EUROPEAN UNIVERSITY INSTITUTE

DEPARTMENT OF ECONOMICS

EUI Working Paper ECO No. 99/17

Rationalizability in Incomplete Information Games

Pierpaolo Battigalli

BADIA FIESOLANA, SAN DOMENICO (FI)

All rights reserved.
No part of this paper may be reproduced in any form
without permission of the author.

©1999 Pierpaolo Battigalli

This work has been submitted to Games and Economic Behavior (Academic Press) for possible publication. Copyright may be transferred without notice after which this version may no longer be accessible.

Printed in Italy in May 1999
European University Institute
Badia Fiesolana

I-50016 San Domenico (FI)
Italy

Rationalizability in Incomplete Information Games

Pierpaolo Battigalli
Department of Economics
European University Institute
battigal@iue.it

Revised: February 1998
Minor revision: April 1999

Abstract

We argue that the rationalizability approach is particularly appropriate to analyze games with genuine incomplete information. We define two nested iterative solution procedures, which do not rely on the specification of a type space *à la* Harsanyi. *Weak rationalizability* is characterized by common certainty of rationality at the beginning of the game. *Strong rationalizability* incorporates a notion of forward induction. The solutions may take as given some extraneous restrictions on players' conditional beliefs. In dynamic games, strong rationalizability is a refinement of weak rationalizability. Existence, regularity properties, equivalence with the set of perfect Bayesian equilibrium outcomes and the set of iteratively interim undominated strategies are proved under standard assumptions. The approach is illustrated by some applications to economic models including reputation, disclosure and signaling.

1. Introduction and Overview[1]

In a n-person game of incomplete information some of the crucial elements governing strategic interaction – such as individual feasibility constraints, how actions are mapped into consequences and individual preferences over consequences – are represented by a vector of parameters θ which is (partially) unknown to some players. For the sake of simplicity, let us assume that θ determines the shape of each player's payoff function and that it can be partitioned into subvectors $\theta_1, ..., \theta_n$ whereby each player $i = 1, ..., n$ knows θ_i. We call θ the *state of Nature* and θ_i the private information or *payoff-type* of player i. The form of the parametric payoff functions $u_i(\cdot, \theta)$ – or, more generally, the form of the mapping Γ associating each conceivable state of Nature θ to the "true" (but unknown) game $G(\theta)$ – is assumed to be common knowledge. In this paper we take this mapping $\theta \mapsto G(\theta)$ as the fundamental description of a strategic situation with incomplete information and we put forward and analyze solution concepts associating to any such mapping a set of possible outcomes. Our approach is related to, but different from Harsanyi's (1967-68) seminal paper on incomplete information games. Harsanyi's Bayesian model is now so entrenched in the literature that only a handful of "pure" game theorists still pay attention to its subtleties. In order to motivate and better understand our contribution it is useful to go through Harsanyi's model in some detail.[2]

1.1. Harsanyi's Bayesian Model

As Harsanyi noticed, one way to provide a Bayesian analysis of incomplete information games is to endow each player with a *hierarchy of beliefs*, that is, (i) a subjective probability measure on the set of conceivable states of Nature, or *first-order belief*, (ii) a subjective probability measure on the set of conceivable first-order beliefs of his opponents, or *second order beliefs*, and so on. In principle, a complete description of every relevant attribute of a player should include, not

[1] This paper is a revision of Battigalli (1995). Helpful comments from Patrick Bolton, Giacomo Bonanno, Tilman Börgers, Françoise Forges, Faruk Gul, Marciano Siniscalchi, Juuso Välimäki, Joel Watson and seminar participants at the University of Valencia, Northwestern University, Caltech, McGill University, SITE (Stanford University), Université de Cergy Pontoise, University of North Carolina and European University Institute are gratefully aknowledged.

[2] For thourough discussion of the Bayesian model see Harsanyi (1995), Gul (1996b), Dekel and Gul (1997).

only his payoff-type, but also his *epistemic type*, that is, an infinite hierarchy of beliefs. Furthermore, (infinitely) many hierarchies of beliefs could be attached to a given payoff-type. This hierarchies-of-beliefs approach is mathematically feasible and theoretically interesting (see e.g. Mertens and Zamir (1985) and the related references mentioned below), but it does not seem to provide a tractable framework to analyze incomplete information games.

Harsanyi's (1967-68) contribution was twofold. On the one hand, he put forward a general notion of "type space" which provides an *implicit*, but relatively parsimonious description of infinite hierarchies of beliefs. On the other hand, he showed how to analyze incomplete information games with the standard tools of game theory. A *type space* can be defined as follows. For each player i and each payoff-type $\theta_i \in \Theta_i$ (Θ_i is the set of i's conceivable payoff-types) we add a parameter e_i corresponding to a purely epistemic component of player i's attributes. In general, different values of e_i can be attached to a given payoff-type θ_i. This way we obtain a set $T_i \subset \Theta_i \times E_i$ of possible attributes, or *Harsanyi-types*, of player i. A Harsanyi-type encodes the payoff-type *and* the epistemic type of a player. In fact, the beliefs of any given player i about his opponents' payoff-types as well as their own beliefs are determined by a function $p_i : T_i \to \Delta(T_{-i})$, where $T_{-i} = \prod_{j \neq i} T_j$. It is assumed that the vector of functions $(p_1, ..., p_n)$ is common knowledge. Therefore every $t_i \in T_i$ corresponds to an infinite hierarchy of beliefs: the first order belief $p_i^1(t_i)$ is simply the marginal of $p_i(t_i)$ on Θ_{-i}; the $(k+1)$-order belief implicit in t_i is derived from $p_i(t_i)$ and knowledge of the $n-1$ functions $p_j^k(\cdot)$, $j \neq i$, mapping the opponents' Harsanyi-types into k-order beliefs. When we add a type space on top of the map $\theta \mapsto G(\theta)$ we obtain a *Bayesian game*. A *Bayesian equilibrium* is a vector of behavioral rules $b_i : T_i \to S_i$ ($i = 1, ..., n$, S_i is the strategy set for player i) such that for each player i and each Harsanyi-type $t_i = (\theta_i, e_i)$, strategy $s_i = b_i(\theta_i, e_i)$ maximizes i's expected payoff given the payoff-type θ_i, the subjective belief $p_i(\theta_i, e_i)$ and the $(n-1)$-tuple of functions b_{-i}. Note that, for any fixed vector of behavioral rules, a vector of Harsanyi-types $(t_1, ..., t_n)$ provides an implicit, but complete description of every relevant aspect of the world: the state of Nature, each player's subjective beliefs about the state of Nature and his opponents' behavior and each player's subjective beliefs about his opponents beliefs.

Within this framework, the players' situation in a game of incomplete information is formally similar to the interim stage of a game with complete, but imperfect and asymmetric information whereby t_i represents the private information of player i about the realization of an initial chance move, such as the cards player i has been dealt in a game of Poker. Harsanyi pushed the analogy even further by assuming that all the subjective beliefs $p_i(t_i)$ $(i = 1, ..., n,\ t_i \in T_i)$ can be derived from a *common prior* $P \in \Delta(\prod_{j=1}^{n} T_j)$ so that $p_i(t_i) = P(\cdot|t_i)$. In this case a Bayesian equilibrium simply corresponds to a Nash equilibrium of a companion game with imperfect information about a fictitiuos chance move selecting the vector of attributes according to probability measure P. This is the so called "random vector model" of the Bayesian game. From the point of view of equilibrium analysis we can equivalently associate to the given Bayesian game a companion game with complete information whereby for each player/role $i = 1, ..., n$ there is a population of potential players characterized by the different attributes $t_i \in T_i$. An actual player is drawn at random from each population i to play the game. The joint distribution of attributes in the n populations is given by the common prior P. This is the "prior lottery model" of the Bayesian game.

1.2. Drawbacks of Standard Bayes-Nash Equilibrium Analysis

Harsanyi's analysis of incomplete information games has offered invaluable insights to economic theorists and applied economists, but its success should not make us overlook some potential drawbacks of this approach and of its standard applications to economic models. These potential drawbacks are all related to the following facts: (a) a Bayesian game provides only an *implicit* and (in general) non exhaustive – or *non-universal* – representation of the conceivable epistemic types; (b) representing a Bayesian game with the "random vector model" or the "prior lottery model" blurs the fundamental distinction between games with *genuine* incomplete information and games with imperfect, asymmetric information: in the former there is *no ex ante stage* at which the players analyze the situation before receiving some piece of information selected at random.

(a) Non-transparent assumptions about beliefs. We mentioned that for every Harsanyi-type in a Bayesian game we can derive a corresponding infinite hierarchy of beliefs. This derivation makes sense if it is assumed that the Bayesian game is common knowledge.[3] Mertens and Zamir (1985) shows that this informal assumption is without loss of generality because (i) the space of n-tuples of (consistent) infinite hierarchies of beliefs is a well- defined type space in the sense of Harsanyi and (ii) every type space is essentially a belief-closed subspace

[3]If we regard the Bayesian game itself as a subjective model of a given player, then we have to assume that this player is certain that everybody shares the same model (cf. Harsanyi (1967-68)).

of the space of infinite hierarchies of beliefs, which is therefore a *universal* type space.[4] This means that the *class* of *all* Bayesian models is sufficiently rich, but whenever we consider a particular (non-universal) model, or a subclass of models, we rule out some epistemic types. This corresponds to making assumptions about players' interactive beliefs, which are often questionable and – due to the implicit representation of epistemic types – non-transparent.

For example, "agreement" and "no-trade" results hold for Bayesian models satisfying the common prior assumption, but the meaning of this assumption as a restriction on players' hierachies of beliefs is not obvious.[5] For the sake of tractability, applied economists often restrict their attention to an even smaller class of Bayesian models by assuming that there is a one-to-one correspondence between payoff-types and Harsanyi-types. These strong and yet only implicit assumptions about players' hierarchies of beliefs may affect the set of equilibrium outcomes in an important way. But we have a hard time reducing these assumptions to more primitive and transparent axioms.

(b1) No ex ante stage and plausibility of assumptions about beliefs. The formal similarity between Bayesian games and games with asymmetric information may be misleading. We are quite ready to accept that in the "random vector model" players assign the same prior probabilities to chance moves.[6] Similarly, assuming a commom probability measure over players' attributes is meaningful and plausible, if not compelling, in the "prior lottery model." For example, it can be justified by assuming that the statistical distribution of characteritics in the population of potential players is commonly known. But in games with *genuine* incomplete information there is no *ex ante* stage and prior probabilities are only a (convenient, but unnecessary) notational device to specify players' infinite hierarchies of beliefs. Thus, the common prior assumption and the conflation of payoff-types and Harsanyi-types are much harder to accept.

(b2) No ex ante stage and learning. The lack of an ex ante stage also makes the equilibrium concept more problematic. A Nash equilibrium of a given "objective" game G may be interpreted as a stationary state of a learning process

[4] See also Brandenburger and Dekel (1993) and references therein. Battigalli and Siniscalchi (1998) provides analogous results for infinite hierachies of systems of *conditional* beliefs in *dynamic* games of incomplete information.

[5] For more on this see Gul (1996b) and Dekel and Gul (1997). Bonanno and Nehring (1996) "makes sense" of the common prior assumption in incomplete information games, characterizing it as a very strong "agreement" property.

[6] For a discussion of the common prior assumption in situations with asymmetric, but complete information see Morris (1995).

as the players repeatedly play G. Furthermore, it is possible to provide sufficient conditions such that learning eventually induces a Nash equilibrium outcome.[7] We cannot provide a similar justification for equilibria of Bayesian games. Let θ be the actual state of Nature in a game of incomplete information Γ and recall that $G(\theta)$ denotes the "true objective game" corresponding to θ. Let us assume that the players interact repeatedly. By the very nature of the problem we are considering, we have to assume that the state of Nature θ is fixed once and for all at the beginning of time rather than being drawn at random according to some i.i.d. process. By repeatedly playing $G(\theta)$ the players can learn (at most) to play a Nash equilibrium of $G(\theta)$, not a Bayesian equilibrium of (some Bayesian game based on) Γ.[8]

1.3. Rationalizable Outcomes of Incomplete Information Games

To summarize what we said so far, in order to analyze an economic model with incomplete information Γ using Harsanyi's approach we have to specify a type space based on Γ and then look for the Bayesian equilibria of the resulting Bayesian game. The specification of the type space is hardly related to the fundamentals of the economic problem and yet may crucially affect the set of equilibrium outcomes. This raises several related theoretical questions. Can we analyze incomplete information games without specifying a type space? Can we provide an independent justification for the Bayesian equilibrium concept? Which results of the Bayesian analysis are independent of the exact specification of the type space? Is it possible to provide a relatively simple characterization of the set of all Bayesian equilibrium outcomes?

The answer to these questions can be found in the literature on rationalizability. Let us consider complete information games first, i.e. games with only one conceivable state of Nature. The set of rationalizable strategies in a static game with complete information is obtained by an iterative deletion procedure which (in two-person games) coincides with iterated strict dominance (Pearce (1984)). Rationalizability exactly characterizes the strategies consistent with common certainty of rationality (Tan and Werlang (1988)) and also the set of subjective correlated equilibrium outcomes (Brandenburger and Dekel (1987)). Note that,

[7] In general, convergence is not guaranteed and, even if the play converges, the limit outcome is a self-confirming (or conjectural) equilibrium, which need not be equivalent to a Nash equilibrium. See Fudenberg and Levine (1998) and references therein.

[8] More generally, their pattern of behavior may converge to what Battigalli and Guaitoli (1996) call "a conjectural equilibrium at θ."

according to the terminology used so far, a subjective correlated equilibrium is simply a Bayesian equilibrium of a model with a unique state of Nature and hence with payoff-irrelevant Harsanyi-types.

This paper puts forward and analyzes some notions of rationalizability for games with genuine *incomplete* information, but the proposed solutions are also relevant for games with asymmetric information where the statistical distribution of attributes in the population of potential players is not known. We focus mainly on the analysis of *dynamic* games where players can signal their types and strategic intent. But the basic idea is more easily understood if we consider static games first. Consider the following procedure: (Basis Step) For every player i, payoff type θ_i and strategy s_i in Γ, we check whether s_i can be justified as a feasible best response for θ_i to some probabilistic beliefs about the opponents' payoff-types and behavior. If the pair (θ_i, s_i) does not pass this test it is "removed." (Inductive Step) For every i, θ_i and s_i we check whether s_i is a feasible best response for θ_i to some probabilistic beliefs about the opponents assigning probability zero to the (vectors of) pairs (θ_{-i}, s_{-i}) removed so far. Note that (epistemic) type spaces are not mentioned. The procedure depends only on the "fundamentals" of the economic model. Not surprisingly, *this solution is equivalent to* an *iterative "interim" dominance* procedure. Furthermore, it turns out that *it exactly characterizes the set of all possible equilibrium outcomes* of the Bayesian games based on Γ. It is also easy to provide an epistemic charaterization *à la* Tan and Werlang (1988) of the rationalizable outcomes as those consistent with common certainty of rationality (see Battigalli and Siniscalchi (1998) in the context of dynamic games).

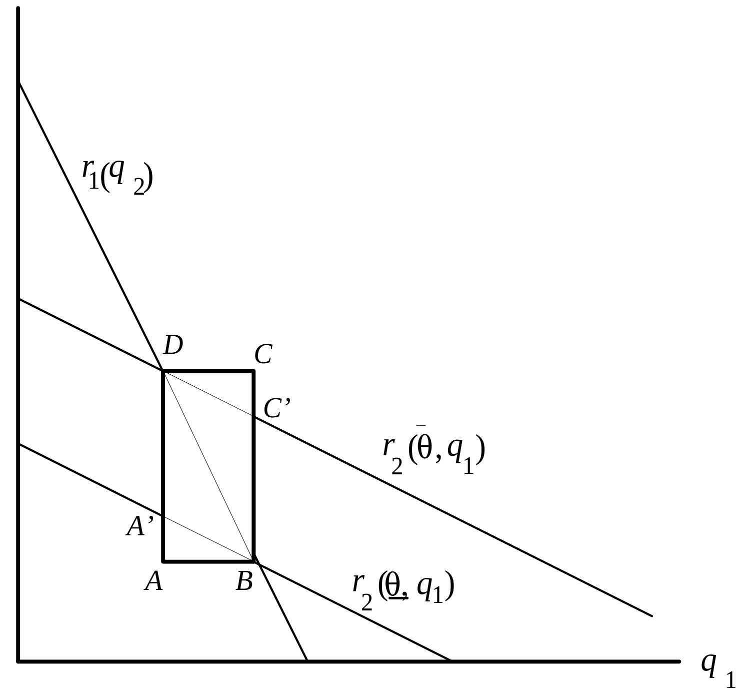

Figure 1

Let us see how the solution procedure works in a textbook example. Consider a Cournot duopoly with one-sided incomplete information. The inverse demand schedule $P(Q)$ is linear and firms have constant marginal cost. The marginal cost firm 1, c_1, is common knowledge, but c_2, the marginal cost of firm 2, is unknown to firm 1. The range of conceivable values of c_2 is a closed interval strictly contained in $[0, P(0)]$ and containing c_1 in its interior. Both firms are expected profit maximizers. Figure 1 shows the reaction functions for firm 1 ($r_1(q_2)$), for the most efficient type of firm 2 ($r_2(\overline{\theta}, q_1)$), and for the least efficient type of firm 2 ($r_2(\underline{\theta}, q_1)$). In this model, there is no loss of generality in considering only best responses to deterministic beliefs.[9] The first step of the rationalizability procedure

[9]This is true in all static games where, for each player i, (1) the utility function $u_i(\theta, s_i, s_{-i})$ is continuous and strictly quasi-concave in its second argument, (2) the strategy space S_i is a

eliminates, for each type of each firm, all the outputs above the monopolistic choice (e.g. $r_2(\theta_2, 0)$ for type θ_2 of firm 2), which is the best response to the most optimisitc conjecture about the opponent (assuming that the opponent might also be irrational). In fact, all the eliminated outputs are strictly dominated for type θ_i by the monopolistic choice of θ_i, while the remaining outputs are best responses to some conjecture. In the second step of the procedure we eliminate, for each type of each firm, all the outputs below the best reponse to the most pessimistic conjecture consistent with rationality of the opponent. For example, for firm 1 we eliminate all the outputs below $r_1(r_2(\overline{\theta}, 0))$. In the third step we eliminate, e.g. for firm 1, all the outputs above $r_1(r_2(\underline{\theta}, r_1(0))$, which is the best response to the most optimistic conjecture consistent with the opponent being rational and certain that everybody is rational. In the limit we obtain a set of rationalizable outcomes represented by the rectangle ABCD in Figure 1.

Let us compare rationalizable outcomes and standard Bayesian equilibrium outcomes. The standard Bayesian model specifies the belief of player 1 about θ_2, say $\pi \in \Delta(\Theta_2)$. It is assumed that it is common certainty that π indeed represents the belief of player 1. The Bayesian equilibrium strategy for player 1 is given by the intersection between the graph of $r_1(\cdot)$ and the graph of $r_2(E(\widetilde{\theta}; \pi), \cdot)$, where $E(\widetilde{\theta}; \pi)$ denotes the expected value of θ_2 given π. The set of Bayesian equilibrium outcomes for all possible $\pi \in \Delta(\Theta_2)$ is the parallelogram A'BC'D in Figure 1. But if we consider *all* the possible specifications of a type space *à la* Harsanyi, the set of Bayesian equilibrium outcomes coincides with the set of rationalizable outcomes.[10]

The procedure described above is relevant if we do not want to rule out any conceivable epistemic type. However, it may be plausible to assume that players' beliefs satisfy some qualitative restrictions. The iterative solution concept can be easily modified to accomodate restrictions on first order beliefs (informally) assumed to be commonly known. In the general definition of the solution procedure

closed interval of the real line, and (3) the set of conceivable payoff types Θ_i is compact and connected.

[10]The proof of Proposition 3.10 shows how to construct a type space such that, in the resulting Bayesian game, each rationalizable outcome is a Bayesian equilibrium outcome. Here we provide a simpler example. Assume that there are two epistemic types for each payoff-type. Thus $T_1 = \{t_1^1, t_1^2\}$ and $T_2 = \Theta_2 \times \{e_2^1, e_2^2\}$. Assume $p_1(t_1^1)$ is degenerate on $(\overline{\theta}, e_2^1)$, $p_1(t_1^2)$ is degenerate on $(\underline{\theta}, e_2^2)$, and $p_2(\theta_2, e_2^j)$ assignes probability one to t_1^j for all θ_2 and j. (These belief functions are consistent with a "correlated" common prior.) In the Bayesian equilibrium where type t_1^1 (t_1^2) chooses the lowest (highest) rationalizable output for firm 1, all the points in the vertical segments AD and BC are equilibrium outcomes.

these *extraneous restrictions* on players' beliefs are parametrically given.

The analysis of incomplete information games is particularly interesting when they have a *dynamic* structure, because in this case a player can make inferences about the types and /or strategic intents of his opponents by observing their behavior in previous stages of the game. 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 we consider two nested solution concepts for (possibly infinite) multi-stage games with incomplete information, called *weak rationalizability* and *strong rationalizability*. Rigourous axiomatizations of these solution concepts involve the definition of extensive form epistemic models and are given elsewhere (Ben Porath (1997), Battigalli and Siniscalchi (1998, 1999a,b)). Intuitively, weak rationalizability simply assumes that players choose sequential best responses to their systems of conditional beliefs, updating *via* Bayes rule whenever possible, and this is common certainty at the beginning of the game. On top of this, strong rationalizability also assumes that each player keeps believing that his opponents are rational even when they behave in an unexpected way, provided that their behavior can be somehow "rationalized" (a more detailed account is provided in Section 3). Thus, unlike weak rationalizability, strong rationalizability incorporates a forward induction criterion.

We apply the rationalizability approach to some economic models. In some cases we are able to obtain the same qualitative results as in the more standard equilibrium analysis based on the common prior assumption and/or the conflation of payoff-types and Harsanyi-types. In other cases, such as the example above, we obtain weaker results. In general, rationalizability emphasizes and clarifies some aspects of strategic reasoning that are either ignored or made obscure by standard equilibrium analysis.

1.4. Related Literature

The solution concepts developed in this paper extend notions of rationalizability for extensive form games with complete information put forward and analyzed by Pearce (1984), Battigalli (1996, 1997) and Ben Porath (1997). The idea of using some notion of rationalizability to analyze games of incomplete information is a quite natural development of Bernheim (1984) and Pearce's (1984) work on complete information games and it appears in some papers in the literature (although several papers take for granted the common prior assumption and/or identify payoff-types and Harsanyi-types). Battigalli and Guaitoli (1996) analyzes the extensive form rationalizable paths of a simple macroeconomic game with incomplete information and no comon prior. This paper also puts forward a notion of conjectural (or self-confirming) equilibrium at a given state of Nature of an incomplete information game. Cho (1994) and Watson (1997) use a notion of subform rationalizability to analyze dynamic bargaining with incomplete information. Watson (1993, 1996) obtains reputation and/or cooperation results for perturbed repeated games under mild restrictions on players' beliefs. Battigalli and Watson (1997) qualify and extend Watson's (1993) analysis of reputation. Perry and Reny (1994) and Khaneman, Perry and Reny (1995) consider some specific social choice problems with incomplete information and propose extensive form mechanisms to implement desirable outcomes in iteratively undominated strategies. Rabin (1994) proposes to combine rationalizability and extraneous restrictions on players' beliefs to introduce behavioral assumptions in game theoretic analysis. Siniscalchi (1997a) applies the approach of the present paper to the analysis of dynamic auctions. A different approach to incomplete information games is proposed in Sákovics (1997). He considers Bayesian models with *finite* hierarchies of beliefs and puts forward a novel solution concept, called "mirage equilibrium."

1.5. Structure of the Paper

The rest of the paper is organized as follows. Section 2 contains the game theoretic set up. Weak and strong rationalizability are defined and analyzed in Section 3 focusing on two-person games with observable actions. Existence and regularity properties are proved for a class of "simple," but possibly infinite games. We define a notion of "weak Bayesian perfect equilibrium" and we show that weak rationalizability characterizes the set of all weak Bayesian perfect equilibrium outcomes. Finally, we extend to the present framework some known results relating rationalizability and iterative dominance. Section 4 shows how the analysis can be extended to n-person games with imperfectly observable actions. Section 5 applies these solution concepts to models of reputation, disclosure and costly signaling. The Appendix contains some details about dynamic games of incomplete information and most of the proofs.

2. Game Theoretic Framework

2.1. Games of Incomplete Information with Observable Actions

A *game of incomplete information with observable actions* is a structure

$$\Gamma = \langle N, (\Theta_i)_{i\in N}, (A_i)_{i\in N}, \mathcal{H}^*(\cdot), (u_i)_{i\in N}\rangle$$

given by the following elements:[11]

- N is a non empty, finite set of *players*.
- For each $i \in N$, $\Theta_i \subset \mathbf{R}^{m_i}$ is a non empty set of possible *types* for player i and $A_i \subset \mathbf{R}^{n_i}$ is a non empty set of possible *actions* for player i ($\mathbf{R}^k$ is the k-dimensional Euclidean space).
- Let $\Theta = \prod_{i\in N}\Theta_i$ and $A = \prod_{i\in N} A_i$. Then

$$A^* = \{\phi\} \cup \left(\bigcup_{t=1}^{t=\infty} A^t\right),$$

that is, A^* is the set of finite and countably infinite sequences of action profiles, including the *empty sequence* ϕ, and

$$\mathcal{H}^*(\cdot) : \Theta \to 2^{A^*}$$

(2^{A^*} is the power set of A^*) is a non empty-valued correspondence assigning to each profile of types θ the set $\mathcal{H}^*(\theta)$ of *feasible histories* given θ. For every history $h \in \mathcal{H}^*(\theta)$ one can derive the set $A(\theta, h) = \prod_{i\in N} A_i(\theta_i, h)$ of *feasible action* profiles. A history $h \in \mathcal{H}^*(\theta)$ is *terminal* at θ if $A(h, \theta) = \emptyset$ (every infinite feasible history is terminal). We let

$$\mathcal{H}(\theta) = \{h \in A^* : A(\theta, h) \neq \emptyset\},$$

$$\mathcal{H}(\theta_i) = \{h \in A^* : \exists\theta_{-i} \in \Theta_{-i}, A((\theta_i, \theta_{-i}), h) \neq \emptyset\},$$

$$\mathcal{H} = \bigcup_{\theta\in\Theta} \mathcal{H}(\theta)$$

respectively denote the set of feasible non terminal histories at θ, or for θ_i, and the set of *a priori feasible* non terminal histories.

[11] The following model generalizes Fudenberg and Tirole (1991, pp 331-332) and Osborne and Rubinstein (1994, pp 231-232). The Appendix provides further details.

- Define the set $\mathcal{Z}$ of *outcomes* as follows:[12]

$$\mathcal{Z} = \{(\theta, h) : h \in \mathcal{H}^*(\theta), A(\theta, h) = \emptyset\}.$$

For all $i \in N$,

$$u_i : \mathcal{Z} \to \mathbf{R}$$

is the *payoff function* for player i ($\mathbf{R}$ denotes the set of real numbers).

Parameter θ_i represents player i's private information about feasibility constraints and payoffs. For brevity, we call θ_i the *"payoff-type"* of player i. It is assumed that Γ is common knowledge. The array $\theta = (\theta_i)_{i \in N}$ is interpreted as a *state of Nature*; it completely specifies the unknown parameters of the game and the players' interactive knowledge about them. Player i at (θ, h) knows (θ_i, h) and whatever can be inferred from history h given that Γ (hence $\mathcal{H}^*(\cdot)$) is common knowledge. Chance moves and residual uncertainty about the environment can be modeled by having a pseudo-player $c \in N$ with a constant payoff function. The "type" θ_c of this pseudo-player represents the residual uncertainty about the state of Nature which would remain after pooling the private information of the real players. Players' common or heterogeneous beliefs about chance moves can be modeled as extraneous restrictions on beliefs (see below).

Game Γ is *static* if for all $\theta \in \Theta$ and $a \in A(\theta, \phi)$, (a) is a terminal history at θ. Game Γ has *private values* if, for all $i \in N$, $u_i(\theta_i, \theta_{-i}, \cdot)$ is independent of θ_{-i}. A player of type θ_i is *active* at history h if $A_i(\theta_i, h)$ contains at least two elements. Γ has *no simultanoeus moves* if for every state of Nature θ and every history $h \in \mathcal{H}(\theta)$ there is only one active player. In this case, Γ can be represented by an extensive form with decision nodes (θ, h), $\theta \in \Theta$, $h \in \mathcal{H}(\theta)$ (pairs (θ, ϕ) are the initial nodes of the arborescence) and information sets for player i of the following form:

$$I(\theta_i, h) = \{(\theta_i, \theta_{-i}, h) : h \in \mathcal{H}(\theta_i, \theta_{-i})\},$$

where θ_i is active at h. Game Γ has (incomplete but) *perfect information* if it has no simultaneous moves and $\mathcal{H}^*(\theta)$ is independent of θ.[13]

Note that the basic model Γ *does not specify players' beliefs about the state of Nature* θ. This is what makes Γ different from the standard notion of a Bayesian

[12] The feasibility correspondence is such that, if $((\theta_i, \theta_{-i}), h) \in \mathcal{Z}$, then $((\theta_i, \theta'_{-i}), h) \in \mathcal{Z}$ for all θ'_{-i}.).

[13] In this case, Γ can also be represented by a game tree (with decision nodes $h \in \mathcal{H}$) featuring perfect information and payoff functions $v_i : \Theta \times Z \to \mathbf{R}$, where Z is the set of terminal nodes.

game. As mentioned in the Introduction, if we want to provide a general (albeit implicit) representation of players' beliefs about the state of Nature and of their hierachies of beliefs, we have to embed each set Θ_i in a possibly richer set T_i of "Harsanyi-types" and specify belief functions $p_i : T_i \to \Delta(T_{-i})$. For more on this see section 3.4.

Turning to the topological properties of Γ, we endow A^* and $\mathcal{Z}$ with the standard "discounting" metrics (see the Appendix) and throughout the paper we rely on the following assumption:

Assumption 0. *A and Θ are closed, $\mathcal{H}^*(\cdot)$ is a continuous correspondence and, for all $i \in N$, u_i is a continuous function.*

2.2. Strategic Forms

A *feasible strategy for type* θ_i is a function $s_i : \mathcal{H} \to A_i$ such that $s_i(h) \in A_i(\theta_i, h)$ for all $h \in \mathcal{H}(\theta_i)$.[14] The set of feasible strategies for type θ_i is denoted $S_i(\theta_i)$ and

$$S_i = \bigcup_{\theta \in \Theta} S_i(\theta_i)$$

denotes the set of *a priori feasible strategies.* (By definition of $\mathcal{H}$, for all $h \in \mathcal{H}$, $A_i(\theta_i, h)$ is nonempty. Therefore $S_i(\theta_i)$ is also nonempty.)

The basic elements of our analysis are feasible type-strategy pairs: (θ_i, s_i) is a *feasible pair* if $s_i \in S_i(\theta_i)$. A generic feasible pair for player i is denoted σ_i and the set of such feasible pairs for player i is the graph of the correspondence $S_i(\cdot) : \Theta_i \to 2^{S_i}$, i.e.

$$\Sigma_i := \{(\theta_i, s_i) \in \Theta_i \times S_i : s_i \in S_i(\theta_i)\}$$

The sets of profiles of feasible pair for all players and for the opponents of a player i are, respectively, $\Sigma = \prod_{j \in N} \Sigma_j$ and $\Sigma_{-i} = \prod_{j \neq i} \Sigma_j$. Each profile $\sigma = [(\theta_i, s_i)]_{i \in N}$ induces a terminal history $\zeta(\sigma) \in \mathcal{H}(\theta)$ and hence an outcome $\zeta^*(\sigma) = (\theta, \zeta(\theta)) \in \mathcal{Z}$. Therefore, for each player i, we can derive the following strategic form payoff function :

$$U_i = u_i \circ \zeta^* : \Sigma \to \mathbf{R}.$$

Furthermore, for each a priori feasible history $h \in \mathcal{H}$ we can define the set of profiles of feasible pairs consistent with h:

$$\Sigma(h) = \{\sigma \in \Sigma : h \text{ is a prefix of } \zeta(\sigma)\}.$$

[14] We let the domain of s_i be $\mathcal{H}$ (instead of $\mathcal{H}(\theta_i)$) only for notational simplicity.

Clearly, $\Sigma(\phi) = \Sigma$. We let $\Sigma_i(h)$ denote the projection of $\Sigma(h)$ on Σ_i, that is, the set of (θ_i, s_i) such that strategy s_i is feasible for type θ_i and does not prevent history h. It can be easily checked that, for all $h \in \mathcal{H}$,

$$\Sigma(h) = \prod_{i \in N} \Sigma_i(h) \neq \emptyset.$$

The information of player i about his opponents at history h is represented in strategic form by $\Sigma_{-i}(h)$, the projection of $\Sigma(h)$ on Σ_{-i}.

We endow the sets Σ_i ($i \in N$) with the standard metrics derived from the metric on $\mathcal{Z}$ (see the Appendix).

Lemma 2.1. *For all $h \in \mathcal{H}$, $\Sigma_i(h)$ is closed.*

2.3. Conditional Beliefs

Players' beliefs in dynamic games can be represented as systems of conditional probabilities. Let Σ be a metric space with Borel sigma-algebra $\mathcal{S}$. Fix a nonempty collection of subsets $\mathcal{B} \subset \mathcal{S} \backslash \{\emptyset\}$, to be interpreted as "relevant hypotheses."

Definition 2.2. *(cf. Rényi (1956) and Myerson (1986)) A* conditional probability system *(or CPS) on $(\Sigma, \mathcal{S}, \mathcal{B})$ is a mapping*

$$\mu(\cdot|\cdot) : \mathcal{S} \times \mathcal{B} \to [0, 1]$$

satisfying the following axioms:

Axiom 1. *For all $B \in \mathcal{B}$, $\mu(B|B) = 1$.*

Axiom 2. *For all $B \in \mathcal{B}$, $\mu(\cdot|B)$ is a probability measure on $(\Sigma, \mathcal{S})$.*

Axiom 3. *For all $A \in \mathcal{A}$, $B, C \in \mathcal{B}$, $A \subset B \subset C \Rightarrow \mu(A|B)\mu(B|C) = \mu(A|C)$.*

The set of probability measures on $(\Sigma, \mathcal{S})$ is denoted by $\Delta(\Sigma)$; the set of conditional probability systems on $(\Sigma, \mathcal{S}, \mathcal{B})$ can be regarded as a subset of $[\Delta(\Sigma)]^{\mathcal{B}}$ (the set of mappings from $\mathcal{B}$ to $\Delta(\Sigma)$) and it is denoted by $\Delta^{\mathcal{B}}(\Sigma)$. The topology on Σ and $\mathcal{S}$ (the smallest sigma-algebra containing this topology) are understood and need not be explicit in our notation. It is also understood that $\Delta(\Sigma)$ is endowed with the topology of weak convergence of measures and $[\Delta(\Sigma)]^{\mathcal{B}}$ is endowed with the product topology.

A relatively simple way to represent the beliefs of a player i in a dynamic game with incomplete information is to consider the set $\Delta^{\mathcal{B}_i}(\Sigma_{-i})$ of conditional probability systems on $(\Sigma_{-i}, \mathcal{S}_{-i}, \mathcal{B}_i)$, where Σ_{-i} is the set of type-strategy profiles for his opponents, $\mathcal{S}_{-i}$ is the Borel sigma algebra of Σ_{-i}, and

$$\mathcal{B}_i = \{B \subset \Sigma_{-i} : \exists h \in \mathcal{H}, B = \Sigma_{-i}(h)\}$$

is the family of "strategic form information sets" for player i.[15] By Lemma 2.1, $\mathcal{B}_i$ is a collection of closed subsets and thus $\Delta^{\mathcal{B}_i}(\Sigma_{-i})$ is indeed a well-defined space of conditional probability systems.

An element of $\Delta^{\mathcal{B}_i}(\Sigma_{-i})$ only describes the *first order* conditional beliefs of player i. Only such beliefs are explicit in the game theoretic analysis of this paper, but the motivations and epistemic foundations of the solution concepts to be proposed below at least implicitly consider higher order beliefs. Battigalli and Siniscalchi (1998) shows how to construct *infinite hierarchies of conditional beliefs* which represent the *epistemic type* of a player, that is, the beliefs that this player would have, conditional on each history, about the state of Nature , his opponents' strategies and his opponents' epistemic types. This construction allows one to define formal notions of *conditional common certainy* and *strong belief* which are informally used in this paper to motivate and clarify the proposed solution concepts. Formal epistemic characterizations of solution concepts in terms of infinite hierachies of conditional beliefs can be found in Battigalli and Siniscalchi (1998, 1999a,b).

[15] Two points are worth discussing. (1) In a situation of incomplete information, when player i forms his beliefs he already knows his private information θ_i. Therefore it would be more germane to the analysis of incomplete information games to consider the set $\Delta^{\mathcal{B}_i(\theta_i)}(\Sigma_{-i})$ of conditional beliefs for type θ_i, where

$$\mathcal{B}_i(\theta_i) = \{B \subset \Sigma_{-i} : \exists h \in \mathcal{H}(\theta_i), B = \Sigma_{-i}(h)\}.$$

(2) A player also has beliefs about himself and they may be relevant when we discuss the epistemic foundations of a solution concept. Once again, we do not explicitly consider such beliefs for notational simplicity. This does not alter the analysis in any essential way.

Our representation of a player's beliefs and our game theoretic analysis are consistent with the following epistemic assumption: at a state of the world where player i's type is θ_i and i's plan is $s_i \in S_i(\theta_i)$, player i would be *certain* of θ_i at each history $h \in \mathcal{H}(\theta_i)$ and would be *certain* to follow plan s_i at each history h consistent with s_i.

2.4. Sequential Rationality

A strategy $\hat{s}_i$ is sequentially rational for a player of type $\hat{\theta}_i$ with conditional beliefs μ^i if it maximises the conditional expected utility of $\hat{\theta}_i$ at every history h consistent with $\hat{s}_i$. Note that this a notion of rationality for plans of actions[16] rather than strategies (see, for example, Reny (1992)). Let

$$\mathcal{H}(\theta_i, s_i) = \{h \in \mathcal{H}(\theta_i) : (\theta_i, s_i) \in \Sigma_i(h)\}$$

and

$$S_i(\theta_i, h) = \{s_i \in S_i(\theta_i) : (\theta_i, s_i) \in \Sigma_i(h)\}$$

respectively denote the set of histories consistent with (θ_i, s_i) and the set of strategies consistent with (θ_i, h). Given a CPS $\mu^i \in \Delta^{\mathcal{B}_i}(\Sigma_{-i})$ and a history $h \in \mathcal{H}(\theta_i, s_i)$, let

$$U_i\left(\theta_i, s_i, \mu^i(\cdot|\Sigma_{-i}(h)\right) = \int_{\Sigma_{-i}(h)} U(\theta_i, s_i, \sigma_{-i})\mu^i\left(d\sigma_{-i} \mid \Sigma_{-i}(h)\right)$$

denote the expected payoff for type θ_i from playing s_i given h, provided that the integral on the right hand side is well-defined.[17]

Definition 2.3. *A strategy $\hat{s}_i$ $(i = 1, 2, ...)$ is sequentially rational for type $\hat{\theta}_i$ with respect to beliefs $\mu^i \in \Delta^{\mathcal{B}_i}(\Sigma_{-i})$, written $(\hat{\theta}_i, \hat{s}_i) \in \rho_i(\mu^i)$ or equivalently $\hat{s}_i \in r_i(\hat{\theta}_i, \mu^i)$, if for all $h \in \mathcal{H}(\hat{\theta}_i, \hat{s}_i)$ where player i is active and all $s_i \in S_i(\hat{\theta}_i, h)$ the following inequality is well-defined and satisfied:*

$$U_i\left(\hat{\theta}_i, \hat{s}_i, \mu^i(\cdot|\Sigma_{-i}(h)\right) \geq U_i\left(\hat{\theta}_i, s_i, \mu^i(\cdot|\Sigma_{-i}(h)\right).$$

Remark 1. *It can be shown that under standard, but somewhat restrictive assumptions the set of maximizers $r_i(\hat{\theta}_i, \mu^i)$ is non-empty. For example, if $S_i(\hat{\theta}_i)$ is compact and $U_i(\hat{\theta}_i, s_i, \sigma_{-i})$ is upper hemicontinuous in s_i, bounded and measurable in σ_{-i}, then $r_i(\hat{\theta}_i, \mu^i) \neq \emptyset$.*

[16] Formally, a *plan of action* is a maximal set of strategies consistent with the same histories and prescribing the same actions at such histories.

[17] Even in well-behaved games (e.g. the Ultimatum Game with a continuum of offers), for some choices of μ^i and/or s_i, the strategic form payoff function U_i is not integrable.

2.5. Extraneous Restrictions on Beliefs

A player's beliefs may be assumed to satisfy some restrictions that are not implied by mutual or common belief in rationality. We call such restrictions *extraneous*, although they may be related to some structural properties of the model. We may distinguish between (i) restrictions on beliefs about the state of Nature and chance moves and (ii) restrictions on beliefs about behavior. Our general theory and the following applications consider both (i) and (ii). Some examples of restrictions of the first kind are the following:

- Some "objective probabilities" of chance moves might be known or satisfy some known restrictions such as positivity or independence across nodes.[18]

- It may be common belief that all the opponents' payoff-types are considered possible *a priori* by each player. Or it may be common belief that the prior probability of a "crazy type" θ_i^* committed to play a strategy s_i^* (either because s_i^* is dominant for θ_i^* or because $S_i(\theta_i^*) = \{\theta_i^*\}$) is either positive or bounded below by a given positive number $\epsilon_i(\theta_i^*)$. This kind of restriction is considered in the analysis of reputation in section 5.1 (see also Watson (1993, 1996) and Battigalli and Watson (1997)).

The following are examples of restrictions of the second kind:

- Specific structural properties of the game such as stationarity (cf. Cho (1994)) or monotonicity may be somehow reflected in players' beliefs. Restrictions of this kind are considered in the analysis of disclosure and signaling in sections 5.2 and 5.3.

- The fact that an action a_i is conditionally weakly dominant at a history h for a type θ_i (e.g. bidding your valuation in a second price auction) may cause everybody to believe that the probability of a conditional on h being reached and θ_i being the true type of player i must be positive or bounded below by a given number $\delta(\theta_i, h, a_i) \in (0, 1]$. Similarly, the fact that action a_i is conditionally weakly dominated for type θ_i at history h (e.g. bidding above your valuation in a first price auction) may cause everybody to believe

[18] Börgers (1991) considers perturbed games with "small trembles" whereby the true trembling probabilities are unknown, but it is common belief that the actual choice is very likely to coincide with the intended choice. He stresses the differerence between correlated and uncorrelated trembles.

that the probability of this action conditional on (θ_i, h) is less than one or bounded above by a given number $\eta(\theta_i, h, a_i) \in [0, 1)$.

- It may be common belief that each player's beliefs about the types and strategies of different opponents satisfy stochastic independence.

- It may be common belief that each player's conditional beliefs have countable support (cf. Watson (1996) and section 5.3).

In general, we assume that, for each state of Nature θ, the conditional probability system of each player i belongs to a given, nonempty subset Δ^i. In order to make sense of the solution concepts discussed in the next section it is sufficient (but not necessary) to assume the following: *for all k=1,2,..., (for every $h \in \mathcal{H}$, every player $i \in N$ would be certain at h that)^k the first order beliefs of every player i* belong to Δ^i. Weaker sufficient epistemic assumptions are discussed in the next section.

3. Weak and Strong Δ-Rationalizability

In this section we define and analyze two nested extensions of the rationalizability solution concept to dynamic games of incomplete information, which take as given some extraneous restrictions on players' beliefs represented by sets of CPSs $\Delta^i \subset \Delta^{\mathcal{B}_i}(\Sigma_{-i})$, $i \in N$. Weak rationalizability is an extension of a solution concept put forward and analyzed by Ben Porath (1997) for games of perfect and complete information.[19] Strong rationalizability is a generalization of the notion of extensive form rationalizability proposed by Pearce (1984) and further analyzed by Battigalli (1996, 1997) (see also Reny (1992)). We focus mainly on two-person games (i.e. $N = \{1, 2\}$) to avoid discussing the issue of correlated *vs* independent beliefs, which would distract the readers' attention from more important points. The analysis is extended to n-person games in Section 4. The two solution concepts are defined by procedures which iteratively eliminate feasible type-strategy pairs and coincide on the class of static games. Epistemic assumptions are crucial for the motivation of these solution concepts. A formal epistemic analysis is beyond the scope of this paper and is provided elsewhere,[20] but we will try to

[19] See also Dekel and Fudenberg (1990), Brandenburger (1992), Börgers (1994) and Gul (1996a).

[20] Ben Porath (1997) analyzes weak rationalizability using finite, non universal, extensive form type spaces. Battigalli and Siniscalchi (1998) analyzes universal and non universal type

be explicit and clear about the epistemic assumptions underlying each solution concept.

A given state of the world describes the state of Nature (hence each player's private information) and the players' *dispositions* to act and to believe conditional on each history, that is, their strategies and their infinite hierarchies of conditional beliefs. Let $\Delta = (\Delta^i)_{i \in N}$. Each Δ-rationalizability solution concept characterizes the feasible type-strategy realized at states where (a) every player $i \in N$ is sequentially rational and has first order beliefs in Δ^i, and (b) the players' higher order conditional beliefs satisfy conditions concerning mutual certainty of (a) and/or robustness of beliefs about (a).

3.1. Weak Δ-Rationalizability

Weak Δ-rationalizability characterizes the set of feasible type-strategy pairs realized at states of the world where *all* the following events are true:[21]

(0) every player i has first order conditional beliefs in Δ^i and is sequentially rational,

(W1) every player i is certain of (0) at the beginning of the game (i.e. conditional on ϕ),

(W2) every player i is certain of (W1) at the beginning of the game,

...

(Wk) every player i is certain of (W(k-1)) at the beginning of the game,

... .

Definition 3.1. *Let $W_i(0, \Delta) = \Sigma_i$, $i = 1, 2$. Assume that the subsets $W_i(k, \Delta)$, $i = 1, 2$, have been defined, $k = 0, 1, \ldots$. Then for each $i = 1, 2$, $W_i(k+1, \Delta)$ is the set of feasible (θ_i, s_i) such that s_i is sequentially rational for θ_i with respect to some CPS $\mu^i \in \Delta^i$ such that $\mu^i(W_{-i}(k, \Delta)|\Sigma_{-i}) = 1$.[22] A feasible pair $(\theta_i, s_i) \in W_i(k, \Delta)$*

spaces for dynamic games of incomplete information and provides epistemic characterizations of solution concepts. Battigalli and Siniscalchi (1999a,b) use an extensive form, universal (or belief-complete) type space to provide an epistemic characterization of strong Δ-rationalizability with correlated and independent beliefs.

[21] The conditions are indexed by the assumed order of mutual certainty of rationality.

[22] It goes without saying that whenever we write a condition like $\mu^i(E|\Sigma_{-i}(h)) \geq \alpha$ and E is not measurable, the condition is *not* satisfied.

is called weakly (k,Δ)-rationalizable. *A feasible pair is* weakly Δ-rationalizable *if it is* weakly (k,Δ)-rationalizable for all $k = 1, 2, \ldots$. *The set of weakly* Δ-*rationalizable pairs for player* i *is denoted by* $W_i(\infty,\Delta)$.

There is a convenient way to reformulate Definition 3.1. For any subset $B_{-i} \subset \Sigma_{-i}$, let

$$\Lambda^i_\Delta(B_{-i}) = \{\mu^i \in \Delta^i : \mu^i(B_{-i}|\Sigma_{-i}) = 1\}.$$

Note that (a) $\Lambda^i_\Delta(B_{-i}) = \emptyset$ whenever B_{-i} is not measurable, (b) operator Λ^i_Δ is monotone[23] on the Borel sigma-algebra of Σ_{-i} and is also monotone with respect to Δ^i, and (c) $W_i(k+1,\Delta) = \rho_i\left(\Lambda^i_\Delta(W_{-i}(k,\Delta))\right)$.

$W_1(k,\Delta) \times W_2(k,\Delta)$ is the set of profiles consistent with assumptions (0)-(k-1) above. Note that these assumptions are silent about how the players would change their beliefs if they observed a history h which they believed impossible at the beginning of the game, even if h is consistent with rationality or mutual certainty of rationality of any order. Therefore weak rationalizability satisfies only a very weak form of backward induction (e.g. in two-stage games with perfect information) and can not capture any kind of forward induction reasoning. This is what makes weak rationalizability different from strong rationalizability.

3.2. Strong Δ-Rationalizability

According to strong rationalizability each player believes that his opponent is rational as long as this is consistent with his observed behavior. More generally, each player bestows on his opponent the highest degree of "strategic sophistication" consistent with his observed behavior (see Remak 2 below). This is a form of forward induction reasoning and it also induces the backward induction path in games of perfect and complete information (cf. Battigalli (1996, 1997)). To make the epistemic assumptions underlying strong rationalizability more transparent recall that a state of the world describes the players' *dispositions* to believe, that is, it describes not only how the players' actual beliefs evolve along the actual path, but also the beliefs the players *would* have at histories off the actual path. We say that player i *strongly believes* an event E if i will or would be certain of E at each history h consistent with E (see Battigalli and Siniscalchi (1999a) and references therein). Strong Δ-rationalizability characterizes the feasible type-strategy pairs realized at states of the world where *all* the following events are true:

[23] A set to set operator Λ is monotone if $E \subset F$ implies $\Lambda(E) \subset \Lambda(F)$.

(0) every player i has first order conditional beliefs in Δ^i and is sequentially rational,

(S1) every player i strongly believes (0),

(S2) every player i strongly believes (0) & (S1),

...

(Sk) every player i strongly believes (0) & (S2) &...& (S(k-1)),

... .

Definition 3.2. *Let* $\Sigma_i(0,\Delta) = \Sigma_i$ *and* $\Phi^i(0,\Delta) = \Delta^i$, $i = 1,2$. *Suppose that* $\Sigma_i(k,\Delta)$ *and* $\Phi^i(k,\Delta)$ *have been defined for each* $i = 1,2$. *Then for each* $i = 1,2$,

$$\Phi^i(k+1,\Delta) =$$

$$\{\mu^i \in \Phi^i(k,\Delta) : \forall h \in \mathcal{H}, \Sigma_{-i}(h) \cap \Sigma_{-i}(k,\Delta) \neq \emptyset \Rightarrow \mu^i(\Sigma_{-i}(k,\Delta)|\Sigma_{-i}(h)) = 1\},$$

$$\Sigma_i(k+1,\Delta) = \rho_i(\Phi^i(k,\Delta)).$$

A feasible pair $(\theta_i, s_i) \in \Sigma_i(k,\Delta)$ *is called* strongly (k,Δ)-rationalizable. *A feasible pair is* strongly Δ-rationalizable *if it is* strongly (k,Δ)-rationalizable for all $k = 1,2,...$. *The set of strongly* Δ*-rationalizable pairs for player* i *is denoted by* $\Sigma_i(\infty,\Delta)$.

Note that $W_i(1,\Delta) = \rho_i(\Delta^i) = \Sigma_i(1,\Delta)$. We show below that under regularity conditions, as the terminology suggests, the set of strongly (k,Δ)-rationalizable profiles is contained in the set of weakly (k,Δ)-rationalizable profiles and that the two sets coincide in static games (in general, it is sufficient that all the sets $W_i(k,\Delta)$ and $\Sigma_i(k,\Delta)$ ($i \in N$, $k = 1,2,..$) are nonempty and measurable).

Remark 2. *The set* $\Phi^i(n+1,\Delta)$ *can be characterized as follows: let* $\kappa(-i,h,n)$ *denote the highest index* $k \leq n$ *such that strongly* (k,Δ)*-rationalizable behavior by* $-i$ *is consistent with* $h \in \mathcal{H}$,[24] *then*

$$\Phi^i(n+1,\Delta) = \left\{\mu^i \in \Delta^i : \forall h \in \mathcal{H}, \mu^i(\Sigma_{-i}(\kappa(-i,h,n),\Delta)|\Sigma_{-i}(h)) = 1\right\} =$$

$$\bigcap_{k=0}^{n} \{\mu^i \in \Delta^i : \forall h \in \mathcal{H}, \Sigma_{-i}(h) \cap \Sigma_{-i}(k,\Delta) \neq \emptyset \Rightarrow \mu^i(\Sigma_{-i}(k,\Delta)|\Sigma_{-i}(h)) = 1\}.$$

[24] That is, $\kappa(-i,h,n)= \max\{k \in \{0,...,n\}$: $\Sigma_{-i}(h) \cap \Sigma_{-i}(k,\Delta) \neq \emptyset\}$.

3.3. Example: "Beer or Quiche" rivisited

In Section 1 we illustrated the rationalizability procedure for static games without extraneous restrictions on beliefs. Now we consider a dynamic game and we introduce extraneous, qualitative restrictions on beliefs about the state of Nature and the opponent's behavior. The game depicted in Figure 2 corresponds to the well-known Beer-Quiche example used by Cho and Kreps (1987) to illustrate the power of some equilibrium refinements in signaling games. Of course, Cho and Kreps analyze a standard extensive form game with a common prior. In their example the **surly** type $\theta(\sigma)$ has prior probability 0.9. They show that only the equilibrium whereby each type chooses B satisfies their "intuitive criterion." Here we replace this common prior assumption with the weaker restriction that (it is common belief that) the prior probability assigned by player 2 to $\theta(\sigma)$ is more than 50%. Furthemore, we 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 (**quiche**). Thus the restricted set of beliefs for player 2 is

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

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

[Insert Figure 2 about here]

It is easy to see that both **quiche** and **beer** are $(1, \Delta)$-rationalizable for both types.[25] But the second restriction on beliefs implies that if 2 fights after 1 has **beer** she also fights after 1 has **quiche**. This in turn implies that, according to any $(1, \Delta)$-rationalizable belief, a **fight** after **beer** is less likely than a **fight** after **quiche**:

$$\Phi^1(1, \Delta) = \{\mu^1 : \mu^1(f \mid B) \leq \mu^1(f \mid Q)\}.$$

Since the only reason for a **surly** type to have **quiche** is to decrease the probability of **fight**, a **surly** type with $(1, \Delta)$-rationalizable beliefs has **beer,** his preferred breakfast. On the other hand, it makes sense for a **wimp** with $(1, \Delta)$-rationalizable beliefs to forgo his preferred breakfast (**quiche**) hoping to avoid a

[25] Note that in a two-person game, if $\Sigma_i(1, \Delta) = \Sigma_i$, then $\Sigma_j(k+1, \Delta) = \Sigma_j(k, \Delta)$ for k odd, $j \neq i$, and $\Sigma_i(k+1, \Delta) = \Sigma_i(k, \Delta)$ for k even. Therefore we may consider only one player at each step.

fight. At this point we have to distinguish beteween weak and strong rationalizability. According to weak rationalizability, if player 2 is *a priori* certain to observe **beer** (a $(2,\Delta)$-rationalizable prior belief), her beliefs conditional on **quiche** are unrestricted. Therefore weak rationalizability has no further behavioral implications. On the other hand, strongly $(2,\Delta)$-rationalizable beliefs reflect a forward induction condition: **quiche** is a $(2,\Delta)$-rationalizable choice for a **wimp**, but not for a **surly** type, hence **beer** is sure evidence that player 1 is a **wimp**. Furthermore, taking into account that for all $(2,\Delta)$-rationalizable beliefs $\mu^2(\theta(\sigma)) > 1/2$ and $\mu^2(B|\theta(\sigma)) = 1$, we must have $\mu^2(\theta(\sigma)|B) \geq \mu^2(\theta(\sigma)) > 1/2$. This implies that the unique strongly Δ-rationalizable strategy for player 2 is "**fight** after **quiche, don't** fight after **beer.**" To summarize:

$$\Phi^2(2,\Delta) = \{\mu \in \Phi^2(1,\Delta) : \ \mu^2(\theta(\omega))|A) = 1, \mu^2(B|\theta(\sigma)) = 1\},$$

$$\Sigma_2(\infty,\Delta) = \Sigma_2(3,\Delta) = \{d \text{ if } B,\ f \text{ if } Q\}.$$

Given this, the only strongly Δ-rationalizable choice for a **wimp** is **beer**.

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

Thus, **quiche** is not rationalizable for either type and yet the "best rationalization" of this message is that it must have been sent by a **wimp.**

3.4. Results

3.4.1. General Properties

It is well-known that even for well-behaved dynamic games with a continuum of actions the strategic form payoff functions need not be continuous or measurable and hence the sequential best response correspondences $r_i(\theta_i, \mu^i)$ $(i \in N)$ need not be well-behaved. We first provide simple conditions on the "fundamentals" implying that the correspondences $r_i(\cdot,\cdot)$ are nonempty-valued and upper-hemicontinuous. Then we show that, if the latter properties are satisified and Δ (extraneous restrictions on beliefs) is "regular," weak and strong Δ-rationalizability are well-behaved.

Definition 3.3. *A game*

$$\Gamma = \langle N, (\Theta_i)_{i\in N}, (A_i)_{i\in N}, \mathcal{H}^*(\cdot), (u_i)_{i\in N}\rangle$$

is simple *if* Θ *is compact and either (a)* A *is finite or (b)* A *is compact and for some integer* T*,*
(b1) Γ *has* T *stages (that is, every terminal history* h *has length* $\ell(h) = T$*),*
(b2) for every $\theta \in \Theta$ *and* $h \in \mathcal{H}(\theta)$*, if* $\ell(h) < T - 1$ *then* $A(\theta, h)$ *is finite.*

Clearly, finite games and infinitely repeated games with a finite stage game are simple. Signaling games with a finite message spaces are simple if A_2 and Θ are compact. Signaling games with a continuum of messages are not simple.

Lemma 3.4. *For every simple game* Σ *is compact and, for each player* i, $r_i(\cdot, \cdot)$ *is an upper-hemicontinuous, nonempty-valued correspondence.*

Even in simple games, the set of (weakly or strongly) Δ-rationalizable profiles may be empty because the extraneous restrictions on beliefs represented by Δ may conflict with common certainty of rationality (or strong belief in rationality). But we can obtain a simple existence result and other regularity properties for the case where Δ only represents restrictions on the marginal prior beliefs about the opponent's type (existence results with a more general set of restrictions on beliefs can be obtained for specific models; see Section 5). For any subset C of a product set $X \times Y$ and for any probability measure μ on C let $\text{proj}_X C$ and $\text{marg}_X \mu$ respectively denote the projection of C on X and the marginal of μ on X, that is,

$$\text{proj}_X C = \{x \in X : \exists y \in Y, (x, y) \in C\}$$

$$(\text{marg}_X \mu)(E) = \mu(\{(x, y) \in C : x \in E\}),\ E \subset X \text{ (measurable)}.$$

Δ is *regular* if, for each player i, Δ^i is nonempty and closed, and there is a set $\Pi \subset \Delta(\Theta_{-i})$ such that

$$\Delta^i = \{\mu^i \in \Delta^{\mathcal{B}_i}(\Sigma_{-i}) : \text{marg}_{\Theta_{-i}} \mu^i(\cdot|\Sigma_{-i}) \in \Pi\}.$$

The following propositions are jointly proved in the Appendix (see proof of Proposition 7.1):

Proposition 3.5. *Suppose that* Δ *and* Δ' *are regular,* Σ *is compact,* $r_i(\cdot, \cdot)$ *is nonempty-valued and upper-hemicontinuous and* $\Delta^i \subset (\Delta^i)'$ *for every player* i*. Then for every player* i *and all* $k = 0, 1, ..., \infty$*,*
(a) the sets $W_i(k, \Delta)$ *and* $\Lambda^i_\Delta(W_i(k, \Delta))$ *of weakly* (k, Δ)*-rationalizable pairs and beliefs are nonempty and compact,* $\text{proj}_{\Theta_i} W_i(k, \Delta) = \Theta_i$*;*
(b) weak (k, Δ) *rationalizability implies weak* (k, Δ')*-rationalizability:* $W_i(k, \Delta) \subset$

$W_i(k, \Delta')$;
(c) $W_1(\infty, \Delta) \times W_2(\infty, \Delta)$ is the largest measurable subset $F_1 \times F_2 \subset \Sigma$ such that

$$F_1 \times F_2 = \rho_1 \left(\Lambda^1_\Delta (F_2) \right) \times \rho_2 \left(\Lambda^2_\Delta (F_1) \right).$$

Proposition 3.6. *Suppose that Δ is regular, Σ is compact and $r_i(\cdot, \cdot)$ is nonempty-valued and upper-hemicontinuous for every player i. Then for every player i and all $k = 0, 1, ..., \infty$,*
(1) the sets $\Sigma_i(k, \Delta)$ and $\Phi^i(k, \Delta)$ of strongly (k, Δ)-rationalizable pair and beliefs are nonempty and compact, $proj_{\Theta_i} \Sigma_i(k, \Delta) = \Theta_i$;
(2) strong (k, Δ)-rationalizability implies weak (k, Δ)-rationalizability: $\Sigma_i(k, \Delta) \subset W_i(k, \Delta)$ (the inclusion holds as an equality if there is only one stage).

Proposition 3.5 (a) (3.6 (1)) says that there is a weakly (strongly) rationalizable strategy for each payoff-type. (b) says that weak rationalizability is monotone with respect to extraneous restrictions on beliefs. This does not hold for strong rationalizability. In fact, if stronger restrictions on beliefs make fewer histories consistent with strongly k-rationalizable strategies, the k-forward induction criterion applies only to this smaller set of histories and the set of $(k+1)$-rationalizable profiles need not be smaller. (c) says the the set of weakly rationalizable profiles is the largest set with the "best reponse property."[26] As an immediate consequence of Lemma 3.4 and Propositions 3.5 and 3.6 we obtain the following:

Corollary 3.7. *In every simple game, if Δ is regular then (a), (c) of Proposition 3.5 and (1), (2) of Proposition 3.6 hold.*

3.4.2. Rationalizability and Bayesian Equilibria

A type in the sense of Harsanyi encodes the player's private information about the external state of Nature (the unknown parameters of the game) and also his epistemic type, that is, his infinite hierarchy of beliefs about the state of Nature and the beliefs of others. Although in the standard model of a Bayesian game these hierachies of beliefs are derived from a common prior on the set of states of the world (recall that a state of the world comprises a state of Nature and an epistemic state for each player), this need not be the case in general. Here we show that if the set of epistemic types is not restricted, rationalizability characterizes the set of outcomes realized in some Bayesian equilibrium. More precisely, we

[26] The equality can be replaced by (weak) inclusion (cf. Pearce (1984)).

prove an equivalence result relating weak rationalizability and a weak refinement of Bayesian equilibrium for dynamic games of incomplete information. For the sake of notational simplicity we do not consider any extraneous restriction on beliefs, i.e. $\Delta^i = \Delta^{\mathcal{B}_i}(\Sigma_{-i})$ for each i. Therefore we suppress any reference to Δ in the notation.

Let us fix an incomplete information game

$$\Gamma = \langle N, (\Theta_i)_{i\in N}, (A_i)_{i\in N}, \mathcal{H}^*(\cdot), (u_i)_{i\in N}\rangle$$

We first have to relate Γ to a Bayesian game. We do this by embedding Θ in a type space *à la* Harsanyi.

Definition 3.8. *A* Bayesian extension of Γ *is a tuple*

$$B_\Gamma = \langle \Gamma, (E_i)_{i\in N}, (T_i)_{i\in N}, (p_i)_{i\in N}\rangle$$

whereby for each player $i \in N$, (1) E_i is a metric space, (2) $T_i \subset \Theta_i \times E_i$ is a measurable set such that $\mathrm{proj}_{\Theta_i} T_i = \Theta_i$, *and (3) $p_i : T_i \to \Delta(T_{-i})$ is a measurable function.*

An element $t_i = (\theta_i, e_i)$ is an Harsanyi-type and e_i is its purely epistemic component which together with the private information θ_i determines the epistemic type $p_i(t_i)$. Since we are considering dynamic games we have to define an appropriate refinement of the Bayesian equilibrium concept. The refinement considered here is a generalization of Reny's (1992) "weakly sequentially rational assessments."

Definition 3.9. *A* weakly perfect Bayesian equilibrium *for a Bayesian extension $B_\Gamma = \langle \Gamma, (E_i)_{i\in N}, (T_i)_{i\in N}, (p_i)_{i\in N}\rangle$ is an array of functions $(b, g) = (b_i, g_i)_{i\in N}$ whereby for each player $i \in N$,*
(1) $b_i : T_i \to S_i$ and $g_i : T_i \to \Delta^{\mathcal{B}_i}(\Sigma_{-i})$ are measurable functions,
(2) (weak sequential rationality) for all $(\theta_i, e_i) \in T_i$, $b_i(\theta_i, e_i) \in r_i(\theta_i, g_i(\theta_i, e_i))$,
(3) (weak consistency) for all $t_i \in T_i$ and measurable subsets $B_{-i} \subset \Sigma_{-i}$

$$g_i(t_i)\left(B_{-i}|\Sigma_{-i}\right) = p_i(t_i)\left(\{(\theta_{-i}, e_{-i}) \in T_{-i} : (\theta_{-i}, b_{-i}(\theta_{-i}, e_{-i})) \in B_{-i}\}\right).$$

We say that a profile $(\theta_i, s_i)_{i\in N}$ is realizable *in (b, g) if, for each i, there is some $e_i \in E_i$ such that $s_i = b_i(\theta_i, e_i)$.*

The phrase "weakly perfect" is due to the fact that the equilibrium profile $(b_i, g_i)_{i\in N}$ only satisfies weak sequential rationality and consistency properties. The weak consistency condition (3) implies that $g_i(t_i)(\cdot|\Sigma_{-i}(h))$ is derived from b_{-i} and $p_i(t_i)$ for each history h with positive probability according to the equilibrium. For a defence of the rationality condition (2) see e.g. Reny (1992).[27] The following proposition says that the sets of weakly perfect Bayesian equilibrium outcomes and of weakly rationalizable outcomes coincide.

Proposition 3.10. [28] *Suppose that* Γ *is a simple game. Then*
(1) for every Bayesian extension B_Γ*, every weakly perfect Bayesian equilibrium* $(b_i, g_i)_{i\in N}$ *of* B_Γ*, and every profile of types* $(\theta_i, e_i)_{i\in N}$ *in* B_Γ*,* $(\theta_i, b_i(\theta_i, e_i))_{i\in N}$ *is weakly rationalizable;*
(2) there exists a Bayesian extension B_Γ *and a weakly perfect Bayesian equilibrium* (b, g) *such that, for every profile* $\sigma \in \Sigma$*,* σ *is weakly rationalizable if and only if* σ *is realizable in* (b, g).

3.4.3. Rationalizability and Dominance

The set of weakly and strongly rationalizable pairs can be further characterized for generic finite games in terms of dominance relations. We say that a game has *no relevant tie* if the following holds: for each player i and all pairs of outcomes $(\theta, h'), (\theta, h'') \in \mathcal{Z}$, if there are $h \in \mathcal{H}(\theta)$, $a', a'' \in A(\theta, h)$ such that $a'_i \neq a''_i$, z' follows (h, a') and z'' follows (h, a''), then $u_i(\theta, h') \neq u_i(\theta, h'')$. This means that if player i, immediately after history h, has deterministic beliefs about the true parameter θ and the continuation of the game, then he cannot be indifferent between any two feasible actions.

A strategy $s_i \in S_i(\theta_i)$ is *weakly dominated* by mixed strategy $m_i \in \Delta(S_i(\theta_i))$

[27] Even if we assumed expected utility maximization at each information set the equivalence result in the text would still hold for equilibrium and rationalizable *outcomes*. The relation to Reny's equilibrium concept is as follows: (i) A finite Bayesian extension of a finite two-person game with incomplete information can be represented as an extensive form game (with possibly heterogeneous and correlated priors on the set of initial nodes). (ii) A weakly perfect Bayesian equilibrium of the finite extension is realization equivalent to a weakly sequentially rational assessment of the extensive form.

[28] Brandenburger and Dekel (1987) prove a similar proposition for finite, static (or normal-form) games with complete information. See Forges (1993) for analogous results about objective correlated equilibria in games of incomplete information.

for type θ_i *on* $B_{-i} \subset \Sigma_{-i}$ if

$$\forall \sigma_{-i} \in B_{-i},\ U_i(\theta_i, s_i, \sigma_{-i}) \leq \sum_{s'_i} m_i(s'_i) U_i(\theta_i, s'_i, \sigma_{-i})$$

and

$$\exists \sigma'_{-i} \in B_{-i},\ U_i(\theta_i, s_i, \sigma_{-i}) < \sum_{s'_i} m_i(s'_i) U_i(\theta_i, s'_i, \sigma_{-i}).$$

The definition of strict dominance is analogous (all weak inequalities are replaced by strict inequalities). For any given rectangular subset $B \subset \Sigma$ let $\mathcal{W}(B)$ ($\mathcal{S}(B)$) denote the set of $(\theta_i, s_i)_{i \in N} \in \Sigma$ such that, for each i, s_i is not weakly (strictly) dominated for θ_i on B_{-i} and let $\mathcal{SW}(B) = \mathcal{S}(B) \cap \mathcal{W}(\Sigma)$. The iterated operator $\mathcal{SW}^n$ is defined in the usual way: $\mathcal{SW}^n(B) = \mathcal{SW}\left(\mathcal{SW}^{n-1}(B)\right)$, where $\mathcal{SW}^0(B) = B$. A subscript p denotes that we only consider *weak* domination by *pure* strategies. Thus $\mathcal{W}_p(B)$ is the set of profiles $(\theta_i, s_i)_{i \in N}$ such that s_i is not weakly dominated for θ_i by another *pure* strategy on B_{-i}, and $\mathcal{SW}_p(B) = \mathcal{S}(B) \cap \mathcal{W}_p(\Sigma)$. Note that $\mathcal{S}$ is a monotone operator. Therefore, also $\mathcal{SW}$ and $\mathcal{SW}_p$ are monotone operators. $W(k)$ and $\Sigma(k)$ denote the subsets of weakly and strongly k-rationalizable profiles, without extraneous restrictions on beliefs. The following proposition extends results proved by Pearce (1984) and Ben Porath (1997) to games with incomplete information.

Proposition 3.11. *(a) (cf. Pearce (1984)) In every finite and static game,*

$$\Sigma(k) = W(k) = \mathcal{S}^k(\Sigma),\ k = 1, 2, \ldots\ .$$

(b) In every finite game with no relevant ties,

$$\Sigma(k) \subset W(k) \subset \mathcal{SW}_p^k(\Sigma),\ k = 1, 2, \ldots\ .$$

(c) (cf. Ben Porath (1997)) In every finite game with no relevant ties, perfect information and private values,

$$\Sigma(k) \subset W(k) = \mathcal{SW}^k(\Sigma),\ k = 1, 2, \ldots\ .$$

An exact characterization of strong rationalizability can be obtained using a notion of iterated *conditional* dominance for each payoff-type. The characterization result can be easily adapted from Shimoji and Watson (1998). These characterizations of rationalizability through iterative dominance procedures can be used to compute the set of rationalizable strategies solving a sequence of linear programming problems (cf. Shimoji and Watson (1998), Section 4). The computation algorithm can also incorporate extraneous restrictions on conditional beliefs (Siniscalchi (1997b)).

4. Generalizations

The solution concepts defined in Section 3 for two-person games with observable actions can be extended to general n-person games with imperfect information about past actions. While the introduction of imperfect information is conceptually straightforward, considering more than two players forces a modeling choice between correlated and independent beliefs and poses the problem of providing a satisfactory definition of independence for conditional probability systems and an appropriate formalization of the forward induction principle for players with multiple opponents. In this section we briefly describe how to deal with these problems. A more complete analysis is provided in Battigalli (1995).

Imperfectly Observed Actions. In a game with observed actions the set of partial histories $\mathcal{H}$ can be regarded as a common collection of information sets for all the players. In games with imperfectly and asymmetrically observed actions each player i has his own collection of information sets $\mathcal{H}_i$, whereby a typical element $h \in \mathcal{H}_i$ now represents a (maximal) *set* of partial histories that player i cannot distinguish. Of course, $\mathcal{H}_i$ need only contain the information sets where player i is active. In order to adapt the analysis of the previous section to this situation it is sufficient to redefine $\Sigma(h)$ as the set of feasible profiles consistent with at least one history contained in h. Perfect recall implies that $\Sigma(h) = \Sigma_i(h) \times \Sigma_{-i}(h)$ for each $h \in \mathcal{H}_i$. The collection $\mathcal{B}_i$ of "relevant hypotheses" for player i is then defined as

$$\mathcal{B}_i = \{B \subset \Sigma_{-i} : \exists h \in \mathcal{H}_i, B = \Sigma_{-i}(h)\}$$

and this determines the space of conditional probability systems $\Delta^{\mathcal{B}_i}(\Sigma_{-i})$. Given these modifications, the other formal definitions are virtually unchanged.

***n*-Person Games and Independent Beliefs.** Extending the previous analysis to n-person games is quite straightforward if it is assumed that each player's beliefs concerning the type and strategy of different opponents may exhibit correlation. Therefore we consider here only the case of independent beliefs.

Recall that in games with observable actions the set $\Sigma(h)$ of feasible profiles consistent with a given history/information set h has a Cartesian structure: $\Sigma(h) = \prod_{i \in N} \Sigma_i(h)$. The same is true whenever h is an information set of a game with *observable deviators*. For the sake of simplicity, we limit our analysis to this class of games.[29] For any two players i and j let

$$\mathcal{B}_{ij} = \{B_j \subset \Sigma_j : \exists h \in \mathcal{H}_i, B_j = \Sigma_j(h)\}$$

[29] For a more general analysis see Battigalli (1995).

be the collection of "strategic form" pieces of information about player j that player i might obtain and let $\Delta^{\mathcal{B}_{ij}}(\Sigma_j)$ be the associated set of i's *marginal* CPS's about player j. A CPS $\mu^i \in \Delta^{\mathcal{B}_i}(\Sigma_{-i})$ is *independent* if there exists a vector of marginal CPS's $(\mu^i_j)_{j\neq i} \in \prod_{j\neq i} \Delta^{\mathcal{B}_{ij}}(\Sigma_j)$ such that, for all $h \in \mathcal{H}_i$, $\mu_i(\cdot|\Sigma_{-i}(h))$ is the product measure obtained from the vector of marginal probability measures $\left(\mu^i_j(\cdot|\Sigma_j(h))\right)_{j\neq i}$ (cf. Rényi (1955), p 303).

Assuming that the players are rational and have independent conditional beliefs and that this is common certainty at the beginning of the game, we obtain a notion of *weak rationalizability with independent beliefs*. The formal definition is essentially the same as in Section 2 except that now it has to be assumed that, for each player i, the restricted set of beliefs Δ^i contains only independent CPS's.[30]

Let us now turn to strong rationalizability. Since we assume that players' conditional beliefs are independent, we also incorporate in the definition of strong rationalizability a principle of *independent best rationalization*: each player i ascribes to every opponent j the "highest degree of strategic sophistication" consistent with j's observed behavior independently of any information about other players.[31] The formal, inductive definition of strong rationalizability (without extraneous restrictions on beliefs beyond independence) can be given as follows. Let μ^i_j denote the marginal on Σ_j of a given independent CPS μ^i.

(0) For all $i \in N$, $\Sigma^0_i = \Sigma_i$ and

$$\Phi^i(0) = \{\mu^i \in \Delta^{\mathcal{B}_i}(\Sigma_{-i}) : \mu^i \text{ is independent}\}.$$

(k+1) For all $i \in N$, $\Sigma^{k+1}_i = \rho_i(\Phi^i(k))$ and

$$\Phi^i(k+1) = \{\mu^i \in \Phi^i(k) : \forall h \in \mathcal{H}_i, \forall j \neq i, \Sigma_j(h) \cap \Sigma^k_j \neq \emptyset \Rightarrow \mu^i_j(\Sigma^k_j|\Sigma_j(h)) = 1\}.$$

5. Applications

In this section we apply weak and strong rationalizability with extraneous restrictions on beliefs to some models with one-sided incomplete information. We obtain results about reputation, disclosure and signaling previously derived for standard Bayesian (perfect) equilibria of Bayesian games whereby payoff-types coincide with Harsanyi-types.

[30] This notion of rationalizability (actually, only the first two steps in the iterative procedure) is used in Battigalli and Watson (1997).

[31] Battigalli and Siniscalchi (1997b) provides a rigorous epistemic axiomatization of the independent best rationalization principle.

5.1. Reputation[32]

Consider an infinitely repeated two-person game with discounting and one-sided incomplete information about feasible strategies. Player 2 does not know the set of feasible strategies of player 1. The finite or countable set of states of Nature Θ corresponds to the set of conceivable feasibility constraints for player 1. Let the stage game be $G = (A_1, A_2; v_1, v_2)$. The set of all (closed loop) strategies for player i in the repeated game is S_i. The set of feasible strategies for player 1's type θ is $S_1(\theta)$. For any state of Nature θ and any infinite history z feasible at θ, the long run average payoff function for player i is $u_i(\theta, z) = (1 - \delta_i) \sum_{t=1}^{\infty} \delta_i^{t-1} v_i(\alpha^t(z))$, where $\alpha^t(z)$ denotes the pair of actions chosen in period t along path z.[33] We are going to show that, under mild assumptions, two steps of weak rationalizability imply that a patient and unconstrained player 1 expects to obtain a long run average payoff approximately as large as the static Stackelberg payoff.

The stage game G satifies the following assumptions:

- Player 2 has a single-valued best response function $BR : A_1 \to A_2$ and

$$\min_{m_1 \in \Delta(A_1)} \max_{a_2 \in A_2} v_2(m_1, a_2) \geq 0.$$

  (The domain of function v_2 is extended to $\Delta(A_1) \times A_2$ *via* expected payoff calculations.)

- Player 1 has a pure maximin action, i.e. there is some "punishing" action a_1^P such that

$$v_2(a_1^P, BR(a_1^P)) = \min_{m_1 \in \Delta(A_1)} \max_{a_2 \in A_2} v_2(m_1, a_2).$$

The first assumption is made only for simplicity, the second is more substantial. Let

$$v_1^* = \max_{a_1 \in A_1} v_1(a_1, BR(a_1)).$$

[32] The model considered here is borrowed from Evans and Thomas (1997), which in turn builds on previous work by Fudenberg and Levine (1989). These papers derive a reputation result for standard equilibria of Bayesian models with one-sided incomplete information. See Sorin (1997) for a unified analysis of reputation and learning results in repeated games with incomplete information.

[33] Note that if Θ is finite the incomplete information game is "simple." Therefore the existence of (strongly and/or weakly) Δ-rationalizable strategies at every state of Nature is guaranteed whenever Δ is "regular" (see Definitions 3.3, 3.5 and Propositions 3.6, 3.7). However, finiteness of Θ is not needed for the following result.

denote player 1's static *Stackelberg payoff* and let a_1^* be a *Stackelberg action*, that is, an action attaining the maximum above. The best response to this action is $a_2^* = BR(a_1^*)$. Finally, $\underline{v}_i$ denotes the worst payoff for player i in G.

The feasibility correspondence $S_1(\cdot)$ satisfies the following assumptions:

- There is a "*normal*" unconstrained type $\theta^0 \in \Theta$ such that $S_1(\theta^0) = S_1$.

- There is a "*committment*" type θ^* such that $S_1(\theta^*) = \{s_1^*\}$ where s_1^* is a strategy "teaching" player 2 to play a_2^*. Strategy s_1^* plays a_1^* in normal phases and instigates punishment phases of incresing length when player 2 fails to play a_2^* in a normal phase (see Evans and Thomas (1997)). More precisely s_1^* is determined by an automaton with a countable set of states

  $$Q = \{\text{Norm}(k), \text{Punish}(k, j); k = 0, 1, ...; j = 0, ..., k\},$$

  where Norm(0) is the initial state; action and transition functions are given by the following table:

State	Action	Transition
Norm(k)	a_1^*	stay in Norm(k) if $a_2 = a_2^*$, go to Punish(k, k) otherwise
Punish($k, 0$)	a_1^P	go to Norm($k+1$)
Punish(k, j), $1 \leq j \leq k$	a_1^P	go to Punish($k, j-1$)

  (Norm(k) is the normal phase after k defections and Punish(k, j) is a the punishment phase after k defections and with j punishments periods to come).

As for extraneous restrictions on players' beliefs, we only assume the following:

- For some $\epsilon \in (0, 1)$ player 2 assigns at least ϵ prior probability to the commitment type θ^* and player 1 is certain of this, i.e. $\Delta = (\Delta^{\mathcal{B}_1}(S_2), \Delta^2(\epsilon))$ where

  $$\Delta^2(\epsilon) = \{\mu^2 \in \Delta^{\mathcal{B}_2}(\Sigma_1) : \mu^2(\{(\theta^*, s_1^*)\}|\Sigma_1) \geq \epsilon\}.$$

According to the following proposition, two steps of weak Δ-rationalizability imply that a patient and unconstrained player 1 should expect (*a priori*) to get a long run average payoff approximately as large as the static Stackelberg payoff. Player 1 can actually achieve this lower bound if he builds up a reputation of behaving like the committment type θ^*.

Proposition 5.1. *There is a positive integer* $M = M(v_2, \delta_2, \epsilon)$ *independent of* δ_1 *such that for any weakly* $(1, \Delta)$*-rationalizable belief* ν^1,

$$\sup_{s_1 \in S_1} U_1(\theta^0, s_1, \nu^1(\cdot|S_2)) \geq U_1(\theta^0, s_1^*, \nu^1(\cdot|S_2)) \geq (1 - \delta_1^M)\underline{v}_1 + \delta_1^M v_1^*.$$

Note that Evans and Thomas (1997, Section 4) prove an analogous result about the lower bound to player 1's equilibrium payoffs in a Bayesian game where player 1 knows the beliefs of player 2 about his type θ and, in particular, knows the prior probability assigned by player 2 to the committment type θ^*.

Sketch of proof.[34] To prove Proposition 5.1 it is sufficient to realize that the proof provided by Evans and Thomas does not use the full force of the foregoing equilibrium assumptions, but rather relies on the following facts: (i) player 2 assigns at least ϵ prior probability to the commitment type θ^* and plays a best response to his beliefs, (ii) player 1 is certain of (i) and also plays a best response to his own beliefs. Of course, these assumptions are implied by $(2, \Delta)$-rationalizability. Thus, one can prove that for any strategy combination (s_1^*, s_2), any real number $\eta \in (0, 1)$ and any positive integer k, player 2 can trigger a punishment and expect with probability at least $(1 - \eta)$ to be punished for less than k periods for at most $N = k + \frac{\log \epsilon}{\log(1-\eta)}$ times (cf. Lemma 1 in Evans and Thomas (1997)). Suppose s_2 is a best response to player 2' beliefs. Then, for any given payoff function v_2 satisfying our assumptions and any discount factor δ_2, we can choose η small enough and k large enough that s_2 fails to play a_2^* at most $K = k + N$ times. This implies that the path induced by (s_1^*, s_2) contains at most $M = (1/2)K(K + 3)$ periods when either a_2^* is not played or player 2 is punished (clearly M depends on ϵ, v_2 and δ_2). Therefore player 1's long run payoff is at least as large as if he got the worst payoff on the first M periods and the Stackelberg payoff afterward. ■

The result stated in Proposition 5.1 holds also when θ^* is a committment type always playing the Stackelberg action and the stage game G has "conflicting interests" in the sense of Schmidt (1993). A similar reputation result is proved by Battigalli and Watson (1997) for games where a patient long run player faces a sequence of short run opponents.

[34] A complete proof is available upon request.

5.2. Disclosure[35]

Consider a two-person signaling game where the sender, player 1, provides *certifiable information* about her own 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 sender's type $\theta \in \Theta$ 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. With this interpretation in mind the following is assumed:

- There is a finite ordered set of sender's types, $\Theta = \{\theta^1, ..., \theta^K\} \subset [0, 1]$ ($\theta^k < \theta^{k+1}$) and a continuum of receiver's actions $A_2 = [0, +\infty)$.
- The set of feasible messages for each type $\theta \in \Theta$ is

  $$M(\theta) = \{m \in 2^\Theta : \theta \in m\}.$$

  $M := \bigcup_{\theta \in \Theta} M(\theta)$ is the set of possible messages.[36]
- The sender's payoff $u : \Theta \times A_2 \to R$ is strictly increasing in its second argument.
- The receiver's expected payoff $v(\pi, a_2) := \sum_\theta \pi(\theta) v(\theta, a_2)$ is such that there is a well-defined best reply function $BR : \Delta(\Theta) \to A_2$ satisfying the following *weak monotonicity* property:

  $$\pi' \neq \pi'' \wedge \max Supp(\pi') \leq \min Supp(\pi'') \implies BR(\pi') < BR(\pi'') \quad (5.1)$$

  (standard conditions such as supermodularity of payoff in (θ, a_2) imply the weak monotonicity property (5.1)).

Fix any Bayesian game obtained from this model by assuming a strictly positive common prior on Θ. It can be shown that any Bayesian perfect equilibrium must satisfy *full disclosure*, that is, for all messages m, the receiver's chooses $a = BR(\min m)$ and no type θ sends a message m with $\min m < \theta$. Full disclosure is also implied by *strong rationalizability* assuming the following restrictions on conditional beliefs:

[35] This model of information transmission is borrowed from Grossman and Hart (1980). See also Okuno-Fujiwara *et al.* (1990), Bolton and Dewatripont (1997, Chapter 5) and the references therein.

[36] More generally, it is suffices to assume that M is *rich* in the following sense: $\forall \theta \in \Theta$, $\exists$ $m \in M(\theta)$, $\theta = \min m$.

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

Let us consider the first few steps in the strong Δ-rationalizability solution procedure. (1) First we eliminate all the receiver's strategies that are not sequential best responses to some $\mu^2 \in \Delta^2$. (2) Then we eliminate, for each type θ, all the messages $m \in M(\theta)$ with $m \neq \{\theta\}$ and $\max m = \theta$. The reason is that mild skepticism and weak monotonicity imply that a rational receiver would respond with a higher action if type θ disclosed sending message $\{\theta\}$. (3) Now apply the rationalization principle (forward induction): since no message m is inconsistent with steps (1) and (2) (it could be sent by type $\theta = \min m$), the receiver must be certain, conditional on every m, that (i) the sender is rational and (ii) the sender is certain that the receiver is rational and has mildly skeptical beliefs. Thus, by step (2), the receiver beliefs must satisfy $\mu^2(\max m|m) = 0$ unless m is a singleton; any strategy s_2 such that $s_2(m) = BR(\max m)$ for some non-singleton m is eliminated. (4) Now eliminate, for each type θ, all the messages/sets m containing at least three elements and such that θ is either the highest or second-to-highest element of m, because θ would induce a higher response by disclosing (the argument here is similar to step (2)). Continuing this way (for exactly $2K - 1$ steps) we obtain the following result:[37]

Proposition 5.2. *Every strongly Δ-rationalizable profile $(((\theta^1, m^1), ..., (\theta^K, m^K)), s_2)$ satisfies full disclosure: for all $k = 1, ..., K$ and $m \in M$, $\theta^k = \min m^k$ and $s_2(m) = BR(\min m)$.*

The argument above is similar to an *intuitive* "unraveling" argument used to show why a perfect Bayesian equilibrium must satisfy full disclosure (see e.g. Chapter 5 of Bolton and Dewatripont (1997)). The compellingness of this argument is due to its inductive structure. But a *rigorous* proof of the equilibrium result, one way or the other, has to proceed by contradiction and therefore is less transparent.

Note also that more general Bayesian extensions of the given economic model (whereby Harsanyi-types and payoff-types do not coincide) have perfect Bayesian

[37] A rigorous proof is available by request and can also be found in Battigalli (1995).

equilibria which satisfy mild skepticism, but do *not* satisfy full disclosure.[38] By Proposition 3.10, this implies that *weakly* Δ-rationalizable strategies need not satisfy full disclosure. This is not surprising since weak rationalizability does not capture forward induction reasoning and rationalizing the sender message is crucial for the proof of Proposition 5.2.

5.3. Costly 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 type θ' or θ'', with $0 < \theta' < \theta''$, chooses an education level $e \in [0, +\infty)$ and has payoff function $u(\theta, e, w) = w - \frac{g(e)}{\theta}$ with g differentiable, strictly increasing, and strictly convex. 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 she "rationally" sets the wage equal to the subjecitvely expected value of $e\theta$ conditional on e.

The restricted set of beliefs for player 1, Δ^1, is the set of (prior) probability measures $\mu^1(\cdot \mid S_2) \in \Delta(S_2)$ with *countable support.* As for player 2, we 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 type, is non-decreasing in e. Countability of supports is merely a technical assumption to simplify the analysis. Monotonicity is similar the "plausibility" property postulated by Kreps and Wilson (1982) in their analysis of reputation and entry deterrence.

Player 2's strategies can be represented by functions $\vartheta(e)$ giving the wage per unit of education. Therefore best reponses 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

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

[38] Here is a very simple example: There are two Harsany-types for the receiver, t_2' and t_2'', while Harsany-types coincide with payoff-types for the sender. The "interim" beliefs are $P_1(t_2'|\theta^k) = 1$ and $P_2(\theta^k|t_2') = 1/K = P_2(\theta^k|t_2'')$ for all k (these beliefs are consistent with a common prior with strictly positive marginal on Θ, that is, $P(\theta^k, t_2') = 1/K$ for all k). The posterior beliefs of type t_2' satisfy $\mu(\min m|m, t_2') = 1$ for all m. The posterior beliefs of type t_2'' are uniform: $\mu(\theta^k|m, t_2'') = 1/\#m$ iff $\theta^k \in m$. Each type θ^k chooses message $\{\theta^k\}$ and, of course, each type t_2 chooses a sequential best response to $\mu(\cdot|\cdot, t_2)$. This a perfect Bayesian equilibrium where the strategy of t_2'' does not satisfy full disclosure.

$\Omega(1,\Delta)$ 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 her 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 $\widehat{\vartheta}(\cdot) \in \Omega(1,\Delta)$.[39] At a subjectively optimal choice of education for type θ, say e^*, $\widehat{\vartheta}(\cdot)$ must be continuous from the right and the marginal rate of substitution $MRS(\theta, e^*) = \frac{1}{\theta}\frac{dg(e^*)}{de}$ must satisfy the first-order condition

$$MRS(\theta, e^*) \geq \widehat{\vartheta}(e^*) + e^* \cdot \frac{d\widehat{\vartheta}(e^*+)}{de}. \tag{5.2}$$

where $\frac{d\widehat{\vartheta}(e^*+)}{de}$ is the right-derivative of $\widehat{\vartheta}(\cdot)$ at e^*.[40]

As in the previous subsection, we focus on *strong* Δ*-rationalizability.* In fact, it is well-known that many patterns of behavior are consistent with Bayesian perfect equilibrium, and hence, *a fortiori*, with weak rationalizability, if no forward induction criterion is applied.

[Insert Figures 3 and 4 about here]

It turns out that the set of strongly Δ-rationalizable strategies 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),
- $\tilde{e}(\theta) = \arg\max_{e\geq 0} u(\theta, e, \widehat{\theta} e)$, $\theta \neq \widehat{\theta}$,
- $\overline{e}(\theta')$ such that $u(\theta', \overline{e}(\theta'), \theta''\overline{e}(\theta')) = u(\theta', e^*(\theta'), \theta' e^*(\theta'))$ and $\overline{e}(\theta'')$ such that $u(\theta'', \overline{e}(\theta''), \theta''\overline{e}(\theta'')) = u(\theta'', \tilde{e}(\theta''), \theta' \tilde{e}(\theta''))$,
- $\hat{e}(\theta'')$ such that $u(\theta'', \hat{e}(\theta''), MRS(\hat{e}(\theta'')) \cdot \hat{e}(\theta'')) = u(\theta'', \overline{e}(\theta'), \theta''\overline{e}(\theta'))$.

[39] Given belief $\mu_2(\cdot \mid S_2)$ about player 2 with countable support $\{s_2^1(\cdot), ..., s_2^k(\cdot), ...\}$ and correspondig salaries per unit of education $\{\vartheta^1(\cdot), ..., \vartheta^k(\cdot), ...\}$, let $\mu^k = \mu_2^k(s_2^k \mid S_2)$. Player 1's expected wage conditional on e is $\sum_k \mu^k s_2^k(e) = e\sum_k \mu^k \vartheta^k(e) = e\widehat{\vartheta}(e)$. Since for each k, $\vartheta^k(\cdot)$ is non decreasing with range in $[\theta', \theta'']$, $\widehat{\vartheta}(\cdot)$ must have the same properties.

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

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

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

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

Proposition 5.3. *The set of Δ-rationalizable choices of education for each type is as follows: If (a) $e^{\sim}(\theta'') > \bar{e}(\theta')$ or (b) $e^{\sim}(\theta'') \leq \bar{e}(\theta') \leq e^*(\theta'')$, then each type $\theta \in \{\theta', \theta''\}$ chooses 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 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 $\Sigma_2(k+1, \Delta) = \Sigma_2(k, \Delta)$, 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 $S_i(k, \Delta, \theta_i)$ denote the set of strongly (k, Δ)-rationalizable strategies for type θ_i of player i. Then

$$S_1(2, \Delta, \theta') = [e^*(\theta'), \bar{e}(\theta')],\ S_1(2, \Delta, \theta'') = [e^{\sim}(\theta''), \bar{e}(\theta'')].$$

To see this, first note that for any conjecture $\widehat{\vartheta}(\cdot) \in \Omega(1, \Delta)$ about player 2, the first order condition (5.2) 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 $e > \bar{e}(\theta')$ 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 $e^* \in [e^*(\theta'), e^{\sim}(\theta')]$ is a best response to the $(1, \Delta)$-rationalizable constant conjecture $\widehat{\vartheta}(e) \equiv MRS(\theta', e^*) \in [\theta', \theta'']$. Every $e^* \in [e^{\sim}(\theta'), \bar{e}(\theta')]$ is a best reply to the $(1, \Delta)$-rationalizable conjecture

$$\widehat{\vartheta}(e) = \begin{cases} \theta' & if \quad e < e^* \\ \theta'' & if \quad e \geq e^* \end{cases}. \tag{5.3}$$

$S_1(2,\Delta,\theta'')$ is obtained in a similar way. Using forward induction and monotonicity, the $(2,\Delta)$-rationalizable beliefs of the firm are (monotonic and) such that

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

Thus one obtains

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

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

Case (a): $\tilde{e}(\theta'') > \bar{e}(\theta')$. 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 strongly $(4,\Delta)$-rationalizable action for type θ is $e^*(\theta)$, $\theta = \theta'$, θ''. The strongly Δ-rationalizable strategies for player 2 are represented by functions in the set

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

■

6. Conclusions

We argued that Harsanyi's analysis of games with incomplete information is in principle very flexible, but for reasons of tractability most of its applications to economic models rely on questionable and non transparent assumptions about players' interactive beliefs, such as the common prior assumption and/or the conflation between payoff-types and Harsanyi-types. *A priori*, it is often not clear whether these assumptions are crucial. But in order to remove them and yet apply Harsanyi's analysis one would have to deal with mind-boggling Bayesian games featuring a universal type space, i.e. including all the conceivable infinite hierarchies of beliefs. Furthermore, unlike the Nash equilibrium concept, if there is *genuine* incomplete information we cannot justify the Bayesian equilibrium concept as a limit outcome of learning in repeated strategic interaction.

In this paper we provide a different, but related methodology. The primitives of our analysis are the sets of conceivable payoff-types, the (parametric) payoff functions and the feasibility correspondences characterizing a dynamic game with incomplete information. We may also take as given some extraneous restrictions on players' (first order) beliefs, but there is no need to specify a type space. In the first part of the paper we provide inductive definitions of a weak and a strong version of the rationalizability solution concept. Weak and strong rationalizability coincide in static games. In dynamic games strong rationalizability incorporates a forward induction principle, weak rationalizability does not. These solution concepts can be given a rigorous epistemic axiomatization, which is only briefly summarized in this paper. Existence and regularity results are provided for "simple," but possibly infinite games. Equivalence with iterated dominance procedures is proved for finite games. It turns out that, as it should be expected, weak rationalizability characterizes the set of all the Bayesian (perfect) equilibrium outcomes obtained by arbitrarily adding a type space *à la* Harsanyi to the given model. While the solution concepts have been defined for games with genuine incomplete information, they can be meaningfully applied to models of asymmetric information featuring an *ex ante* stage. In particular, the proposed solutions make sense for games representable with the "prior lottery model," provided that the statistical distribution of characteristics in the population of potential players is unknown. In the second part of the paper we apply weak and strong rationalizability to some economic models with one-sided incomplete information deriving results about reputation, disclosure and costly signaling.

The proposed methodology has several advantages with respect to the traditional one. First, the inductive solution can be computed without specifying an epistemic type space. Second, assumptions about interactive beliefs are typically weaker, more intuitive and transparent. Third, we can test the robustness of the results of the received Bayesian theory with respect to the specification of the type space. Fourth, the applications show that looking at rationalizable outcomes may clarify some aspects of strategic thinking that are overlooked or even obscured by standard equilibrium analysis.

7. Appendix

7.1. Incomplete Information Games: Feasibility Correspondence and Topological Structure

The sets of feasible actions for a given state of Nature θ and (feasible) history h are derived from the feasibility correspondence $\mathcal{H}^*(\cdot) : \Theta \to 2^{A^*}$ as follows:

$$A(\theta, h) = \{a \in A : (h, a) \in \mathcal{H}^*(\theta)\},$$

$$A_i(\theta_i, h) = \{a_i \in A_i : \exists a_{-i} \in A_{-i}, \exists \theta_{-i} \in \Theta_{-i}, (a_i, a_{-i}) \in A((\theta_i, \theta_{-i}), h)\}.$$

The feasibility correspondence satisfies the following properties (recall that A^* is the set of finite or countably infinite sequences of action profiles):

(1) for every $h \in A^*$ and $\theta \in \Theta$, if $h \in \mathcal{H}^*(\theta)$, every initial subsequence (prefix) of h belongs to $\mathcal{H}^*(\theta)$, in particular, $\phi \in \mathcal{H}^*(\theta)$ for all $\theta \in \Theta$,

(2) for every infinite sequence $h^* \in A^\infty$ and every $\theta \in \Theta$, if for every finite initial subsequence h of h^*, $h \in \mathcal{H}^*(\theta)$, then $h^* \in \mathcal{H}^*(\theta)$,

(3) for every $\theta = (\theta_i)_{i \in N} \in \Theta$, $h \in A^*$

$$A(\theta, h) = \prod_{i \in N} A_i(\theta_i, h),$$

$$A(\theta, h) = \emptyset \text{ if and only if for all } i \in N,\ A_i(\theta_i, h) = \emptyset.$$

We endow A^* and the set of outomes $\mathcal{Z} \subset \Theta \times A^*$ with the following metrics d_{A^*} and $d_{\mathcal{Z}}$: Recall that Θ_i and A_i are subsets of $\mathbf{R}^{m_i}$ and $\mathbf{R}^{n_i}$ respectively ($i \in N$). Let d_k be the Euclidean metric in $\mathbf{R}^k$ and $m = \sum_{i \in N} m_i$, $n = \sum_{i \in N} n_i$. Denote by $\ell(h)$ the length of a history ($\ell(h) = \infty$ if h is an infinite history) and let $\alpha^t(h)$ be the action profile at position t in history h ($t \leq \ell(h)$). If $\ell(h) \leq \ell(h')$, then

$$d_{A^*}(h, h') = \sum_{t=1}^{\ell(h)} (1/2)^t d_n(\alpha^t(h), \alpha^t(h')) + \sum_{t=\ell(h)+1}^{\ell(h')} (1/2)^t$$

(the second summation is zero if $\ell(h) = \ell(h')$),

$$d_{\mathcal{Z}}((\theta, h), (\theta', h')) = d_m(\theta, \theta') + d_{A^*}(h, h').$$

d_{A^*} is the natural metric for games with discounting. It can be checked that (A^*, d_{A^*}) and $(\mathcal{Z}, d_{\mathcal{Z}})$ are complete, separable, metric spaces.

The sets of strategies and strategy-type pairs are endowed with the "discounted" sup-metrics d_{S_i}, d_{Σ_i} and $d_{\Sigma_J}(i \in N,\ \emptyset \neq J \subset N,\ \Sigma_J = \prod_{i\in J} \Sigma_i)$:

$$d_{S_i}(s_i, s_i') = \sum_{t=0}^{\infty} (1/2)^t \left(\sup_{h:\ell(h)=t} d_{n_i}(s_i(h), s_i'(h)) \right),$$

$$d_{\Sigma_i}((\theta_i, s_i), (\theta_i', s_i')) = d_{m_i}(\theta_i, \theta_i') + d_{S_i}(s_i, s_i'),$$

$$d_{\Sigma_J}(\sigma_J, \sigma_J') = \sum_{i\in J} d_{\Sigma_i}(\sigma_i, \sigma_i').$$

7.2. Proofs

Proof of Lemma 2.1. Let $S_i(h)$ be the set of strategies consistent with history h. Clearly $S_i(h)$ is closed. Since

$$\Sigma_i(h) = \{(\theta_i, s_i) : s_i \in S_i(\theta) \cap S_i(h)\},$$

we only have to show that $S_i(\theta_i)$ is upper-hemicontinuous in θ_i. Suppose that $(\theta_i^k, s_i^k) \to (\theta_i, s_i)$ and $s_i^k \in S_i(\theta_i^k)$ for all k. Then for all $h' \in \mathcal{H}$, $s_i^k(h') \to s_i(h')$ and $s_i^k(h) \in A_i(\theta_i^k, h')$ for all k. Since $\mathcal{H}^*(\cdot)$ is continuous, each $A_i(\cdot, h')$ $(h' \in \mathcal{H})$ is also continuous. Therefore for all $h' \in \mathcal{H}$, $s_i(h') \in A_i(\theta_i, h')$ and $s_i \in S_i(\theta_i)$. ■

Proof of Lemma 3.4. In a simple game Θ and A are compact and either A is finite (case (a)) or $\mathcal{H}$ is finite (case (b)). If A is finite, S is a totally bounded, complete metric space. Therefore S is compact. If $\mathcal{H}$ is finite, S is topologically equivalent to a compact subset of a Euclidean space. In both cases $\Sigma \subset \Theta \times S$ is compact. By Lemma 2.1 each $\Sigma(h)$ is closed, hence compact.

We consider the rest of the proof for case (b) (A compact, finite horizon, finite sets of feasible actions through the second to last stage). The proof for case (a) is similar. Since $\Sigma_i(h)$ is the graph of the correspondence $S_i(\cdot, h)$, this correspondence is nonempty-compact-valued and upper-hemicontinuous. Now we show that it is also lower-hemicontinuous. Fix $h \in \mathcal{H}$ and suppose that $\theta_i^k \to \theta_i$ and $s_i \in S_i(\theta_i, h)$. By Assumption 0, each $A_i(\cdot, h')$, $(h' \in \mathcal{H})$ is continuous, hence lower-hemicontinuous. Therefore we can find a sequence of actions $(a_{i,h'}^k)_{k=1}^{\infty}$ such that $a_{i,h'}^k \to s_i(h')$ and $a_{i,h'}^k \in A_i(\theta_i^k, h')$. Let $s_i^k(h') = a_{i,h'}^k$ for all $h' \in \mathcal{H}$. By construction $s_i^k \in S_i(\theta_i)$ and $(s_i^k)_{k=1}^{\infty}$ converges pointwise to s_i. Since $\mathcal{H}$ is finite $s_i^k \to s_i$. If $h' \neq h$ is a prefix of h, then by assumption all $A_i(\theta_i^k, h')$ and $A_i(\theta_i, h')$ are finite. Thus, by continuity of $A_i(\cdot, h')$, $A_i(\theta_i^k, h') = A_i(\theta_i, h')$ and $s_i^k(h') = s_i(h')$ for k large. This implies that $s_i^k \in S_i(\theta_i^k, h)$. Therefore $S_i(\cdot, h)$ is lower-hemicontinuous.

The outcome function $\zeta^* : \Sigma \to \mathcal{Z}$ is continuous: suppose that $(\theta_i^k, s_i^k)_{i\in N}$ converges to $(\theta_i, s_i)_{i\in N}$, then for k large s^k and s induce the same action profile through the second-to-last stage and in the last stage the action profile induced by s^k converges to the action profile induced by s. Therefore the strategic payoff functions $U_i = u_i \circ \zeta^*$ are also continuous and (by compactness of Σ) bounded.

Since $S_i(\cdot, h)$ is nonempty-compact-valued and continuous and U_i is continuous and bounded, the conditional expected payoff $U_i\left(\theta_i, s_i, \mu^i(\cdot|\Sigma_{-i}(h)\right)$ is always well-defined and continuous in (θ_i, s_i, μ^i) and the correspondence

$$r_i(\theta_i, \mu^i, h) = \arg\max_{s_i \in S_i(\theta_i, h)} U_i\left(\theta_i, s_i, \mu^i(\cdot|\Sigma_{-i}(h)\right)$$

is nonempty-valued (for $h \in \mathcal{H}(\theta_i)$) and upper-hemicontinuous in (θ_i, μ^i). By a standard dynamic programming argument it can be shown that $r_i(\theta_i, \mu^i)$ is nonempty. We show that $r_i(\cdot, \cdot)$ is upper-hemicontinuous. Suppose that $(\theta_i^k, \mu^{i,k}, s_i^k) \to (\theta_i, \mu^i, s_i)$ and, for all k, $s_i^k \in r_i(\theta_i^k, \mu^{i,k})$. Since the game is simple, for k large s_i^k and s_i prescribe the same action through the second-to-last stage, which implies that $\mathcal{H}(\theta_i^k, s_i^k) = \mathcal{H}(\theta_i, s_i)$. This and upper-hemicontinuity of each correspondence $r_i(\cdot, \cdot, h)$ $(h \in \mathcal{H})$ imply that, for each history $h \in \mathcal{H}(\theta_i, s_i)$, $s_i \in r_i(\theta_i, \mu^i, h)$. Therefore $s_i \in r_i(\theta_i, \mu^i)$. ■

The following result summarizes Propositions 3.5 and 3.6.

Proposition 7.1. *Suppose that Δ and Δ' are regular, Σ is compact, $r_i(\cdot, \cdot)$ is nonempty-valued and upper-hemicontinuous and $\Delta^i \subset (\Delta^i)'$ for every player i. Then for every player i and all $k = 0, 1, ..., \infty$,*
(a) the sets $W_i(k, \Delta)$ and $\Sigma_i(k, \Delta)$ of weakly and strongly (k, Δ)-rationalizable profiles are nonempty and compact with $proj_{\Theta_i} W_i(k, \Delta) = proj_{\Theta_i} \Sigma_i(k, \Delta) = \Theta_i$, the sets $\Lambda_\Delta^i(W_i(k, \Delta))$ and $\Phi^i(k, \Delta)$ are nonempty and compact as well;
(b) $\Sigma_i(k, \Delta) \subset W_i(k, \Delta)$,
(c) $W_i(k, \Delta) \subset W_i(k, \Delta')$;
(d) $W_1(\infty, \Delta) \times W_2(\infty, \Delta)$ is the largest measurable subset $F_1 \times F_2 \subset \Sigma$ such that

$$F_1 \times F_2 = \rho_1\left(\Lambda_\Delta^1(F_2)\right) \times \rho_2\left(\Lambda_\Delta^2(F_1)\right).$$

Proof. First note that compactness of Σ and regularity of Δ imply that each set Δ^i is compact as well. Then observe that for every measurable subset $\emptyset \neq E_{-i} \subset \Sigma_{-i}$ such that $\text{proj}_{\Theta_{-i}} E_{-i} = \Theta_{-i}$ the following holds:

$$\emptyset \neq \{\mu^i \in \Delta^{\mathcal{B}_i}(\Sigma_{-i}) : \forall h \in \mathcal{H}, E_{-i} \cap \Sigma_{-i}(h) \neq \emptyset \Rightarrow \mu^i(E_{-i}|\Sigma_{-i}(h)) = 1\} \cap \Delta^i \subset$$

$$\{\mu^i \in \Delta^{\mathcal{B}_i}(\Sigma_{-i}) : \mu^i(E_{-i}|\Sigma_{-i}) = 1\} \cap \Delta^i = \Lambda^i_\Delta(E_{-i});$$

nonemptiness follows from measurability and the fact that, since $\text{proj}_{\Theta_{-i}} E_{-i} = \Theta_{-i}$ and Δ is regular, we are taking the intersection of nonempty sets characterized by logically independent properties. The inclusion holds because $\Sigma_{-i}(\phi) = \Sigma_{-i}$ and $E_{-i} \cap \Sigma_{-i}(\phi) \neq \emptyset$. The last equality is true by definition. Finally note that (a) , (b) and (c) are true by definition for $k = 0$. Assume that (a), (b) and (c) hold for all $k = 0, ..., n$.

(a, n+1) By the inductive hypothesis, the argument above implies the sets of weakly and strongly (n, Δ)-rationalizable beliefs $\Lambda^i_\Delta(W_{-i}(n, \Delta))$ and

$$\begin{aligned} \Phi^i(n, \Delta) &= \bigcap_{k=0}^{n} \{\mu^i \in \Delta^i : \\ \forall h &\in \mathcal{H}, \Sigma_{-i}(h) \cap \Sigma_{-i}(k, \Delta) \neq \emptyset \Rightarrow \mu^i(\Sigma_{-i}(k, \Delta)|\Sigma_{-i}(h)) = 1\} \end{aligned}$$

are nonempty and compact.

Since $r_i(\cdot, \cdot)$ is a nonempty-valued, upper-hemicontinuous and Θ_i is closed, each set $\rho_i(\mu^i) = \bigcup_{\theta_i \in \Theta_i} \{\theta_i\} \times r_i(\theta_i, \mu^i)$ is nonempty and closed and correspondence $\rho_i(\cdot)$ is upper-hemicontinuous. Therefore the sets of weakly and strongly $(n+1, \Delta)$-rationalizable pairs

$$\rho_i\left(\Lambda^i_\Delta(W_{-i}(n, \Delta))\right)$$

and

$$\Sigma_i(n+1, \Delta) = \rho_i\left(\Phi^i(n, \Delta)\right)$$

are nonempty and compact. Furthermore, nonemptiness of $r_i(\cdot, \cdot)$ implies that their projections on Θ_i coincide with Θ_i. This proves that (a) holds for all non-negative integers k. Clearly, compactness and the projection property hold also for $k = \infty$. Since the sequences of weakly and strongly (k, Δ)-rationalizable sets are nested, nonemptiness of

$$W_i(\infty, \Delta) = \bigcap_{k \geq 0} W_i(k, \Delta)$$

and

$$\Sigma_i(\infty,\Delta) = \bigcap_{k\geq 0} \Sigma_i(k,\Delta)$$

follows from the finite intersection property of compact sets.

(b, n+1) By the inductive hypothesis $\Sigma_{-i}(n,\Delta) \subset W_{-i}(n,\Delta)$ and both sets are measurable and nonempty. Therefore

$$\Lambda^i_\Delta(\Sigma_{-i}(k,\Delta)) \subset \Lambda^i_\Delta(W_{-i}(k,\Delta))$$

and

$$\Phi^i(k,\Delta) \subset \{\mu^i \in \Delta^i : \forall h \in \mathcal{H}, \Sigma_{-i}(h) \cap \Sigma_{-i}(n,\Delta) \neq \emptyset \Rightarrow \mu^i(\Sigma_{-i}(n,\Delta)|\Sigma_{-i}(h)) = 1\}$$
$$\subset \Lambda^i_\Delta(\Sigma_{-i}(k,\Delta)).$$

Thus we obtain

$$\Sigma_i(n+1,\Delta) = \rho_i(\Phi^i(n,\Delta)) \subset$$
$$\rho_i\left(\Lambda^i_\Delta(\Sigma_{-i}(n,\Delta))\right) \subset \rho_i\left(\Lambda^i_\Delta(W_{-i}(n,\Delta))\right) = W_i(n+1,\Delta).$$

Clearly the inclusion holds in the limit as $k \to \infty$.

(c, n+1) By the inductive hypothesis and part (a) $W_{-i}(n,\Delta) \subset W_{-i}(n,\Delta')$ and both sets are measurable. By monotonicity of operator $\rho_i \circ \Lambda^i_\cdot(\cdot)$ we obtain

$$W_i(n+1,\Delta) = \rho_i\left(\Lambda^i_\Delta(W_{-i}(n,\Delta))\right) \subset$$

$$\rho_i\left(\Lambda^i_\Delta(W_{-i}(n,\Delta'))\right) \subset \rho_i\left(\Lambda^i_{\Delta'}(W_{-i}(n,\Delta'))\right) = W_i(n+1,\Delta').$$

(d) (The following argument is a simple generalization of the proof of Proposition 3.1 in Bernheim (1984).) We first show that $W_1(\infty,\Delta) \times W_2(\infty,\Delta)$ contains every "fixed set" $F_1 \times F_2$. By definition $F_1 \times F_2 \subset W_1(0,\Delta) \times W_2(0,\Delta)$. Suppose that $F_1 \times F_2 \subset W_1(k,\Delta) \times W_2(k,\Delta)$. By part (a) each set $W_i(k,\Delta)$ is measurable. Thus, monotonicity of the operator $\rho_i \circ \Lambda^i_\Delta(\cdot)$ on the Borel sigma algebra of Σ_{-i} $(i = 1,2)$ implies

$$F_1 \times F_2 = \rho_1\left(\Lambda^1_\Delta(F_2)\right) \times \rho_2\left(\Lambda^2_\Delta(F_1)\right) \subset$$

$$\rho_1\left(\Lambda^1_\Delta(W_2(k,\Delta))\right) \times \rho_2\left(\Lambda^2_\Delta(W_1(k,\Delta))\right) = W_1(k+1,\Delta) \times W_2(k+1,\Delta).$$

Clearly the inclusion holds in the limit as $k \to \infty$. Now we show that $W_1(\infty,\Delta) \times W_2(\infty,\Delta)$ *is* a "fixed set." By part (a) $W_i(k,\Delta)$ is measurable for all $k =$

$0, 1, ..., \infty$. Thus monotonicity of $\rho_i \circ \Lambda^i_\Delta(\cdot)$ and $W_{-i}(\infty, \Delta) \subset W_{-i}(k, \Delta)$ ($k = 0, 1, ...$) yield

$$\rho_i \circ \Lambda^i_\Delta\left(W_{-i}(\infty, \Delta)\right) \subset \bigcap_{k \geq 0} \rho_i \circ \Lambda^i_\Delta\left(W_{-i}(k, \Delta)\right) = \bigcap_{k \geq 0} W_i(k+1, \Delta) = W_i(\infty, \Delta).$$

Therefore

$$\rho_1\left(\Lambda^1_\Delta(W_2(\infty, \Delta))\right) \times \rho_2\left(\Lambda^2_\Delta(W_1(\infty, \Delta))\right) \subset W_1(\infty, \Delta) \times W_2(\infty, \Delta).$$

Now suppose that $\sigma_i \in W_i(\infty, \Delta)$. Then there exists a sequence of CPSs $(\mu^{i,k})_{k=0}^{\infty}$ such that for all k, $\mu^{i,k} \in \Delta^i$, $\mu^{i,k}(W_{-i}(k, \Delta)|\Sigma_{-i}) = 1$ and $\sigma_i \in \rho_i(\mu^{i,k})$. Since $\Delta^{\mathcal{B}_i}(\Sigma_{-i})$ is compact, we may assume w.l.o.g. that $\mu^{i,k} \to \mu^i$. Since Δ^i is closed, $\mu^i \in \Delta^i$. Furthermore, it must be the case that $\mu^i(W_{-i}(k, \Delta)|\Sigma_{-i}) = 1$ for all k (otherwise, $\mu^{i,k}$ could not converge to μ^i) and thus (by continuity of the measure $\mu^i(\cdot|\Sigma_{-i})$) $\mu^i(W_{-i}(\infty, \Delta)|\Sigma_{-i}) = 1$. Since ρ_i is upper-hemicontinuous, $\sigma_i \in \rho_i(\mu^i)$. This shows that

$$W_1(\infty, \Delta) \times W_2(\infty, \Delta) \subset \rho_1\left(\Lambda^1_\Delta(W_2(\infty, \Delta))\right) \times \rho_2\left(\Lambda^2_\Delta(W_1(\infty, \Delta))\right).$$

■

Remark 3. *The the proof of part (b) uses only the fact that the sets of weakly and strongly (k, Δ)-rationalizable profiles are measurable and nonempty. The proof of part (c) relies only on measurability of the sets of weakly (k, Δ)-rationalizable profiles.*

Proof of Proposition 3.10. Recall that we are assuming no extraneous restrictions on beliefs. The set of weakly k-rationalizable strategies for player i is denoted $W_i(k)$ and the set of CPSs assigning prior probability one to some subset $B_{-i} \subset \Sigma_i$, is denoted $\Lambda^i(B_{-i})$.

(1) We show that every profile realizable in a weakly perfect Bayesian equilibrium (b, g) of any Bayesian extension $B_\Gamma = \langle \Gamma, (E_i)_{i \in N}, (T_i)_{i \in N}, (p_i)_{i \in N} \rangle$ is weakly rationalizable. Let Σ^b_i be the set of realizable pairs given b, that is,

$$\Sigma^b_i = \{(\theta_i, s_i) : \exists e_i, s_i = b_i(\theta_i, e_i)\}.$$

By definition,

$$\Sigma^b_i \subset \Sigma_i = W_i(0),\ g_i(T_i) \subset \Delta^{\mathcal{B}_i}(\Sigma_{-i}) = \Lambda^i(W_{-i}(0)),\ i = 1, 2.$$

Suppose that for each i, $\Sigma_i^b \subset W_i(k)$ and $g_i(T_i) \subset \Lambda^i(W_{-i}(k))$. Since (b, g) satisfies weak sequential rationality, the inductive hypothesis yields

$$\Sigma_i^b \subset \rho_i(g_i(T_i)) \subset \rho_i(\Lambda^i(W_{-i}(k))) = W_i(k+1),\ i = 1, 2.$$

The inclusion above and weak consistency of (b, g) yield

$$g_i(T_i) \subset \Lambda^i(\Sigma_{-i}^b) \subset \Lambda^i(W_{-i}(k+1))$$

as desired (note that Σ_{-i}^b and $W_{-i}(k+1)$ are both measurable).

(2) We construct a Bayesian extension B_Γ and a weakly perfect Bayesian equilibrium (b, g) such that $W_i(\infty) = \Sigma_i^b$. Since Γ is simple, Proposition 7.1 (d) yields

$$W_1(\infty) \times W_2(\infty) = \rho_1\left(\Lambda^1(W_2(\infty))\right) \times \rho_2\left(\Lambda^2(W_1(\infty))\right).$$

Therefore, for each $(\theta_i, s_i) \in W_i(\infty)$ there is a corresponding belief $g_i(\theta_i, s_i) \in \Lambda^i(W_{-i}(\infty))$ such that $s_i \in r_i(\theta_i, g_i(\theta_i, s_i))$. Since r_i is an upper-hemicontinuous correspondence, we may assume without loss of generality that function g_i is measurable. B_Γ is constructed as follows: for every i,

- $E_i = S_i$, $T_i = W_i(\infty)$,
- for all $t_i \in T_i$, $B_{-i} \subset T_{-i}$ (measurable) $p_i(t_i)(B_{-i}) = g_i(t_i)(B_{-i}|\Sigma_{-i})$.

By Proposition 7.1 (a), T_i is nonempty compact and $proj_{\Theta_i} T_i = \Theta_i$, $i = 1, 2$. Thus B_Γ is a well-defined Bayesian extension of Γ. Let $b_i(\theta_i, s_i) = s_i$ for all $(\theta_i, s_i) \in T_i$, $i = 1, 2$. b_i is obviously measurable. Weak sequential rationality and weak consistency are satisfied by construction. Therefore B_Γ is indeed a Bayesian extension of Γ and (b, g) is the desired weakly perfect Bayesian equilibrium. ■

Proof of Proposition 3.11. By Proposition 7.1 (b) we only have to consider the relationship between weak rationalizability and dominance. Take an arbitrary finite game. If $(\theta_i, s_i) \in \rho_i(\mu^i)$, then s_i is a best reply to the (prior) belief $\mu^i(\cdot|\Sigma_{-i})$ for type θ_i. This implies that s_i cannot be strictly dominated for type θ_i. Thus for every rectangular subset $B \subset \Sigma$

$$\rho_1(\Lambda^1(B_2)) \times \rho_2(\Lambda^2(B_1)) \subset \mathcal{S}(B).$$

(a) If the game is static, then it is also true that

$$\mathcal{S}(B) \subset \rho_1(\Lambda^1(B_2)) \times \rho_2(\Lambda^2(B_1))$$

(the proof can be easily adapted from Pearce (1984, Lemma 3)) and a standard inductive argument proves (a).

(b) If we assume that the game has no relevant ties, then $W(1) \subset \mathcal{W}_p(\Sigma)$ (the proof can be adapted from Battigalli (1997, Lemma 3). Thus $W(1) \subset \mathcal{S}(\Sigma) \cap \mathcal{W}_p(\Sigma) = \mathcal{SW}_p(\Sigma)$. Suppose that

$$W(n) \subset \mathcal{SW}_p^n(\Sigma).$$

Then

$$W(n+1) = \rho_1(\Lambda^1(W_2(n))) \times \rho_2(\Lambda^2(W_1(n))) \subset \mathcal{S}(\mathcal{SW}_p^n(\Sigma)) \cap \mathcal{W}_p(\Sigma) = \mathcal{SW}_p^{n+1}(\Sigma).$$

This proves statement (b).

(c) In every perfect information game with private values, $\mathcal{W}_p(\Sigma) = \mathcal{W}(\Sigma)$ (Battigalli (1997, Lemma 4) shows this result for games with perfect and complete information, the proof can be easily adapted to cover the present more general case). Thus, if the game has no relevant tie, part (b) implies $W(k) \subset \mathcal{SW}^k(\Sigma)$ for all k.[41] Suppose that

$$W(n) = \mathcal{SW}^n(\Sigma)$$

and let $(\theta_1, s_1, \theta_2, s_2) \in \mathcal{SW}^{n+1}(\Sigma)$. By the induction hypothesis and the definition of operator $\mathcal{SW}$, $(\theta_1, s_1, \theta_2, s_2) \in \mathcal{S}(\Sigma_\phi^n) \cap \mathcal{W}(\Sigma) \subset \Sigma_\phi^n$. Thus for each i, there are $\nu', \nu'' \in \Delta(\Sigma_j)$ such that $\nu'(W_{-i}(n)) = 1$, ν'' is strictly positive and s_i is a best response to ν' and ν'' for type θ_i (Pearce (1984, Lemmata 3 and 4)). Construct $\mu^i \in [\Delta(\Sigma)]^{\mathcal{B}_i}$ as follows: for all $h \in \mathcal{H}$, $B_{-i} \subset \Sigma_{-i}(h)$,

$$\mu^i(B_{-i}|\Sigma_{-i}(h)) = \frac{\nu(B_{-i})}{\nu(\Sigma_{-i}(h))},$$

where $\nu = \nu'$, if $\nu'(\Sigma_j(h)) > 0$, and $\nu = \nu''$ otherwise. It can be checked that μ^i is indeed a CPS ($\mu^i \in \Delta^{\mathcal{B}_i}(\Sigma_{-i})$), $\mu^i(W_{-i}(n)|\Sigma_{-i}) = 1$ and $(\theta_i, s_i) \in \rho_i(\mu^i)$. Thus $(\theta_i, s_i) \in W_i(n)$. ■

Proof of Proposition 5.3 (b), (c).

Case (b): $e^{\sim}(\theta'') \leq \bar{e}(\theta') \leq e^*(\theta'')$. In this case $S_1(4, \Delta, \theta') = \{e^*(\theta')\} \cup [e^{\sim}(\theta''), \bar{e}(\theta')]$. In fact, 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

[41] Ben Porath (1997, Lemma 2.1) independently proved that, in generic games with perfect (and complete) information, $W(1) \subset \mathcal{W}(\Sigma)$.

(5.3) with $e^* > \bar{e}(\theta')$. Every choice $e^* \in [\tilde{e}(\theta''), \bar{e}(\theta')]$ is justified by the $(3, \Delta)$-rationalizable conjecture (5.3). $S_1(4, \Delta, \theta'') = \{e^*(\theta'')\}$ 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 $S_1(4, \Delta, \theta') = \{e^*(\theta')\} \cup [\tilde{e}(\theta''), \bar{e}(\theta')]$ as in case (b), but $S_1(4, \Delta, \theta'') = [\hat{e}(\theta'')), \bar{e}(\theta')]$. In fact, by choosing $e > \bar{e}(\theta')$ player 1 is revealed as type θ''. Thus any choice $e^* > \bar{e}(\theta')$ is dominated by $e \in (\bar{e}(\theta'), e^*)$. 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 $\theta''-$ indifference curve through $(\bar{e}(\theta'), \theta''\bar{e}(\theta'))$ (see Figure 4). Choices $e^* \in [\hat{e}(\theta''), e^*(\theta''))$ are justified by $(3, \Delta)$-rationalizable conjectures

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

Choices $e^* \in [e^*(\theta''), \bar{e}(\theta')]$ are justified by $(3, \Delta)$-rationalizable conjectures (5.3). 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 (5.2).

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] BATTIGALLI, P. (1995): "Non-Equilibrium Analysis of Incomplete Information Games," Department of Economics, Princeton University.

[2] BATTIGALLI, P. (1996): "Strategic Rationality Orderings and the Best Rationalization Principle," *Games and Economic Behavior,* **13**, 178-200.

[3] BATTIGALLI, P. (1997): "On Rationalizability in Extensive Games," *Journal of Economic Theory*, **74**, 40-61.

[4] BATTIGALLI, P. and D.GUAITOLI (1996): "Conjectural Equilibria and Rationalizability in a Game with Incomplete Information," in (P.Battigalli, A. Montesano and F.Panunzi, Eds.) *Decisions, Games and Markets*. Dordrecht: Kluwer Academic Publishers.

[5] BATTIGALLI, P. and M. SINISCALCHI (1998): "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," Working Paper ECO 98/29, European University Institute.

[6] BATTIGALLI, P. and M. SINISCALCHI (1999a): "Interactive Beliefs and Forward Induction," Working Paper ECO 99/15, European University Institute.

[7] BATTIGALLI, P. and M. SINISCALCHI (1999b): "Interactive Beliefs, Epistemic Independence and Strong Rationalizability," forthcoming, *Research in Economics,* **53**, No 3.

[8] BATTIGALLI, P. and J. WATSON (1997): "On 'Reputation' Refinements with Heterogeneous Beliefs," *Econometrica*, **65**, 369-374.

[9] BEN PORATH, E. (1997): "Rationality, Nash Equilibrium and Backwards Induction in Perfect Information Games," *Review of Economic Studies,* **64**, 23-46.

[10] BERNHEIM, D. (1984): "Rationalizable Strategic Behavior," *Econometrica,* **52**, 1002-1028.

[11] BOLTON, P. and M. DEWATRIPONT (1997): *Introduction to Contract Theory*, mimeo.

[12] BONANNO, G. and K. NEHRING (1996): "How to Make Sense of the Common Prior Assumption Under Incomplete Information," working paper, University of California Davis. [Extended abstract in: I. Gilboa (ed.), *Theoretical Aspects of Rationality and Knowledge* (TARK 1998), Morgan Kaufman, San Francisco, 1998, 147-160.]

[13] BÖRGERS, T. (1991): "On the Definition of Rationalizability in Extensive Form Games," Discussion Paper 91-22, University College London.

[14] BÖRGERS, T. (1994): "Weak Dominance and Approximate Common Knowledge," *Journal of Economic Theory*, **64**, 265-276.

[15] BRANDENBURGER, A. (1992): "Lexicographic Probabilities and Iterated Admissibility," in Dasgupta, Gale, Hart and Maskin (eds.) *Economic Analysis of Markets and Games*. Cambridge MA: MIT Press.

[16] BRANDENBURGER, A. and E. DEKEL (1987): "Rationalizability and Correlated Equilibria," *Econometrica,* **55,** 1391-1402.

[17] BRANDENBURGER, A. and E. DEKEL (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory,* **59**, 189-198.

[18] CHO, I.K. (1994): "Stationarity, Rationalizability and Bargaining," *Review of Economic Studies,* **61**, 357-374.

[19] CHO, I.-K. and D. KREPS (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, **102**, 179-221.

[20] DEKEL, E. and D. FUDENBERG (1990): "Rational Behavior with Payoff Uncertainty," *Journal of Economic Theory,* **52**, 243-67.

[21] DEKEL, E. and F. GUL (1996): "Rationality and Knowledge in Game Theory," forthcoming in *Advances in Economics and Econometrics* (D. Kreps and K. Wallis, Eds.). Cambridge UK: Cambridge University Press.

[22] EVANS, R. and J.P. THOMAS (1997): "Reputation and Experimentation in Repeated Games with Two Long-Run Players," *Econometrica*, **65**, 1153-1173.

[23] FORGES, F. (1993): "Five Legitimate Definitions of Correlated Equilibrium in Games with Incomplete Information," *Theory and Decision*, **35**, 277-310.

[24] FUDENBERG, D. and D.K. LEVINE (1989): "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, **57**, 759-778.

[25] FUDENBERG, D. and D.K. LEVINE (1998): *The Theory of Learning in Games.* Cambridge MA: MIT Press.

[26] FUDENBERG D. and J. TIROLE (1991): *Game Theory.* Cambridge MA: MIT Press.

[27] GROSSMAN, S. and O. HART (1980): "Disclosure Laws and Takeover Bids," *Journal of Finance*, **35**, 323-3??.

[28] GUL, F. (1996a): "Rationality and Coherent Theories of Strategic Behavior," *Journal of Economic Theory,* **70**, 1-31.

[29] GUL, F. (1996b): "A Comment on Aumann's Bayesian View," *Econometrica*, forthcoming.

[30] HARSANYI, J. (1967-68): "Games of Incomplete Information Played by Bayesian Players. Parts I, II, III," *Management Science,* **14**, 159-182, 320-334, 486-502.

[31] HARSANYI, J. (1995): "Games with Incomplete Information," *American Economic Review*, **85**, 291-303.

[32] KHANEMAN, M., M. PERRY and P. RENY (1995): "Imperfect Information without Common Prior," mimeo.

[33] KREPS, D. and R.WILSON (1982): "Reputation and Imperfect Information," *Journal of Economic Theory*, **27**, 253-279.

[34] MERTENS J.F. and S. ZAMIR (1985): "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, **14**, 1-29.

[35] MORRIS, S. (1995): "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, **11**, 227-253.

[36] MYERSON, R. (1986): "Multistage Games with Communication," *Econometrica,* **54**, 323-358.

[37] OKUNO-FUJIWARA, M., A. POSTELWAITE, and K. SUZUMURA (1990): "Strategic Information Revelation," *Review of Economic Studies,* **57**, 25-47.

[38] OSBORNE, M. and A.RUBINSTEIN (1994): *A Course in Game Theory.* Cambridge MA: MIT Press.

[39] PEARCE, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica,* **52**, 1029-1050.

[40] PERRY, M. and P. RENY (1999): "A General Solution to King Solomon's Dilemma," *Games and Economic Behavior,* **26**, 279-285.

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

[42] RENY, P. (1992): "Backward Induction, Normal Form Perfection and Explicable Equilibria," *Econometrica,* **60**, 626-649.

[43] RÊNYI, A. (1955): "On a New Axiomatic Theory of Probability," *Acta Mathematica Academiae Scientiarum Hungaricae,* **6**, 285-335.

[44] SÁKOVICS, J. (1997): "Games of Incomplete Information without Common Knowledge Priors," mimeo, Insitute of Economic Analysis, Barcelona.

[45] SCHMIDT, K. (1993): "Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests," *Econometrica*, **61**, 325-351.

[46] SINISCALCHI, M (1997a): "Rationalizable Bidding Behavior in Dynamic Auctions," work in progress, Stanford University.

[47] SINISCALCHI, M (1997b): Personal communication.

[48] SHIMOJI, M. and J. WATSON (1998): "Conditional Dominance, Rationalizability, and Game Forms," *Journal of Economic Theory,* **83**, 161-195.

[49] SORIN, S. (1997): "Merging, Reputation and Repeated Games with Incomplete Information," working paper 467, Laboratoire d'Économétrie, École Polytechnique, Paris.

[50] TAN, T. and S. WERLANG (1988): "The Bayesian Foundation of Solution Concepts of Games," *Journal of Economic Theory,* **45**, 370-391.

[51] WATSON, J. (1993): "A 'Reputation' Refinement without Equilibrium," *Econometrica,* **61**, 199-205.

[52] WATSON, J. (1996): "Reputation in Repeated Games with no Discounting," *Games and Economic Behavior*, **15**, 82-109.

[53] WATSON, J. (1998): "Alternating-Offer Bargaining with Two-Sided Incomplete Information," *Review of Economic Studies,* **65**, 573-594.