by Δ-*rationalizability* assuming the following weak restriction on conditional beliefs:

- The (first-order) beliefs of the Sender are unrestricted. The restricted set of conditional systems Δ^2 is characterized by a *mild skepticism* condition: the Receiver never rules out the lowest type consistent with a given message, that is,

$$\Delta^2 = \{\mu^2 \in \Delta^*(\Theta, M) : \forall m \in M,\ \mu^2(\min m \mid m) > 0\}.$$

Proposition 5 *Δ-rationalizability implies full disclosure, that is, m is rationalizable for θ only if $\theta = \min m$, and the only rationalizable strategy of the Receiver is $s_2^*(m) = BR_2(\min m)$ for all m.*

Proof

Preliminary Remark 1. I first show by induction that for every $k = 0, 1, 2, ...$ the strategy s_2^* defined by $s_2^*(m) = BR_2(\min m)$ is (k, Δ)-rationalizable [i.e., $s_2^* \in S_2(k, \Delta)$] and each message m is (k, Δ)-rationalizable for payoff-type $\theta = \min m$ [i.e., $\min m \in \Theta(m, k, \Delta)$]. By definition, $s_2^* \in S_2(0, \Delta) = S_2$. Suppose by way of induction that $s_2^* \in S_2(k, \Delta)$ and $\min m \in \Theta(m, k, \Delta)$ for each m. Let μ^* be the conditional probability system defined by

$$\begin{aligned} \forall \theta &\in \Theta,\ \mu^*((\theta, \{\theta\})|\phi) = 1, \\ \forall m &\in M,\ \mu^*(\min m|m) = 1 \end{aligned}$$

[i.e., the Receiver initially believes that the Sender will just reveal the state, hence he believes what the Sender says if $m = \{\theta\}$, and after non-singleton (vague) messages, which falsify the initial belief, his revision rule is to assign probability one to the smallest state consistent with the message]. By definition, μ^* satisfies mild skepticism ($\mu^* \in \Delta^2$). By the inductive hypothesis, $\min m \in \Theta(m, k, \Delta)$ for all m, therefore

$$\forall m \in M,\ \mu^*(\Theta(m, k, \Delta)|m) = \mu^*(\min m|m) = 1.$$

By definition, $s_2^*(m) = BR_2(\min m) = BR_2(\mu^*(\cdot|m))$ for all m. Therefore s_2^* satisfies all the conditions to survive step $k+1$: $s_2^* \in S_2(k+1, \Delta)$. Furthermore, for every payoff-type θ, the set of best responses to s_2^* is $BR_1(\theta, s_2^*) = \{m : \theta = \min m\}$. Since $s_2^* \in S_2(k, \Delta)$ (inductive hypothesis), it follows that $\min m \in \Theta(m, k+1, \Delta)$ for all m. This proves the claim.

Preliminary Remark 2: Rationality of the Sender (only) implies that he always tells the truth:

$$\begin{aligned} \Sigma_1(1, \Delta) &= \{(\theta, m) : \theta \in m\} \\ \text{i.e. } \Theta(m, 1, \Delta) &= m \end{aligned}$$

It follows from these preliminary remarks that, *for each step k, each message m is consistent with* (k, Δ) *rationalizability*, and a $(k+1, \Delta)$-rationalizable strategy s_2 must select, for each m, a best reply to a belief $\mu(\cdot|m)$ such that $Supp(\mu(\cdot|m)) \subseteq \Theta(m, k, \Delta) \subseteq m$. Thus,

$$
\begin{aligned}
S_2(k+1, \Delta) &\subseteq \{s_2 : BR_2(\min \Theta(m, k, \Delta)) \leq s_2(m) \leq BR_2(\max \Theta(m, k, \Delta))\} \\
&\subseteq \{s_2 : BR_2(\min m) \leq s_2(m) \leq BR_2(\max m)\}
\end{aligned}
$$

Main Proof. For any message m with at least k elements, let θ_m^k denote the kth element of m *in decreasing order:* $\theta_m^1 = \max m$, $\theta_m^2 = \max(m \backslash \{\max m\})$, *etc.* I stipulate by *convention that, if m has less than k elements, then* $\theta_m^k = \min m$. I prove that $\forall k \geq 0$,

$$\forall m \in M, \max \Theta(m, 2k+1, \Delta) \leq \theta_m^{k+1}. \tag{4}$$

By the preliminary remarks, this implies that

$$S_2(2k+2, \Delta) \subseteq \left\{ s_2 : BR_2(\min m) \leq s_2(m) \leq BR_2(\theta_m^{k+1}) \right\};$$

hence s_2^* is the only Δ-rationalizable strategy of the Sender. Since $u(\theta, m, a)$ is strictly increasing in its third argument, the complete result easily follows.

The second preliminary remark implies that Eq. (4) holds for $k = 0$. Suppose by way of induction that Eq. (4) holds for a given k. It must be shown that $\max \Theta(m, 2k+3, \Delta) \leq \theta_m^{k+2}$. This is true by convention if m has less than $k+2$ elements. Thus, suppose that m has at least $k+2$ elements and consider a type $\theta' \in m$ such that $\theta' > \theta_m^{k+2}$, that is, $\theta' \geq \theta_m^{k+1}$. I prove that m is not $(2k+3, \Delta)$-rationalizable for θ'.

A message is $(2k+3, \Delta)$-rationalizable for θ' if it is a best response for θ' to a belief μ^1 with $\mu^1 (S_2(2k+2, \Delta)) = 1$. I prove that the (revealing) message $m' = \{\theta'\}$ is a strictly better response for θ' to such a belief μ^1 than message m.

Every strategy $s_2 \in S_2(2k+2, \Delta)$ is a sequential best response to some conditional probability system μ^2 that satisfies mild skepticism [$\mu^2(\min m'|m') > 0$ for every m'] and is such that $\mu^2(\Theta(m', 2k+1, \Delta)|m') = 1$ for every m'. (The Preliminary Remarks shows that these two conditions are mutually consistent, therefore such beliefs do exist.) Thus, mild skepticism, the inductive hypothesis and the choice of θ' yield

$$\mu^2(\cdot|m) \neq \mu^2(\cdot|\{\theta'\}) \;\wedge\; \max \text{Supp}\mu^2(\cdot|m) \leq \theta_m^{k+1} \leq \theta' = \min \text{Supp}\mu^2(\cdot|\{\theta'\}).$$

By the weak monotonicity assumption (3), this implies $BR_2(\mu^2(\cdot|m)) < BR_2(\theta')$. Therefore $\mu^1 (S_2(2k+1, \Delta)) = 1$ yields

$$\max \text{Supp}\pi(\cdot|m; \mu^1) < \min \text{Supp}\pi(\cdot|\{\theta'\}; \mu^1),$$

which in turn implies that, for θ', $m' = \{\theta'\}$ is a strictly better response to μ^1 than m. This proves that $\theta' \notin \Theta(m, 2k+2, \Delta)$, as desired. ■

Example of Bayesian equilibrium (with payoff-irrelevant Harsanyi types) that does not exhibit full disclosure.

There are two Harsanyi-types for the Receiver, $B_2 = \{b_2', b_2''\}$, while Harsanyi-types coincide with payoff-types for the Sender (with the notation of Section 2.6, B_1 is a singleton). The belief functions are given by $\tau_1(\theta^k)(b_2') = 1$ and $\tau_2(b_2')(\theta^k) = 1/K = \tau_2(b_2'')(\theta^k)$ for all k, that is, each type of the Sender is certain that the (Harsanyi) type of the receiver is b_2', and each type of the receiver has a uniform belief on Θ. [Note that these beliefs are consistent with a common prior $p \in \Delta(\Theta \times B_2)$ with strictly positive marginal on Θ, that is, $p(\theta^k, b_2') = 1/K$ for all k.] In equilibrium, each type θ^k chooses the revealing message $m^k = \{\theta^k\}$. The posterior beliefs of type b_2' satisfy $\mu(\min m | m, b_2') = 1$ for all m. Since the Sender is certain that the Receiver's (Harsanyi) type is b_2', the Sender expects him to play strategy $s_2(m) = BR_2(\min m)$. Hence, sending the revealing message is indeed a best response. The posterior beliefs of type b_2'' are uniform on m and b_2'' plays the sequential best response to such system of beliefs. Note that in this example posterior beliefs cannot be derived *via* Bayes rule if m is not a singleton; therefore posterior beliefs do not violate Bayes rule. Of course, each type b_2 chooses a sequential best response to $\mu(\cdot|\cdot, b_2)$. This a perfect Bayesian (or sequential) equilibrium where the strategy of Harsanyi-type b_2'' does not satisfy full disclosure.

Job market signaling: proof of Proposition 4 (b), (c).

Case (b): $e^{\sim}(\theta'') \leq \bar{e}(\theta') \leq e^*(\theta'')$. In this case the set of $(4, \Delta)$-rationalizable messages for the low type θ' is

$$M(4, \Delta, \theta') = \{e^*(\theta')\} \cup [e^{\sim}(\theta''), \bar{e}(\theta')].$$

To see this, note that any education choice $e < e^{\sim}(\theta'')$ reveals Player 1 as type θ' and can be optimal only if $e = e^*(\theta')$. The latter is justified by any conjecture like (2) with $e^* > \bar{e}(\theta')$ (see Section 4). Every choice $e^* \in [e^{\sim}(\theta''), \bar{e}(\theta')]$ is justified by the $(3, \Delta)$-rationalizable conjecture (2). $M(4, \Delta, \theta'') = \{e^*(\theta'')\}$ as in case (a). Thus the only $(5, \Delta)$-rationalizable strategy for Player 2 and $(6, \Delta)$-rationalizable conjecture for both types of Player 1 are given by the function

$$\overline{\vartheta}(e) = \begin{cases} \theta' & if \quad e \leq \bar{e}(\theta') \\ \theta'' & if \quad e > \bar{e}(\theta') \end{cases} .$$

The best reply to $\overline{\vartheta}(\cdot)$ for type θ is $e^*(\theta)$, $\theta = \theta', \theta''$.

Case (c): $\bar{e}(\theta') > e^*(\theta'')$. In this case $M(4, \Delta, \theta') = \{e^*(\theta')\} \cup [e^{\sim}(\theta''), \bar{e}(\theta')]$ as in case (b), but, unlike case (b),

$$M(4, \Delta, \theta'') = [\hat{e}(\theta'')), \bar{e}(\theta')].$$

To see this, note that by choosing $e > \bar{e}(\theta')$ Player 1 is revealed as type θ''. Thus any choice $e^* > \bar{e}(\theta')$ is dominated by $e \in (\bar{e}(\theta'), e^*)$ for θ''. Similarly, any

$e^* < \bar{e}(\theta')$ must be justified by a conjecture $\widehat{\vartheta}(\cdot)$ such that $u(\theta'', e^*, \widehat{\vartheta}(e^*)e^*) \geq u(\theta'', \bar{e}(\theta'), \theta''\bar{e}(\theta'))$, i.e. the point $(e^*, \widehat{\vartheta}(e^*)e^*)$ must lie on or above the θ''-indifference curve through point $(\bar{e}(\theta'), \theta''\bar{e}(\theta'))$ (see Figure 4). Any choice $e^* \in [\hat{e}(\theta''), e^*(\theta''))$ is justified for θ'' by the $(3, \Delta)$-rationalizable conjecture

$$\widehat{\vartheta}(e) = \begin{cases} \theta' & if \quad e < \hat{e}(\theta'') \\ MRS(\theta'', e^*) & if \quad e \in [\hat{e}(\theta''), e^*(\theta'')] \\ \theta'' & if \quad e > e^*(\theta'') \end{cases} .$$

Any choice $e^* \in [e^*(\theta''), \bar{e}(\theta')]$ is justified for θ'' by the $(3, \Delta)$-rationalizable conjecture (2). Choices $e^* < \hat{e}(\theta'')$ cannot be justified by $(3, \Delta)$-rationalizable conjectures: By way of contradiction, let $\widehat{\vartheta}(\cdot)$ be a $(3, \Delta)$-rationalizable conjecture justifying $e^* < \hat{e}(\theta'')$. Since $(e^*, \widehat{\vartheta}(e^*))$ must lie above the θ''-indifference curve through $(\bar{e}(\theta'), \theta''\bar{e}(\theta'))$, $\hat{\theta}(e^*) \geq MRS(\theta'', \hat{e}(\theta''))$ (see Figure 4). $MRS(\theta'', e)$ is strictly increasing in e, thus $MRS(e^*, \theta'') < MRS(\hat{e}(\theta''), \theta'')$. These inequalities jointly violate the first order condition (1).

Therefore Player 2's Δ-rationalizable strategies and Player 1's rationalizable conjectures are the functions $\widehat{\vartheta}(\cdot) \in \Omega(3, \Delta)$ such that $\widehat{\vartheta}(e) = \theta'$ if $e < \hat{e}(\theta'')$, and $\widehat{\vartheta}(e) = \theta''$ if $e > \bar{e}(\theta')$, which implies the thesis. ■

References

[1] Aumann, R.J. (1998). "Reply to Gul". *Econometrica,* Vol. 66, No 4, 929–938.

[2] Battigalli, P. (2001). "A Note on Rationalizability and Reputation with Two Long-Run Players". *RISEC (International Review of Economics and Business)*, Vol. 48, No. 3, 383-393.

[3] Battigalli, P. (2003). "Rationalizability in Infinite, Dynamic Games of Incomplete Information". *Research in Economics*, Vol. 57, No. 1, 1-38.

[4] Battigalli, P. and M. Siniscalchi (1999). "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games". *Journal of Economic Theory*, Vol. 88, No. 1, 188-230.

[5] Battigalli, P. and M. Siniscalchi (2002). "Strong Belief and Forward Induction Reasoning". *Journal of Economic Theory*, Vol. 106, No. 2, 356-391.

[6] Battigalli, P. and M. Siniscalchi (2003a). "Rationalization and Incomplete Information". *Advances in Theoretical Economics*, Vol. 3, No. 1, Article 3. (http://www.bepress.com/bejte/advances/vol3/iss1/art3).

[7] Battigalli, P. and M. Siniscalchi (2003b). "Rationalizable Bidding in First Price Auctions". *Games and Economic Behavior*, Vol. 45, No. 1, 38-72.

[8] Battigalli, P. and J. Watson (1997). "On 'Reputation' Refinements with Heterogeneous Beliefs". *Econometrica*, Vol. 65, No. 2, 369-374.

[9] Ben Porath, E. (1997). "Rationality, Nash Equilibrium and Backwards Induction in Perfect Information Games". *Review of Economic Studies,* Vol. 64, No. 1, 23-46.

[10] Bergemann, D. and S. Morris (2003). "Robust Implementation: the Role of Large Type Spaces". mimeo, Yale University.

[11] Bergemann, D. and S. Morris (2004). "Robust Mechanism Design". Cowles Foundation Discussion Paper No. 1421R.

[12] Bolton, P. and M. Dewatripont (2005). *Contract Theory.* Cambridge MA: MIT Press, forthcoming.

[13] Bonanno, G. and K. Nehring (1999). "How to Make Sense of the Common Prior Assumption Under Incomplete Information". *International Journal of Game Theory,* Vol. 28, No. 3, 409-434.

[14] Brandenburger, A. and E. Dekel (1987). "Rationalizability and Correlated Equilibria". *Econometrica,* Vol. 55, No. 6, 1391-1402.

[15] Cho, I.K. (1994). "Stationarity, Rationalizability and Bargaining". *Review of Economic Studies,* 61, No. 2, 357-374.

[16] Cho, I.K. (2003). "Monotonicity and Rationalizability in Large First Price Auctions", mimeo, University of Illinois.

[17] Cho, I.K. (2004). "Monotonicity and Rationalizability in Large Uniform Price and Double Auctions", mimeo, University of Illinois.

[18] Cho, I.-K. and D. Kreps (1987). "Signaling Games and Stable Equilibria". *Quarterly Journal of Economics*, Vol. 102, No. 2, 179-221.

[19] Fudenberg D. and J. Tirole (1991). *Game Theory.* Cambridge MA: MIT Press.

[20] Grossman, S. (1981). "The Informational Role of Warranties and Private Disclosure About Product Quality". *Journal of Law and Economics,* Vol. 24, No. 2, 461-489.

[21] Grossman, S. and O. Hart (1980). "Disclosure Laws and Takeover Bids". *Journal of Finance*, Vol. 35, No. ?, 323-334.

[22] Gul, F. (1998). "A Comment on Aumann's Bayesian View". *Econometrica*, Vol. 66, No. 2, 923-927.

[23] Dekel, E., D. Fudenberg and S. Morris (2003). "Interim Rationalizability", mimeo.

[24] Dekel, E. and A. Wolinsky (2003). "Rationalizable Outcomes of Large Independent Private-Value First-Price Discrete Auctions". *Games and Economic Behavior*, Vol. 43, No. 2, 175-188.

[25] Ely, J. and M. Peski (2004). "Hierarchies of Beliefs and Interim Rationalizability," mimeo, Northwestern University.

[26] Feinberg, Y. (2000). "Characterizing Common Priors in the Form of Posteriors". *Journal of Economic Theory*, Vol. 91, No. 2, 127-179.

[27] Harsanyi, J. (1967-68). "Games of Incomplete Information Played by Bayesian Players. Parts I, II, III". *Management Science,* Vol. 14, No. 3, 5 , 7, 159-182, 320-334, 486-502.

[28] Harsanyi, J. (1995). "Games with Incomplete Information". *American Economic Review*, Vol. 85, No. 3, 291-303.

[29] Hu, T.-W. (2004). "General Rationalizability and Its Robustness for Strategic Form Games with Incomplete Information," mimeo, National Taiwan University.

[30] Kreps, D. and R.Wilson (1982). "Reputation and Imperfect Information". *Journal of Economic Theory*, Vol. 27, No. 2, 253-279.

[31] Mertens J.F. and S. Zamir (1985). "Formulation of Bayesian Analysis for Games with Incomplete Information". *International Journal of Game Theory*, Vol. 14, No 1, 1-29.

[32] Morris, S. (1995). "The Common Prior Assumption in Economic Theory". *Economics and Philosophy*, Vol. 11, No. ?, 227-253.

[33] Okuno-Fujiwara, M., A. Postlewaite, and K. Suzumura (1990). "Strategic Information Revelation". *Review of Economic Studies,* Vol. 57, No. 1, 25-47.

[34] Pearce, D. (1984). "Rationalizable Strategic Behavior and the Problem of Perfection". *Econometrica,* Vol. 52, No. 4, 1029-1050.

[35] Perry, M. and P. Reny (1999). "A General Solution to King Solomon's Dilemma". *Games and Economic Behavior,* Vol. 26, No. 2, 279-285.

[36] Rabin, M. (1994). "Incorporating Behavioral Assumptions into Game Theory". In J. Friedman (ed.) *Problems of Coordination in Economic Activity.* Dortrecht: Kluwer Academic Publishers.

[37] Sobel, J., L. Stole and I. Zapater (1990). "Fixed-Equilibrium Rationalizability in Signaling Games". *Journal of Economic Theory,* Vol. 52, No. 2, 304-331.

[38] Spence A. M. (1974). *Market Signaling.* Cambridge MA: Harvard University Press.

[39] Watson, J.(1993). "A 'Reputation' Refinement without Equilibrium". *Econometrica,* Vol. 61, No. 1, 199-205.

[40] Watson, J. (1996). “Reputation in Repeated Games with no Discounting”. *Games and Economic Behavior*, Vol. 15, No. 1, 82-109.

[41] Watson, J. (1998). “Alternating-Offer Bargaining with Two-Sided Incomplete Information”. *Review of Economic Studies,* Vol. 65, No. 3, 573-594.