Robust Dynamic Mechanism Design

Antonio Penta*
University of Wisconsin, Madison
Dept. of Economics

First Version: May, 2009   This version: October 17, 2011

Abstract

In situations in which the social planner has to make several decisions over time, before agents have observed all the relevant information, static mechanisms may not suffice to induce agents to reveal their information truthfully. This paper focuses on questions of partial and full implementation in dynamic mechanisms, when agents’ beliefs are unknown to the designer (hence the term “robust”). It is shown that a social choice function (SCF) is (partially) implementable for all models of beliefs if and only if it is ex-post incentive compatible. Furthermore, in environments with single crossing preferences, strict ex-post incentive compatibility and a “contraction property” are sufficient to guarantee full robust implementation. This property limits the interdependence in agents’ valuations.

Full robust implementation requires that, for all models of agents beliefs, all the perfect Bayesian equilibria of a mechanism induce outcomes consistent with the SCF. This paper shows that, for a weaker notion of equilibrium and for a general class of dynamic games, the set of all such equilibria can be computed by means of a “backwards procedure” which combines the logic of rationalizability and backward induction reasoning. It thus provides foundation to a tractable approach to the implementation question, allowing at the same time stronger implementation results.

Keywords: backward induction reasoning – dynamic mechanism design – implementation – rationalizability – robustness

*[email: apenta@ssc.wisc.edu]. I thank my advisor, George Mailath, for his unique dedication, and committee members Andrew Postlewaite and Qingmin Liu. I am indebted to all of them for constantly challenging me throughout this project, as well as for their encouragement. I also thank participants to several seminars and conferences. I am specially grateful to Pierpaolo Battigalli, Dirk Bergemann, Stephen Morris, Alessandro Pavan, Marzena Rostek and Bill Sandholm for their helpful comments.
1 Introduction.

Several situations of economic interest present problems of mechanism design that are inherently dynamic. Consider the problem of a public authority (or “social planner”) who wants to assign yearly licenses for the provision of a public good to the most productive firm in each period. Firms’ productivity is private information and may change over time; it may be correlated over time, and later productivity may depend on earlier allocative choices (for example, if there is learning-by-doing). Hence, the planner’s choice depends on private information of the firms, and the design problem is to provide firms with the incentives to reveal their information truthfully. But firms realize that the information revealed in earlier stages can be used by the planner in the future, affecting the allocative choices of later periods. Thus, in designing the mechanism (e.g. a sequence of auctions), the planner has to take into account “intertemporal effects” that may alter firms’ static incentives.

A rapidly growing literature has recently addressed similar problems of dynamic mechanism design, in which the planner has to make several decisions over different periods, with the agents’ information changing over time. In the standard approach, some commonly known distribution over the stochastic process generates payoffs and signals.\(^1\) Hence, it is implicitly assumed that the designer knows the agents’ beliefs about their opponents’ private information and their beliefs, conditional on all possible realizations of agents’ private information. In that approach, classical implementation questions can be addressed: For any given “model of beliefs”, we can ask under what conditions there exists a mechanism in which agents reveal their information truthfully in a Perfect Bayesian Equilibrium (PBE) of the game (partial implementation), or whether there exists a mechanism such that all the PBE of the induced game induce outcomes consistent with the social choice function (full implementation).

It is commonly accepted that the assumption that the designer knows the agents’ infinite hierarchies of beliefs is too strong. Not only are these assumptions strong, but the sensitivity of game theoretic results to the fine details of agents’ higher order beliefs is also well documented.\(^2\) Weakening the reliance of game theoretic analysis on common

---

\(^1\) Among others, see Bergemann and Valimaki (2010), Athey and Segal (2007), Pavan, Segal and Toikka (2011). Gershkov and Moldovanu (2009a,b) depart from the “standard” approach described above in that the designer does not know the “true” distribution, combining learning with incentive compatibility problems.

\(^2\) For recent work in this area, see for instance Weinstein and Yildiz (2007), or Penta (2011a) and Weinstein and Yildiz (2011) for dynamic games. In the context of mechanism design, a classical reference
knowledge assumptions seems thus crucial to enable us “to conduct useful analysis of practical problems” (Wilson, 1987, p.34).3

This paper focuses on the question of whether partial and full implementation can be achieved, in dynamic environments, when agents’ beliefs are unknown to the designer (hence the term “robust”). For the partial implementation question, building on the existing literature on static robust mechanism design (particularly, Bergemann and Morris, 2005) it is not difficult to show that a Social Choice Function (SCF) is PBE-implementable for all models of beliefs if and only if it is ex-post incentive compatible.4

The analysis of the full implementation question instead raises novel problems. On the one hand, the existing work on dynamic mechanism design has focused solely on problems of partial implementation.5 This is a serious limitation, but difficult to overcome, because even when the agents’ beliefs are known to the designer, characterizing the set of PBE of a given mechanism can be very difficult. On top of this, for “robust” full implementation we need a mechanism in which, for any model of beliefs, all the PBE induce outcomes consistent with the SCF. The direct approach to the question is thus to compute the set of PBE for each model of beliefs, and then take the union of all such equilibria. But given the difficulties discussed above, it might seem that adding the “robustness” requirement to the already difficult full implementation problem is doomed to make the problem untractable.

This paper introduces and provides foundations to a methodology that avoids the difficulties of the direct approach. The key ingredient is the notion of interim perfect equilibrium (IPE). IPE weakens Fudenberg and Tirole’s (1991) PBE allowing a larger set of beliefs off-the-equilibrium path. The advantage of weakening PBE in this context is twofold: on the one hand, full implementation results are stronger if obtained under a

---

3 In the context of mechanism design, this approach (sometimes referred to as the Wilson doctrine) has been put forward in a series of papers by Bergemann and Morris (2005, 2007 2009a,b, 2010, 2011), who developed a belief-free approach to classical implementation questions, known as “robust” mechanism design. In the same spirit, Chung and Ely (2007) provided foundations to dominant strategy mechanisms.

4 In environments with dynamic revelation of information, agents’ signals are intrinsically multidimensional. The negative result by Jehiel et al. (2006) thus set tight limits for the Wilson’s doctrine in dynamic environments. However, the literature provides several examples of important economic environments where ex-post implementation with multidimensional signals is possible (e.g. Picketty, 1999; Eso and Maskin, 2002; Bikhchandani, 2006. The point is further discussed in Bergemann and Morris, 2009).

5 In the best of my knowledge, this is the first work that studies full implementation in environments with dynamic revelation of information. The references in footnote 1 all focus on partial implementation problems. The closest work to dynamic full implementation is Lee and Sabourian’s (2011), who study repeated full implementation. The main difference between “dynamic” and “repeated” implementation is that, in the latter, the distribution of types, SCF and mechanism are the same in every period (hence they do not depend on the previous history.)
weaker solution concept (if all the IPE induce outcomes consistent with the SCF, then so do all the PBE, or any other refinement of IPE); on the other hand, the weakness of IPE is crucial to making the problem tractable. In particular, it is shown that the set of IPE-strategies across models of beliefs can be computed by means of a “backwards procedure” that combines the logic of rationalizability and backward induction reasoning: For each history, compute the set of rationalizable continuation-strategies, treating private histories as “types”, and proceed backwards from almost-terminal histories to the beginning of the game. (Refinements of IPE would either lack such a recursive structure, or require more complicated backwards procedures.)

These results are then applied to study conditions for full implementation in environments with monotone aggregators of information: In these environments information is revealed dynamically, and while agents’ preferences may depend on their opponents’ information (interdependent values) or on the signals received in any period, in each period all the available information (across agents and current and previous periods) can be summarized by one-dimensional statistics. In environments with single-crossing preferences, sufficient conditions for full implementation in direct mechanisms are studied: these conditions bound the amount of interdependence in agents’ valuations, and require that the “intertemporal effects” be sufficiently well-behaved.

The rest of the paper is organized as follows: Section 2 discusses an introductory example to illustrate the main concepts and insights. Section 3 introduces the notion of environments, which define agents’ preferences and information structure (allowing for information to be obtained over time). Section 4 introduces mechanisms. Models of beliefs, used to represent agents’ higher order uncertainty, are presented in Section 5. Section 6 is the core of the paper, and contains the main solution concepts and results for the proposed methodology. Section 7 focuses on the problem of partial implementation, while Section 8 analyzes the problem of full implementation in direct mechanisms. Proofs are in the Appendices.

2 A Dynamic Public Goods Problem.

I discuss here an example that introduces the main ideas and results, abstracting from some technicalities. The section ends with a brief discussion of the suitable generalizations of the example’s key features.

Consider an environment with two agents \((n = 2)\) and two periods \((T = 2)\). In each period \(t = 1, 2\), agents privately observe a signal \(\theta_{i,t} \in [0, 1]\), \(i = 1, 2\), and the planner

\footnote{This restriction helps overcoming the limits of ex-post incentive compatibility with multidimensional signals discussed in footnote 4.
chooses some quantity \( q_t \) of a public good. The cost function for the production of the public good is \( c(q_t) = \frac{1}{2} q_t^2 \) in each period, and for each realization \( \theta = (\theta_{i,1}, \theta_{i,2}, \theta_{j,1}, \theta_{j,2}) \), \( i, j = 1, 2 \) and \( i \neq j \), agent \( i \)'s valuation for one unit of the public goods \( q_1 \) and \( q_2 \) are, respectively,

\[
\alpha_{i,1}(\theta_1) = \theta_{i,1} + \gamma \theta_{j,1}
\]

and

\[
\alpha_{i,2}(\theta_1, \theta_2) = \varphi(\theta_{i,1}, \theta_{i,2}) + \gamma \varphi(\theta_{j,1}, \theta_{j,2})
\]

where \( \gamma \geq 0 \) and \( \varphi : [0, 1]^2 \rightarrow \mathbb{R} \) is assumed continuously differentiable and strictly increasing in both arguments. Notice that if \( \gamma = 0 \), we are in a private-values setting; for any \( \gamma > 0 \), agents have interdependent values. Also, since \( \varphi \) is strictly increasing in both arguments, there are “intertemporal effects”: the first period signal affects the agents’ valuation in the second period.

The notation \( \alpha_{i,t} \) is mnemonic for “aggregator”: functions \( \alpha_{i,1} \) and \( \alpha_{i,2} \) aggregate all the information available up to period \( t = 1, 2 \) into real numbers \( a_{i,1}, a_{i,2} \), which uniquely determine agent \( i \)'s preferences. Agent \( i \)'s utility function is

\[
\begin{align*}
    u_i(q_1, q_2, \pi_{i,1}, \pi_{i,2}, \theta) &= \alpha_{i,1}(\theta_1) \cdot q_1 + \pi_{i,1} \\
    &\quad + [\alpha_{i,2}(\theta_1, \theta_2) \cdot q_2 + \pi_{j,2}],
\end{align*}
\]

where \( \pi_{i,1} \) and \( \pi_{i,2} \) represent the quantity of private good in period \( t = 1, 2 \). The optimal provision of public good in each period is therefore

\[
\begin{align*}
    q_1^*(\theta_1) &= \alpha_{i,1}(\theta_1) + \alpha_{j,1}(\theta_1) \\
    q_2^*(\theta_1, \theta_2) &= \alpha_{i,2}(\theta_1, \theta_2) + \alpha_{j,2}(\theta_1, \theta_2).
\end{align*}
\]

Consider now the following direct mechanism: agents publicly report messages \( m_{i,t} \in [0, 1] \) in each period, and for each profile of reports \( m = (m_{i,1}, m_{i,2}, m_{j,1}, m_{j,2}) \), agent \( i \) receives transfers

\[
\begin{align*}
    \pi_{i,1}^*(m_{i,1}, m_{j,1}) &= -(1 + \gamma) \left[ \gamma \cdot m_{i,1} \cdot m_{j,1} + \frac{1}{2} m_{i,1}^2 \right] \\
    \pi_{i,2}^*(m_{i,1}, m_{j,1}) &= -(1 + \gamma) \left[ \gamma \cdot \varphi(m_{i,1} m_{i,2}) \cdot \varphi(m_{j,1} m_{j,2}) + \frac{1}{2} \varphi(m_{i,1} m_{i,2})^2 \right],
\end{align*}
\]

and the allocation is chosen according to the optimal rule, \( (q_1^*(m_1), q_2^*(m_1, m_2)) \).

If we complete the description of the environment with a model of agents’ beliefs, then the mechanism above induces a dynamic Bayesian game. The solution concept that will be used for Bayesian games is “interim perfect equilibrium” (IPE), a weaker version of PBE.
in which agents’ beliefs at histories immediately following a deviation are unrestricted (they are otherwise obtained via Bayesian updating).

“Robust” implementation though is concerned with the possibility of implementing a social choice function (SCF) irrespective of the model of beliefs. So, consider the SCF $f = (q_i^*, \pi_{i,t}^*, \pi_{j,t}^*)_{t=1,2}$ that we have just described: We say that $f$ is partially robustly implemented by the direct mechanism if, for any model of beliefs, truthfully reporting the private signal in each period is an “interim perfect equilibrium” (IPE) of the induced game.

For each $\theta = (\theta_{i,1}, \theta_{i,2}, \theta_{j,1}, \theta_{j,2})$ and $m = (m_{i,1}, m_{i,2}, m_{j,1}, m_{j,2})$, define

$$\Delta_i(\theta, m) = \varphi(m_{i,1}, m_{i,2}) - \varphi(\theta_{i,1}, \theta_{i,2})$$
$$- \gamma \cdot [\varphi(\theta_{j,1}, \theta_{j,2}) - \varphi(m_{j,1}, m_{j,2})]$$
$$= \alpha_{i,2}(m) - \alpha_{i,2}(\theta).$$

In words: given payoff state $\theta$ and reports $m$ (for all agents and periods), $\Delta_i(\theta, m)$ is the difference between the value of the aggregator $\alpha_{i,2}$ under the reports profile $m$, and its “true” value if payoff-state is $\theta$.

For given first period (public) reports $\hat{m}_1 = (\hat{m}_{i,1}, \hat{m}_{j,1})$ and private signals $\left(\hat{\theta}_{i,1}, \hat{\theta}_{i,2}, \right)$, and for point beliefs $(\theta_{j,1}, \theta_{j,2}, m_{j,2})$ about the opponent’s private information and report in the second period, if we ignore problems with corner solutions, then the best response $m_{i,2}^*$ of agent $i$ at the second period in the mechanism above satisfies:

$$\Delta_i\left(\theta_{i,1}, \theta_{i,2}, \theta_{j,1}, \theta_{j,2}, \hat{m}_1, m_{i,2}^*, m_{j,2}\right) = 0. \quad (6)$$

Also, given private signal $\hat{\theta}_{i,1}$, and point beliefs about $(\theta_{i,2}, \theta_{j,1}, \theta_{j,2}, m_{j,1}, m_{i,2}, m_{j,2}) \equiv (\theta_{i\setminus(i),1}, m_{i\setminus(i),1})$, the first period best-response satisfies:

$$m_{i,1}^* - \hat{\theta}_{i,1} = \gamma (\theta_{j,1} - m_{j,1})$$
$$+ \frac{\partial \varphi(m_{i,1}, m_{i,2})}{\partial m_{i,1}} \cdot \Delta\left(\theta_{i,1}, \theta_{i\setminus(i),1}, m_{i,1}^*, m_{i\setminus(i),1}\right) \quad (7)$$

This mechanism satisfies **ex-post incentive compatibility**: For each possible realization of $\theta \in [0, 1]^4$, conditional on the opponents reporting truthfully, if agent $i$ has reported truthfully in the past (i.e. $m_{i,1} = \theta_{i,1}$), then equation (6) is satisfied if and only if $m_{i,2} = \theta_{i,2}$. Similarly, given that $\Delta(\theta, m) = 0$ in the second period, the right-hand side of (7) is zero if the opponents report truthfully in the first period, and so it is optimal.

---

7We ignore here the possibility of corner solutions, which do not affect the fundamental insights. Corner solutions will be discussed in Section 8.
to report \( m_{i,1} = \theta_{i,1} \) (independent of the realization of \( \theta \)). Notice that this is the case for any \( \gamma \geq 0 \). Since such incentive compatibility is realized ex-post, conditioning to all information being revealed, incentive compatibility will also be realized with respect to any model of beliefs. Thus, for any such model of beliefs, there always exist an IPE that induces truthful revelation, that is, \( f \) is robustly partially implementable if \( \gamma \geq 0 \).

Even with ex-post incentive compatibility, it is still possible that, for some model of beliefs, there exists an IPE which does not induce truthful revelation: To achieve full robust implementation in this mechanism we must guarantee that all the IPE for all models of beliefs induce truthful revelation. We approach this problem indirectly, applying a “backwards procedure” to the belief-free dynamic game that will be shown to characterize the set of IPE-strategies for all models of beliefs. In the procedure, for each public history \( \tilde{m}_1 \) (profile of first-period reports), apply rationalizability in the continuation game, treating the private histories of signals as types; then, apply rationalizability at the first stage, maintaining that continuation strategies are rationalizable in the corresponding continuations.

Before illustrating the procedure, notice that equation (6) implies that, conditional on having reported truthfully in the first period \( (m_{i,1} = \theta_{i,1}) \), truthful revelation in the second period is a best-response to truthful revelation of the opponent irrespective of the realization of \( \theta \). Now, maintain that the opponent is revealing truthfully \( (m_{j,t} = \theta_{j,t} \text{ for } t = 1,2) \); if \( m_{i,1} \neq \theta_{i,1} \), i.e. if \( i \) has misreported in the first period, the optimal report in the second period is a further misreport \( (m_{i,2} \neq \theta_{i,2}) \), such that the implied value of the aggregator \( \alpha_{i,2} \) is equal to its true value (i.e.: \( \Delta(\theta, \tilde{m}_1, m_2) = 0 \).) This is the notion of self-correcting strategy, \( s^c_i \): a strategy that reports truthfully at the beginning of the game and at every truthful history, but in which earlier misreports (which do not arise if \( s^c_i \) is played) are followed by further misreports, to “correct” the impact of the previous misreports on the value of the aggregator \( \alpha_{i,2} \). It will be shown next that, if \( \gamma < 1 \), then the self-correcting strategy profile is the only profile surviving the “backward procedure” described above. Hence, given the results of Section 6, the self-correcting strategy is the only strategy played in all IPE for all “models of beliefs”. Since \( s^c \) induces truthful revelation, this implies that, if \( \gamma < 1 \), SCF \( f \) is fully robustly implemented.

For given \( \tilde{m}_1 \) and \( \theta_i = (\theta_{i,1}, \theta_{i,2}) \), let \( x_i(\theta_i) = [\varphi(\tilde{m}_{i,1}, m_{i,2}) - \varphi(\theta_{i,1}, \theta_{i,2})] \) denote type \( \theta_i \)'s “implied over-report” of the value of \( \varphi \). Then equation (6) can be interpreted as saying that “the optimal over-report of \( \varphi \) is equal to \( -\gamma \) times the (expected) opponent’s under-report of \( \varphi \).” Let \( x^0_j \) and \( \bar{x}^0_j \) denote the minimum and maximum possible values of \( x_j \). Then, if \( i \) is rational, his over-reports are bounded by \( x_i(\theta_i) \leq \bar{x}_i \equiv \gamma \cdot \bar{x}^0_j \) and \( x_i(\theta_i) \geq \underline{x}_i \equiv -\gamma \cdot \bar{x}_j \). Recursively, define \( \bar{x}_i^k = -\gamma \cdot \bar{x}_j^{k-1} \) and \( \underline{x}_i^k = -\gamma \sum_{j \neq i} \bar{x}_j^{k-1} \). Also, for each \( k \) and \( i \), let \( y_i^k = [\bar{x}_i^k - \underline{x}_i^k] \) denote the distance between the maximum and lowest
possible over-report at step \( k \). Then, substituting, we obtain the following system of difference equations:

\[
\mathbf{y}^k = \Gamma \cdot \mathbf{y}^{k-1}
\]

where

\[
\mathbf{y}^k = \begin{pmatrix} y_i^k \\ y_j^k \end{pmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix} 0 & \gamma \\ \gamma & 0 \end{bmatrix}
\]

(8)

Notice that the continuation game from \( \hat{m}_1 \) is dominance solvable if and only if \( \mathbf{y}^k \to 0 \) as \( k \to \infty \). In that case, for each \( \theta_i, x_i(\theta_i) \to 0 \), and so truthful reporting of \( \varphi \) (i.e., the continuation of the self-correcting strategy) is uniquely rationalizable in the continuation game. Thus, it suffices to study conditions for the dynamic system above to converge to the steady state \( \mathbf{0} \). In this example, \( \mathbf{0} \) is an asymptotically stable steady state if and only if \( \gamma < 1 \). Hence, if \( \gamma < 1 \), the only rationalizable outcome in the continuation from \( \hat{m}_1 \) guarantees that \( \Delta = 0 \). Given this, the first period best response simplifies to

\[
m^*_i,1 - \hat{\theta}_{i,1} = \gamma (\theta_{j,1} - m_{j,1}).
\]

The same argument can be applied to show that truthful revelation is the only rationalizable strategy in the first period if and only if \( \gamma < 1 \) (cf. Bergemann and Morris, 2009). Then, if \( \gamma < 1 \), the self-correcting strategy is the only “backwards rationalizable” strategy, hence the only strategy played as part of IPE for all models of beliefs.

**Key Properties and their Generalizations.** The analysis in Section 8 generalizes several features of this example: The notions of aggregator functions and of self-correcting strategy have a fairly straightforward generalization. An important feature of the example is that, in each period, the marginal rate of substitutions between \( q_t \) and the other goods is increasing in \( \alpha_{i,t} \) for each \( i \). This property implies that, for given beliefs about the space of uncertainty and the opponents’ messages, higher types report higher messages. Such monotonicity allowed us to construct the recursive system (8). A single-crossing condition will generalize this property in Section 8. Finally, the generalization of the idea that \( \gamma < 1 \) takes the form of a “contraction property”.

Consider the first period: for any \( \theta_1 \) and \( m_1 \),

\[
\alpha_{i,1}(m_{i,1}, m_{j,1}) - \alpha_{i,1}(\theta_{i,1}, \theta_{j,1}) = (m_{i,1} - \theta_{i,1}) + \gamma (m_{j,1} - \theta_{j,1}).
\]

Thus, if \( \gamma < 1 \), for any set of possible “deceptions” \( D \) there exists at least one agent \( i \in \{1, 2\} \) who can unilaterally sign \( \alpha_{i,1}(m'_{i,1}, m_{j,1}) - \alpha_{i,1}(\theta_{i,1}, \theta_{j,1}) \) by reporting some

\[8\]The name, borrowed from Bergemann and Morris (2009), is evocative of the logic behind the system (8).
message $m'_{i,1} \neq \theta_{j,1}$, irrespective of $\theta_{j,1}$ and $m_{j,1}$. That is, for all $\theta_{j,1}$ and $m_{j,1}$ in $D$:

$$\text{sign} \left[ \alpha_{i,1} \left( m'_{i,1}, m_{j,1} \right) - \alpha_{i,1} \left( \theta_{i,1}, \theta_{j,1} \right) \right] = \text{sign} \left[ m'_{i,1} - \theta_{i,1} \right].$$

Similarly, for public history $\hat{m}_1$ in the second period, $\gamma < 1$ guarantees that there exists at least one agent that can unilaterally sign $\left[ \alpha_{i,2} \left( \hat{m}_1, m'_{i,2}, m_{j,2} \right) - \alpha_{i,2} \left( \theta_i, \theta_j \right) \right]$ (uniformly over $\theta_j$ and $m_{j,2}$), by reporting some message $m'_{i,2}$ other than the one implied by the self-correcting strategy, $s^c_{i,2}$:

$$\text{sign} \left[ m'_{i,2} - s^c_{i,2} \right] = \text{sign} \left[ \alpha_{i,2} \left( \hat{m}_1, m'_{i,2}, m_{j,2} \right) - \alpha_{i,2} \left( \theta_i, \theta_j \right) \right].$$

This property will be required to hold at all histories.$^{9}$

3 Environments.

Consider an environment with $n$ agents and $T$ periods. In each period $t = 1, ..., T$, each agent $i = 1, ..., n$ observes a signal $\theta_{i,t} \in \Theta_{i,t}$. For each $t$, $\Theta_t := \Theta_{1,t} \times ... \times \Theta_{n,t}$ denotes the set of period-$t$ signals profiles. For each $i$ and $t$, the set $\Theta_{i,t}$ is assumed non-empty and compact subset of a finitely dimensional Euclidean space. For each agent $i$, $\Theta_i := \times_{t=1}^T \Theta_{i,t}$ is the set of $i$’s payoff types: a payoff-type is a complete sequence of agent $i$’s signals in every period. A state of nature is a profile of agents’ payoff types, and the set of states of nature is defined as $\Xi := \times_{i \in N} \Theta_i$.

In each period $t$, the social planner chooses an allocation from a non-empty subset of a finitely dimensional Euclidean space, $\Xi_t$ (possibly a singleton). The set $\Xi = \times_{t=1}^T \Xi_t$ denotes the set of feasible sequences of allocations. Agents have preferences over sequences of allocations that depend on the realization of $\Theta$: for each $i = 1, ..., n$, preferences are represented by utility functions $u_i : \Xi \times \Theta \to \mathbb{R}$. Thus, the states of nature characterize everybody’s preferences over the sets of feasible allocations.

An environment is defined by a tuple

$$\mathcal{E} = \langle N, \Xi, \Theta, (u_i)_{i \in N} \rangle,$$

assumed common knowledge.

Notice that an environment only represents agents’ information and preferences: it does not encompass agents’ beliefs. Each agent’s payoff-type $\theta_i \in \Theta_i$ represents his knowledge of the state of nature at the end of period $T$. That is, his information about everyone’s preferences over the feasible allocations.

$^{9}$The general formulation (Section 8) allows to accommodate the case analogous to the possibility of corner solutions in the example above.
For each $t$, let $Y^t_i := \times^t_{\tau=1} \Theta^t_i$ denote the set of possible histories of player $i$'s signals up to period $t$. For each $t$ and private signals $y^t_i = (\theta_{i,1}, ..., \theta_{i,t}) \in Y^t_i$, agent $i$ knows that the “true” state of nature $\theta^* \in \Theta$ belongs to the set $(y^t_i) \times (\times_{\tau=t+1}^T \Theta_{i,\tau}) \times \Theta_{-i}$.

At any point in time, agents form beliefs about the features of the environment they do not know. These beliefs should be interpreted here as purely subjective. Since robust mechanism design is concerned with problems of implementation as agents’ model of beliefs change, we maintain the description of the agents’ beliefs separate from the description of their information (which is part of the environment, and held constant). Models of beliefs are presented in Section 5.

**Social Choice Functions.** The description of the primitives of the problem is completed by the specification of a social choice function (SCF), $f : \Theta \rightarrow \Xi$.

Given the constraints of the environment, a necessary condition for a SCF to be implementable is that period-$t$ choices be measurable with respect to the information available in that period. We thus assume the following condition:

**Condition 1 (Measurability)** A SCF $f : \Theta \rightarrow \Xi$ is **measurable** if there exist functions $f_t : Y^t \rightarrow \Xi_t$, $t = 1, ..., T$, such that for each $\theta = (\theta_1, ... \theta_T)$, $f(\theta) = (f_t(\theta_1, ..., \theta_t))_{t=1}^T$.

In the following, we will only consider SCF that satisfy such necessary condition. We thus write $f = (f_t)_{t=1}^T$.

### 4 Mechanisms

A mechanism is a tuple

$$\mathcal{M} = \left( (M_{i,t})_{i \in N} \right)_{t=1}^T, (g_t)_{t=1}^T$$

where each $M_{i,t}$ is a non-empty set of messages available to agent $i$ at period $t$ ($i \in N$ and $t = 1, ..., T$); $g_t$ are “outcome functions”, assigning allocations to each history at each stage. As usual, for each $t$ we define $M_t = \times_{i \in N} M_{i,t}$. It is assumed that the reported messages and the chosen allocation are publicly observed at the end of each period.

Formally, let $H^0 := \{\phi\}$ (\phi denotes the empty history). For each $t = 1, ..., T$, the period-$t$ outcome function is a mapping $g_t : H^{t-1} \times M_t \rightarrow \Xi_t$, where for each $t$, the set of public histories of length $t$ is defined as:

$$H^t = \{(h^{t-1}, m_t, \xi_t) \in H^{t-1} \times M_t \times \Xi_t : \xi_t = g_t (h^{t-1}, m_t)\}.$$

The set of public histories is defined as $\mathcal{H} = \cup_{t=0}^T H^t$. Throughout the paper we focus on compact mechanisms, in which the sets $M_{i,t}$ are compact subsets of finitely dimensional Euclidean spaces.
4.1 Belief-Free Dynamic Games

An environment $\mathcal{E}$ and a mechanism $\mathcal{M}$ determine a belief-free dynamic game, that is a tuple

$$(\mathcal{E}, \mathcal{M}) = \langle N, (\mathcal{H}_i, \Theta_i, u_i)_{i \in N} \rangle.$$ 

Sets $N$, $\Theta$, and payoff functions $u_i$ are as defined in $\mathcal{E}$, while sets $\mathcal{H}_i$ are defined as follows: the set of player $i$’s private signals is given by $Y^t_i = (\times_{\tau=1}^t \Theta_{i,\tau})$; sets $H^t_t (t = 0, 1, \ldots, T)$ are defined as in $\mathcal{M}$; player $i$’s set of private histories of length $t$ ($t = 1, \ldots, T$) is defined as $H^t_t := H^{t-1} \times Y^t_i$, and finally $\mathcal{H}_i := \{ \phi \} \cup (\bigcup_{\tau=1}^T H^t_{i,\tau})$ denotes the set of $i$’s private histories. Thus, each private history of length $t$ is made of two components: a public component, made of the previous messages of the agents and the allocations chosen by the mechanism in periods 1 through $t - 1$; and a private component, made of agent $i$’s private signals from period 1 through $t$.

It is convenient to introduce notation for the partial order representing the precedence relation on the sets $\mathcal{H}$ and $\mathcal{H}_i$: $h^\tau \prec h^t$ indicates that history $h^\tau$ is a predecessor of $h^t$ (similarly for private histories: $(h^{\tau-1}_i, y^\tau) \prec (h^{t-1}_i, y^t)$ if and only if $h^\tau \prec h^t$ and $y^\tau \prec y^t$).

Remark 1 The tuple $(\mathcal{E}, \mathcal{M})$ is not a Bayesian game (Harsanyi, 1967-68), because it does not encompass a specification of agents’ interactive beliefs. A Bayesian game is obtained by appending a model of beliefs $\mathcal{B}$, introduced in Section 5. Concepts and notation for structures $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ will be introduced in Section 6.1.

4.1.1 Strategic Forms

Agents’ strategies in the game $(\mathcal{E}, \mathcal{M})$ are measurable functions $s_i : \mathcal{H}_i \rightarrow M_i$ such that $s_i(h^t_i) \in M_i, t$ for each $h^t_i \in \mathcal{H}_i$. The set of player $i$’s pure strategies is denoted by $S_i$, and as usual we define the sets $S = \times_{i \in N} S_i$ and $S_{-i} = \times_{j \neq i} S_j$. Payoffs are defined as in $\mathcal{E}$, as functions $u_i : \Xi \times \Theta \rightarrow \mathbb{R}$. For any strategy profile $s \in S$, each realization of $\theta \in \Theta$ induces a terminal allocation $g^s(\theta) \in \Xi$. Hence, we can define strategic-form payoff functions $U_i : S \times \Theta \rightarrow \mathbb{R}$ as $U_i(s, \theta) = u_i(g^s(\theta), \theta)$ for each $s$ and $\theta$.

As the game unfolds, agents learn about the environment observing the private signals, but they also learn about the opponents’ behavior through the public histories: for each public history $h^t_i$ and player $i$, let $S_i(h^t_i)$ denote the set of player $i$’s strategies that are consistent with history $h^t_i$ being observed. Clearly, since $i$’s private histories are only informative of the opponents’ behavior through the public history, for each $i$, $h^t_i = (h^{t-1}_i, y^t_i) \in \mathcal{H}_i$ and $j \neq i$, $S_j(h^t_i) = S_j(h^{t-1}_i)$. Sets $\mathcal{H}_i$ and $S_i$ are endowed with the standard metrics derived from the Euclidean metric on $H^T \times \Theta$.

\[\text{See Appendix A.1 for details.}\]
4.2 Direct Mechanisms

The notion of direct mechanism is based on the (necessary) Measurability Condition 1 (p. 10):

**Definition 1** A direct mechanism for SCF \( f = (f_t)_{t=1,...,T} \), denoted by \( \mathcal{M}(f) \), is a mechanism such that for each \( i \) and for each \( t = 1,...,T \), \( M_{i,t} = \Theta_{i,t} \), and \( g_t = f_t \).

In a direct mechanism, agents are asked to announce their signals at every period. Based on the reports, the mechanism chooses the period-\( t \) allocation according to the function \( f_t : Y^t \to \Xi^t \), as specified by the SCF. The truthtelling strategies are those that, conditionally on having reported truthfully in the past, report truthfully the period-\( t \) signal, \( \theta_{i,t} \). Truthtelling strategies may differ in the behavior they prescribe at histories following past misreports, but they all are outcome equivalent and induce truthful revelation in each period. The set of such strategies is denoted by \( S^*_i \), while \( S^* \) denotes the set of strategy profiles that are outcome equivalent to truthtelling strategy profiles.

5 Models of Beliefs.

A model of beliefs for an environment \( \mathcal{E} \) is a tuple

\[ B = \langle N, \Theta, (B_i, \beta_i)_{i \in N} \rangle \]

such that for each \( i \), \( B_i \) is the set of types, assumed Polish, and \( \beta_i : B_i \to \Delta(\Theta \times B_{-i}) \) is a continuous function.\(^{11}\)

At period 0 agents have no information about the environment. Their (subjective) priors about the payoff state and the opponents’ beliefs are implicitly represented by means of types \( b_i \), as the beliefs \( \beta_i(b_i) \in \Delta(\Theta \times B_{-i}) \). In periods \( t = 1,...,T \), agents update their beliefs using their private information (the history of payoff signals), and other information possibly disclosed by the mechanism set in place. The main difference with respect to standard (static) type spaces with payoff types, as in Bergemann and Morris (2005) for example, is that players here do not know their own payoff-type at the interim stage: payoff-types are disclosed over time, and known only at the end of period \( T \). Thus, an agent’s type at the beginning of the game is completely described by a “prior” belief over the payoff states and the opponents’ types.

Standard models of dynamic mechanism design (e.g. Bergemann and Valimaki, 2008, Athey and Segal, 2007, Pavan et al., 2009) correspond to the case where, for each \( i \), \( B_i \)

\(^{11}\)A Polish space is a complete separable metric space. For measurable space \( X \), \( \Delta(X) \) denotes the set of probability measures on \( X \), endowed with the corresponding Borel sigma-algebra.
is a singleton and \( \text{supp}(\text{marg}_{\Theta_i} \beta_i(b_i)) = \Theta_i \). (In fact, standard models assume that such prior is common: \( \beta_i(b_i) = p^* \) for all \( i \).)

To summarize our terminology, in an environment with beliefs \((\mathcal{E}, \mathcal{B})\) we distinguish the following “stages”: in period 0 (the interim stage) agents have no information, their (subjective) prior is represented by types \( b_i \), with beliefs \( \beta_i(b_i) \in \Delta(\Theta \times B_{-i}) \); \( T \) different period-\( t \) interim stages, for each \( t = 1, \ldots, T \), when a type’s beliefs after a history of signals \( y_t^i \) will be concentrated on the set

\[
\{y_t^i\} \times \left( \prod_{\tau=t+1}^T \Theta_{i,\tau} \right) \times \Theta_{-i} \times B_{-i}.
\]

The term “ex-post” refers to hypothetical situations in which interim profiles are revealed: period-\( t \) ex-post stage refers to a situation in which everybody’s signals up to period \( t \) are revealed. By ex-post stage (or, period-\( T \) ex-post stage) we refer to the final realization, when payoff-states are fully revealed.

6 Solution Concepts.

This section is organized in two parts: the first, introduces the main solution concept for environments with a model of beliefs, interim perfect equilibrium (IPE); the second introduces a “backward procedure” for belief-free environments, which will be used in the analysis of the full-implementation problem in Section 8. Proposition 1 shows that the backwards procedure characterizes the set of IPE strategies for all models of beliefs.

6.1 Mechanisms in Environments with Beliefs: \((\mathcal{E}, \mathcal{M}, \mathcal{B})\).

A tuple \((\mathcal{E}, \mathcal{M}, \mathcal{B})\) determines a dynamic Bayesian game. Strategies in a Bayesian game are measurable mappings \( \sigma_i : B_i \to S_i \), and the set of (pure) strategies is denoted by \( \Sigma_i \). Agent \( i \)'s information sets in the Bayesian games are \( B_i \times \mathcal{H}_i \), with generic element \((b_i, h_i)\). At period 0, agents only know their own type. Period 0 histories are therefore \( h_0^i = (b_i, \phi) \in B_i \times \{\phi\} \).

A system of beliefs consists of collections \( (p_i(b_i, h_i^t))(b_i, h_i^t) \in B_i \times \mathcal{H}_i \) for each agent \( i \), such that \( p_i(b_i, h_i^t) \in \Delta(\Theta \times B_{-i}) \): a belief system represents agents’ conditional beliefs about the payoff state and the opponents’ types at each information set of the Bayesian game. A strategy profile and a belief system \((\sigma, p)\) form an assessment. For each agent \( i \), a strategy profile \( \sigma \) and conditional beliefs \( p_i \) induce, at each information set \((b_i, h_i^{t-1})\), a probability measure \( P^{\sigma, p_i}(b_i, h_i^{t-1}) \) over the private histories of length \( t \), \( h_i^t \in H^{t-1} \times Y_i^t \).

Definition 2 Fix a strategy profile \( \sigma \in \Sigma \). A beliefs system \( p \) is weakly preconsistent
with $\sigma$ if for each $i \in N$:
\begin{equation}
\forall h_i^0 \in B_i, p_i(b_i, \phi) = \beta_i(b_i) \tag{9}
\end{equation}

\begin{equation}
\forall h_i^t = (y_i^t, h_i^{t-1}) \in \mathcal{H}_i \setminus \{\phi\}, \\
\text{supp}[p_i(b_i, h_i^t)] \subseteq \{y_i^t\} \times (\times_{\tau=t+1}^T \Theta_{i,\tau}) \times \Theta_{-i} \times B_{-i} \tag{10}
\end{equation}

and for each $h_i^t$ such that $h_i^{t-1} \prec h_i^t$,
\begin{equation}
p_i(b_i, h_i^t) [E] \cdot P_{\sigma, p_i} (b_i, h_i^{t-1}) [h_i^t] = p_i(b_i, h_i^{t-1}) [E]. \tag{11}
\end{equation}

where the event $[h_i^t] \subseteq \Theta \times B_{-i}$ is defined as:
\begin{equation}
[h_i^t] = \{y_i^t\} \times (\times_{\tau=t+1}^T \Theta_{i,\tau}) \times B_{-i}.
\end{equation}

Condition (9) requires each agent’s beliefs conditional on observing type $b_i$ to agree with that type’s beliefs as specified in the model $B$; condition (10) requires conditional beliefs at each information set to be consistent with the information about the payoff state contained in the history itself; finally, condition (11) requires that belief system $p_i$ be consistent with Bayesian updating whenever possible.

From the point of view of each $i$, for each $(b_i, h_i^t) \in B_i \times \mathcal{H}_i$ and strategy profile $\sigma$, the induced terminal history is a random variable that depends on the realization of the state of nature and opponents’ types profile (agent $i$’s type $b_i$ is known to agent $i$ at $(b_i, h_i^t)$). This is denoted by $g^{\sigma(b_i, h_i^t)}(\theta, b_{-i})$. As done for the belief-free games in section 4.1, we can define strategic-form payoff functions as follows:
\begin{equation}
U_i(\sigma, \theta, b_{-i}; b_i, h_i^t) = u_i \left( g^{\sigma(b_i, h_i^t)}(\theta, b_{-i}), \theta \right).
\end{equation}

**Definition 3** Fix a belief system $p$. A strategy profile is sequentially rational with respect to $p$ if for every $i \in N$ and every $(b_i, h_i^t) \in B_i \times \mathcal{H}_i$, the following inequality is satisfied for every $\sigma_i^t \in \Sigma_i$:
\begin{equation}
\int_{\Theta \times B_{-i}} U_i(\sigma, \theta, b_{-i}; b_i, h_i^t) \cdot dp_i(b_i, h_i^t) \tag{12}
\end{equation}

\begin{equation}
\geq \int_{\Theta \times B_{-i}} U_i(\sigma_i^t, \sigma_{-i}, \theta, b_{-i}; b_i, h_i^t) \cdot dp_i(b_i, h_i^t).
\end{equation}

**Definition 4** An assessment $(\sigma, p)$ is an Interim Perfect Equilibrium (IPE) if:

1. $\sigma$ is sequentially rational with respect to $p$; and

14
2. \( p \) is weakly preconsistent with \( \sigma \).

If inequality (12) is only imposed at private histories of length zero, the solution concept coincides with *interim equilibrium* (Bergemann and Morris, 2005). IPE refines interim equilibrium imposing two natural conditions: first, sequential rationality; second, weak preconsistency of the belief system.

Weak preconsistency imposes no restrictions on the beliefs held at histories that receive zero probability at the preceding node.\(^{12}\) Hence, even if agents’ initial beliefs admit a common prior, IPE is weaker than Fudenberg and Tirole’s (1991) Perfect Bayesian Equilibrium. Also, notice that any player’s deviation is a zero probability event, and treated the same way. In particular, if history \( h^t_i \) is precluded by \( \sigma_i (b_i, h^{t-1}_i) \) alone, then \( P^{s_i, s_j} (b_i, h^t_i) [h^t_i] = 0 \), and agent \( i \)’s beliefs at \( (b_i h^t_i) \) are unrestricted the same way they would be after an unexpected move of the opponents. This feature of IPE is not entirely standard, but it is key to the result that the set of all IPE-strategies (taking the union over all models of beliefs) can be computed by means of a convenient “backwards procedure” (proposition 1). Treating own deviations the same as the opponents’ is key to the possibility of considering continuation games “in isolation”, necessary for that result.\(^ {13} \)

### 6.2 Mechanisms in Belief-Free Environments: \((\mathcal{E}, \mathcal{M})\).

This section introduces a backwards procedure for belief-free dynamic games and shows that it characterizes the set of all IPE-strategies (taking the union over all models of beliefs). The formal definition is notationally cumbersome, but the idea is that illustrated in the example of Section 2.

Fix a public history \( h^{T-1} \) of length \( T - 1 \). For each payoff-type \( y^T_i \in \Theta_i \) of each agent, the continuation game is a static game, to which we can apply belief-free rationalizability (e.g., Bergemann and Morris, 2009a). For each \( h^{T-1} \), let \( R_i (h^{T-1}) \) denote the set of pairs \( (y^T_i, s_i | h^{T-1}) \) such that continuation strategy \( s_i | h^{T-1} \) is rationalizable in the continuation game from \( h^{T-1} \) for type \( y^T_i \). We now proceed backwards: for each public history \( h^{T-2} \) of length \( T - 2 \), we apply again rationalizability to the continuation game from \( h^{T-2} \) (in normal form), restricting continuation strategies \( s_i | h^{T-2} \in S_{h^{T-2}} \) to be rationalizable in the continuation games from histories of length \( h^{T-1} \). \( R_i (h^{T-2}) \) denotes the set of pairs

\(^{12}\)Unlike other notions of weak perfect Bayesian equilibrium, in IPE agents’ beliefs are consistent with Bayesian updating also off-the-equilibrium path. In particular, in complete information games, IPE coincides with subgame-perfect equilibrium. See Penta (2011b) for a thorough discussion of the solution concept.

\(^{13}\)In Penta (2009) I consider a minimal strengthening of IPE, in which agents’ beliefs are not upset by unilateral own deviations, and I show how the analysis adapts to that case. The resulting “backwards procedure” is more complex, essentially undoing the advantages of the *indirect approach* developed below.
\((y_i^{T-1}, s_i|h^{T-2})\) such that continuation strategy \(s_i|h^{T-2}\) is rationalizable in the continuation game from \(h^{T-2}\) for “type” \(y_i^{T-1}\). Inductively, this is done for each \(h^{t-1}\), until the initial node \(\phi\) is reached, for which the set \(R_i^\phi\) is computed.

**The “Backwards Procedure”**. To formally define the backwards procedure, we need extra notation for continuation strategies: For each \(h_i^t\), \(S^{h_i^t}\) denotes the set of player \(i\)'s strategies in the subform starting from \(h_i^t\). For each public history \(h^{t-1}\), let \(S^{h^{t-1}} = \{s^{h^{t-1}} = (y_i^{t-1}, s_i) : s_i \in S^{h^{t-1}}(y_i^{t-1})\}\): an element of \(S^{h^{t-1}}\) is a function assigning to each \(y_i^{t-1} \in Y_i^t\) a continuation strategy \(s_i \in S^{h_i^t}\), where \(h_i^t = (h_i^{t-1}, y_i^{t-1})\).

For each \(s_i \in S_i\) and each \(h^{t-1} \in \mathcal{H}\), \(s_i|h^{t-1} \in S^{h^{t-1}}\) denotes the continuation of \(s_i\) starting from \(h^{t-1}\). The notation \(g^{s|h^{t-1}}(\theta)\) refers to the terminal history induced by strategy profile \(s\) from the public history \(h^{t-1}\) if the realized state of nature is \(\theta\). Strategic-form payoff functions can be defined for continuations from a given public history: for each \(h \in \mathcal{H}\) and each \((s, \theta) \in S \times \Theta\), \(U_i(s, \theta; h) = u_i(g^{s|h}(\theta), \theta)\). For the initial history \(\phi\), it will be written \(U_i(s, \theta)\) instead of \(U_i(s, \theta; \phi)\).

For any \(h_i^t = (h^{t-1}, y_i^{t-1})\) and \(\pi \in \Delta(\Theta \times S^{h^{t-1}})\), let \(\rho(\pi; h_i^t)\) denote the set of continuation strategies from \(h_i^t\), \(s_i \in S^{h_i^t}\), that are best response to conjectures \(\pi\) over the payoff states and the opponents’ continuation strategies. That is:

\[
BR_i(\pi; h_i^t) = \arg\max_{s_i \in S^{h_i^t}} \int_{(s, s_{-i}) \in \Omega \times S^{h^{t-1}}} U_i(s_i, s_{-i}, \theta; h^{t-1}) \cdot d\pi
\]

We can now introduce the backwards procedure formally. The definition is recursive, starting from the last stage and proceeding backwards:

- **[t = T]** For each \(h_i^T = (h^{T-1}, y_i^T)\), let \(R_i^0(h_i^T) = S_i^{h_i^T}\), and for each \(k = 1, 2,\ldots,\) let
  \[
  R_i^k(h^{T-1}) = (y_i^T, s_j) : s_j \in R_j^{k-1}(h^{T-1}, y_j^T)\}
  \]
  \[
  R_i^{k-1}(h^{T-1}) = \times_j R_j^{k-1}(h^{T-1})\}
  \]
  \[
  R_i^k(h^{T-1}, y_i^T) = \left\{s_i \in R_i^{k-1}(h^{T-1}, y_i^T) : \begin{array}{ll}
  (R.1) & \exists \pi \in \Delta(\{y_i^T\} \times \Theta_{-i} \times R_{-i}^{k-1}(h^{T-1})) \\
  (R.2) & s_i \in BR_i(\pi; h^{T-1}, y_i^T)
  \end{array} \right\}
  \]

  Notice that \(R_i(h^{T-1})\) consists of pairs of types \(y_i^T\) and continuation strategies \(s_i \in S_i^{(h^{T-1}, y_i^T)}\), hence \(R_i(h^{T-1}) \subseteq S_i^{h^{T-1}}\).

- **[t = T - 1, ... 0]** For each \(h_i^t = (h^{t-1}, y_i^t)\), let
  \[
  R_i^0(h^{t-1}, y_i^t) = \left\{s_i \in S_i^{h_i^t} : \forall h^t \text{ s.t. } h^{t-1} < h^t, \quad \forall y_i^{t+1} = (y_i^t, \theta_i, t+1), s_i(h^t, y_i^{t+1}) \in R_i(h^t, y_i^{t+1}) \right\}
  \]
and for each $k = 1, 2, \ldots$, let

$$R_{i}^{k-1}(h^{t-1}) = \left\{ (y^t_i, s_i) : s_j \in R_{j}^{k-1}(h^{t-1}, y^t_j) \right\},$$

$$R_{i}^{k-1}(h^{t-1}) = \times_{j \neq i} R_{j}^{k-1}(h^{t-1}),$$

$$R_{i}^{k}(h^{t-1}, y^t_i) = \left\{ s_i \in R_{i}^{k-1}(h^{t-1}, y^t_i) : \begin{array}{l}
(R.1) \exists \pi \in \Delta \left( \{y^t_i\} \times \Theta \times R_{i}^{k-1}(h^{t-1}) \right), \\
(R.2) s_i \in BR_i(\pi, h^{t-1}, y^t_i) \end{array} \right\},$$

$$R_{i}^{k}(h^{t-1}, y^t_i) = \bigcap_{k=1}^{\infty} R_{i}^{k}(h^{t-1}, y^t_i) \quad \text{and} \quad R_{i}(h^{t-1}) = \left\{ (y^t_i, s_i) : s_i \in R_{i}^{k-1}(h^{t-1}, y^t_i) \right\}.$$

Finally: $R_{i}^{0} = \{ s_i \in S_i : s_i | y^1_i \in R_{i}(y^1_i) \}$ for each $y^1_i \in Y^1_i$.

### 6.3 Characterization and Computation of the set of IPE

The following result is central to the analysis of the full implementation problem conducted in Section 8.

**Proposition 1** Fix a game $(\mathcal{E}, \mathcal{M})$. For each $i$: $\hat{s}_i \in R_{i}^{0}$ if and only if $\exists \mathcal{B}, \hat{b}_i \in B_i$ and $(\hat{s}, \hat{p})$ such that: (i) $(\hat{s}, \hat{p})$ is an IPE of $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ and (ii) $\hat{s}_i = \hat{s}_i \left( \hat{b}_i \right)$.

**Proof.** (See Appendix B) ■

The proof of the proposition requires the introduction of several new concepts. In the next subsection I introduce these concepts and I discuss the intuition for the result. The content of Section 6.3.1 is not required for understanding the results of Sections 7 and 8. The uninterested reader may thus skip this section altogether.

### 6.3.1 Sketch of the Proof of Proposition 1

The proof of proposition 1 is based on a solution concept for belief-free games in extensive form, Backwards Extensive Form Rationalizability $(\mathcal{B} \mathcal{R})$. Proposition 1 obtains by first showing that $\mathcal{B} \mathcal{R}$ characterizes the set of IPE strategies (Proposition 2) and then showing that $\mathcal{B} \mathcal{R}$ can be computed via the backwards procedure $R^{0}$ (Proposition 3). In Penta (2011b) I provide epistemic foundations for $\mathcal{B} \mathcal{R}$ and argue that it captures precisely the implications of backward induction reasoning in games with incomplete information. I discuss some of these ideas in the following.

**Backwards Extensive Form Rationalizability.** Similar to rationalizability, $\mathcal{B} \mathcal{R}$ is a non-equilibrium solution concept: no coordination of beliefs on some equilibrium strategy is imposed. Rather, agents form conjectures about everyone’s behavior, which may or may not be consistent with each other. To avoid confusion, we refer to this kind of beliefs as “conjectures”, retaining the term “beliefs” only for those introduced in Section 5.
Agents entertain conjectures about the space $\Theta \times S$. As the game unfolds, and agents observe their private histories, their conjectures change. For each private history $h_t^i = (h_{t-1}^i, y_t^i) \in H_i$, define the event $[h_t^i] \subseteq \Theta \times S$ as:

$$[h_t^i] = \{y_t^i\} \times (\times_{\tau=t+1}^{\infty} \Theta_{i,\tau}) \times \Theta_{-i} \times S(h_{t-1}^i).$$

(Notice that, by definition, $[h_t^i] \subseteq [h_{t-1}^i]$ whenever $h_{t-1}^i \leq h_t^i$.)

**Definition 5** A conjecture for agent $i$ is a conditional probability system (CPS hereafter), that is a collection $\mu^i = (\mu^i(h_t^i))_{h_t^i \in H_i}$ of measures $\mu^i(h_t^i) \in \Delta (\Theta \times S)$ that satisfy the following conditions:

C.1 For all $h_t^i \in H_i$, $\mu^i(h_t^i) \in \Delta ([h_t^i])$;

C.2 For every measurable $A \subseteq [h_t^i] \subseteq [h_{t-1}^i]$, $\mu^i(h_t^i[A] \cdot \mu^i(h_{t-1}^i)[h_t^i] = \mu^i(h_{t-1}^i)[A]$.

The set of agent $i$’s conjectures is denoted by $\Delta^{H_i} (\Theta \times S)$.

Condition C.1 states that agents’ are always certain of what they know; condition C.2 states that agents’ conjectures are consistent with Bayesian updating whenever possible. Notice that in this specification agents entertain conjectures about the payoff state, the opponents’ and their own strategies. This point is discussed below.

Strategy $s_i$ is sequentially rational with respect to conjectures $\mu^i$ if, at each history $h_t^i \in H_i$, it prescribes optimal behavior with respect to $\mu^i(h_t^i)$ in the continuation of the game. Formally: Given a CPS $\mu^i \in \Delta^{H_i} (\Theta \times S)$ and a history $h_t^i = (h_{t-1}^i, y_t^i)$, strategy $s_i$ expected payoff at $h_t^i$, given $\mu^i$, is defined as:

$$U_i(s_i, \mu^i; h_t^i) = \int_{\Theta \times S_{-i}} U_i(s_i, s_{-i}, \theta; h_{t-1}^i) \cdot d\text{marg}_{\Theta \times S_{-i}} \mu^i(h_t^i).$$

(13)

**Definition 6** Strategy $s_i$ is sequentially rational with respect to $\mu^i \in \Delta^{H_i} (\Theta \times S)$, written $s_i \in r_i(\mu^i)$, if and only if for each $h_t^i \in H_i$ and each $s_t^i \in S_i$ the following inequality is satisfied:

$$U_i(s_i, \mu^i; h_t^i) \geq U_i(s_t^i, \mu^i; h_t^i).$$

(14)

If $s_i \in r_i(\mu^i)$, we say that conjectures $\mu^i$ “justify” strategy $s_i$.

We can now introduce Backwards Rationalizability ($\mathcal{BR}$):
Definition 7 For each $i \in N$, let $\mathcal{BR}_i^0 = S_i$. Define recursively, for $k = 1, 2, \ldots$

\[
\mathcal{BR}_i^k = \begin{cases} 
\exists \mu^i \in \Delta^\mathcal{H}_i(\Theta \times S) \text{ s.t.} \\
(1) \hat{s}_i \in r_1(\mu^i) \\
(2) \text{supp}(\mu^i(\phi)) \subseteq \Theta \times \{\hat{s}_i\} \times \mathcal{BR}_i^{k-1} \\
(3) \text{for each } h^*_i = (h^{k-1}_i, y^*_i) \in \mathcal{H}_i:\n\quad s \in \text{supp}(\text{marg}_{\Theta} \mu^i(h^*_i)) \text{ implies:} \\
\quad (3.1) s_i|h^*_i = \hat{s}_i|h^*_i, \text{ and} \\
\quad (3.2) \exists s'_{i-1} \in \mathcal{BR}_{i-1}^{k-1}: s'_{i-1}|h^{t-1} = s_{i-1}|h^{t-1}
\end{cases}
\]

Finally, $\mathcal{BR} := \bigcap_{k \geq 0} \mathcal{BR}_i^k.$

$\mathcal{BR}$ consists of an iterated deletion procedure. At each round, strategy $\hat{s}_i$ survives if it is justified by conjectures $\mu^i$ that satisfy two conditions: condition (2) states that at the beginning of the game, the agent must be certain of his own strategy $\hat{s}_i$ and have conjectures concentrated on opponents’ strategies that survived the previous rounds of deletion; condition (3) restricts the agent’s conjectures at unexpected histories: condition (3.1) states that agent $i$ is always certain of his own continuation strategy; condition (3.2) requires conjectures to be concentrated on opponents’ continuation strategies that are consistent with the previous rounds of deletion. However, at unexpected histories, agents’ conjectures about $\Theta$ are essentially unrestricted. Thus, condition (3) embeds two conceptually distinct kinds of assumptions: the first concerning agents’ conjectures about $\Theta$; the second concerning their conjectures about the continuation behavior. For ease of reference, they are summarized as follows:

- **Unrestricted-Inference Assumption (UIA):** At unexpected histories, agents’ conjectures about $\Theta$ are essentially unrestricted. In particular, agents are free to infer anything about the opponents’ private information (or their own future signals) from the public history.

For example, conditional conjectures may be such that $\text{marg}_{\Theta} \mu^i(h^*_i)$ is concentrated on a “type” $y'_{i-1}$ for which some of the previous moves in $h^{t-1}$ are irrational. Nonetheless, condition (3.2) implies that it is believed that $y'_{i-1}$ will behave rationally in the future. From an epistemic viewpoint, it can be shown that $\mathcal{BR}$ can be interpreted as common certainty of future rationality at every history.

\^14It goes without saying that whenever we write a condition like $\mu^i(X|h^*_i) \geq \gamma$ and $X$ is not measurable, the condition is not satisfied.
• **Common Certainty in Future Rationality (CCFR):** at every history (expected or not), agents share common certainty in future rationality.\footnote{See Penta (2011b) for a formal characterization. Penta (2011b) also discuss how CCFR can be interpreted as a condition of belief persistence on the continuation strategies. In games of complete information, an instance of the same principle is provided by subgame perfection, where agents believe in the equilibrium continuation strategies both on- and off-the-equilibrium path. The belief persistence hypothesis goes hand in hand with the logic of backward induction, allowing to envision each subgame “in isolation” (see also footnote 16.)}

We introduce next the first result, showing that $BR$ characterizes the set of IPE for all models of beliefs. As emphasized above, in $BR$ agents hold conjectures about both the opponents’ and their own strategies. First, notice that conditions (2) and (3.2) in the definition of $BR$ maintain that agents are always certain of their own strategy; furthermore, the definition of sequential best response (def. 6) depends only on the marginals of the conditional conjectures over $\Theta \times S_{-i}$. Hence, this particular feature of $BR$ does not affect the standard notion of rationality. The fact that conjectures are elements of $\Delta^H_i (\Theta \times S)$ rather than $\Delta^H_i (\Theta \times S_{-i})$ corresponds to the assumption, discussed in Section 6.1, that IPE treats all deviations the same. Its implication is that both histories arising from unexpected moves of the opponents and from one’s own deviations represent zero-probability events, allowing the same set of conditional beliefs about $S_i$, with essentially the same freedom that IPE allows after anyone’s deviation. This is the main insight behind the following result (the proof is in Appendix B):

**Proposition 2** Fix a game $(E, M)$. For each $i$: $\hat{s}_i \in BR_i$ if and only if $\exists B, \hat{b}_i \in B_i$ and $(\hat{s}, \hat{p})$ such that: (i) $(\hat{s}, \hat{p})$ is an IPE of $(E, M, B)$ and (ii) $\hat{s}_i = \hat{\sigma}_i (\hat{b}_i)$.

Next we show that $BR$ can be computed via the backwards procedure $R^\phi$ introduced above. That is:

**Proposition 3 (Computation)** $BR_i = R_i^\phi$ for each $i$.

The proof of the proposition can be found in Appendix B. To grasp the fundamental insights, notice that an implication of this proposition is that $BR$ (hence the set of IPE) has a recursive structure analogous to that of subgame perfect equilibrium in games of complete information. That is, $BR$ can be computed analyzing continuation games in isolation. To see why this is the case, consider properties UIA and CCFR. First, notice that under UIA, the set of beliefs agents are allowed to entertain about the opponents’ payoff types (i.e. the support of their marginal beliefs over $\Theta_{-i}$) is the same at every history (equal to $\Theta_{-i}$). Hence, in this respect, their information about the opponents’
types in the subform starting from (public) history \( h_{t-1} \) is the same as if the game started from \( h_t \). Also, CCFR implies that agents’ epistemic assumptions about everyone’s behavior in the continuation is also the same at every history. Thus, UIA and CCFR combined imply that, from the point of view of \( BR \), a continuation from history \( h_{t-1} \) is equivalent to a game with the same space of uncertainty and strategy spaces equal to the continuation strategies, which justifies the possibility of analyzing continuation games “in isolation”. We call this property of \( BR \) “continuation-game consistency”, because it generalizes an analogous property of subgame perfect equilibrium, which is key to the possibility of solving the game backwards.\(^{16}\)

Hence, taken together, Propositions 2 and 3 (i.e., proposition 1) imply that the set of IPE is continuation-game consistent. That is: for each \( h \), the set of IPE strategies in the continuation game from \( h \) coincides with the set of IPE strategies of the continuation game considered in isolation.

7 Partial Implementation.

The notion of (partial) implementation adopted by the classical literature on static mechanism design is that of interim incentive compatibility:

**Definition 8** SCF \( f \) is interim incentive compatible on \( \mathcal{B} = \langle N, \Theta, (B_i, \beta_i)_{i \in N} \rangle \) if truthful revelation is an interim equilibrium of \((\mathcal{E}, \mathcal{M}(f), \mathcal{B})\). That is, \( \exists \sigma^* \in \Sigma^* \) such that for each \( i \in N \) and \( b_i \in B_i \), for all \( \sigma_i \in \Sigma_i \),

\[
\int_{\Theta \times B_{-i}} U_i(\sigma^*, \theta, b_{-i}; b_i) \cdot d\beta_i(b_i) \\
\geq \int_{\Theta \times B_{-i}} U_i(\sigma_i, \sigma^*_{-i}; \theta, b_{-i}; b_i) \cdot d\beta_i(b_i).
\]

(Recall that \( \Sigma^* \) denotes the set of truth-telling strategies.) Bergemann and Morris (2005) showed that a SCF is interim incentive compatible on all type spaces, if and only if it is ex-post incentive compatible, that is:

**Definition 9** SCF \( f \) is ex post incentive compatible if for each \( i \), for each \( \theta \in \Theta \) and \( s'_i \in S_i \)

\[
U_i(s^*, \theta) \geq U_i(s'_i, s^*_{-i}; \theta).
\]

\(^{16}\)Harsanyi and Selten’s (1988) notion of “subgame consistency” provides one of the most compelling arguments in favor of subgame perfect equilibrium:

‘It is natural to require that a solution function for extensive form games is subgame consistent in the sense that the behavior prescribed on a subgame is nothing else than the solution of the subgame itself’ (Harsanyi and Selten, ibid., p.90).
We say that a SCF is Strictly Ex-Post Incentive Compatible if for any \( s' \notin S^*_i \), the inequality holds strictly.

Interim incentive compatibility imposes no requirement of perfection: If players cannot commit to their strategies, more stringent incentive compatibility requirements must be introduced, to account for the dynamic structure of the problem. We thus apply the solution concept introduced in Section 6.1, IPE: A mechanism is interim perfect incentive compatible if the truth­telling strategy is an IPE of the direct mechanism.

**Definition 10** SCF \( f \) is interim perfect incentive compatible on \( B = \langle N, \Theta, (B_i, \beta_i)_{i \in N} \rangle \) if there exist beliefs \( (p^i)_{i \in N} \) and \( \sigma^* \in \Sigma^* \) such that, \( (\sigma^*, p) \) is an IPE of \( (E, M(f), B) \).

For a given model of beliefs, interim perfect incentive compatibility is clearly more demanding than interim incentive compatibility. But, as the next result shows, the requirement of “perfection” is no more demanding than the “ex-ante” incentive compatibility if it is required for all models of beliefs:

**Proposition 4 (Partial Implementation)** SCF \( f \) is interim perfect incentive compatible on all models of beliefs if and only if it is ex post incentive compatible.

**Proof.** (See Appendix C.1)

Hence, as far as “robust” partial implementation is concerned, assuming that agents can commit to their strategies is without loss of generality: The dynamic mechanism can be analyzed in its normal form.

### 8 Full Implementation in Direct Mechanisms.

This section provides sufficient conditions for full robust implementation in direct mechanisms. In general, focusing on direct mechanisms is not without loss of generality for full implementation problems. This is why the classical literature on Bayesian Implementation typically adopts complex “augmented mechanisms”, in which agents report more than their own type.\(^\text{17}\) Compared to such augmented mechanisms, direct mechanisms have the advantage of being simple, which is an important desideratum from the viewpoint of the Wilson doctrine. Section 8.4 considers mechanisms with slightly richer message spaces, which allow to simplify the analysis yet avoiding the intricacies of the classical augmented mechanisms. The restrictiveness of direct mechanisms is further discussed in Section 9.

\(^{17}\) The classical reference is Maskin (1999) or, for Bayesian settings, Postlewaite and Schmeidler (1988), Palfrey and Srivastava (1989) and Jackson (1991).
Definition 11 SCF \( f \) is fully perfectly implementable in the direct mechanism if for every \( B \), every IPE-strategy profile \( \sigma^* \) of the Bayesian game \((\mathcal{E}, \mathcal{M}(f), B)\) is such that \( \sigma^*(b) \in S^* \) for all \( b \in B \).

The following proposition follows immediately from proposition 1.

Proposition 5 SCF \( f \) is (fully) robustly perfect-implementable in the direct mechanism if and only if \( R^s \subseteq S^* \).

8.1 Environments with Monotone Aggregators of Information.

In this Section it is maintained that each set \( \Theta_{i,t} = [\theta_{i,t}, \theta_{i,t}^h] \subseteq \mathbb{R} \), so that, for each \( t = 1, ..., T \), \( Y_t \subseteq \mathbb{R}^{nt} \). Environments with monotone aggregators are characterized by the property that for each agent, in each period, all the available information (across time and agents) can be summarized by a one-dimensional statistic. Furthermore, such \( T \) statistics (one per period) uniquely determine an agent’s preferences. (This notion generalizes properties of preferences discussed in the example in Section 2).

Definition 12 An Environment admits monotone aggregators (EMA) if, for each \( i \), and for each \( t = 1, ..., T \), there exists an aggregator function \( \alpha^t_i : Y^t \rightarrow \mathbb{R} \) and a valuation function \( v_i : \Xi \times \mathbb{R}^T \rightarrow \mathbb{R} \) that satisfy the following conditions:

1. for each \( (\xi^*, \theta^*) \in \Xi \times \Theta \), \( u_i(\xi^*, \theta^*) = v_i \left( \xi^*, (\alpha^t_i(y^\tau(\theta^*)))_{\tau=1}^T \right) \).

2. \( \alpha^t_i \) and \( v_i \) are continuous and \( \alpha^t_i \) is strictly increasing in \( \theta_{i,t} \);

3. For any \( y^t_i, \hat{y}^t_i \in Y^t_i \), if \( \alpha^t_i(y^t_i, y^t_{-i}) > \alpha^t_i(\hat{y}^t_i, y^t_{-i}) \) for some \( y^t_{-i} \in Y^t_{-i} \), then \( \alpha^t_i(y^t_i, \hat{y}^t_{-i}) \geq \alpha^t_i(\hat{y}^t_i, \hat{y}^t_{-i}) \) for any \( \hat{y}^t_{-i} \in Y^t_{-i} \).

Assuming the existence of the aggregators and the valuation functions (condition 1), per se, entails little loss of generality. The bite of the representation derives mainly from the continuity and monotonicity conditions (2), and from condition (3) which guarantees that \( i \)'s private histories of signals \( Y^t_i \) can be (weakly) ordered in terms of the induced values of the period-\( t \) aggregator \( \alpha^t_i \).

---

18 Recall that \( S^* \) denote the set of strategy profiles that are outcome equivalent to the truthful strategy profiles.

19 The notation \( y^\tau(\theta^*) \) refers to the realized sequence of signals from period 1 through \( \tau \) if the realized state is \( \theta^* \). Formally: For any \( \theta = (\theta^1, ..., \theta^T) \in \Theta \) and \( t = 1, ..., T \), \( y^t(\theta) = (\theta^1, ..., \theta^t) \).
The self-correcting strategy. The analysis is based on the notion of self-correcting strategy, $s^c$, which generalizes what we have already described in the leading example of Section 2: at each period-$t$ history, $s^c_t$ reports a message such that the implied period-$t$ valuation is “as correct as it can be”, given the previous reports. That is: conditional on past truthful revelation, $s^c_t$ truthfully reports $i$’s period-$t$ signal; at histories that come after previous misreports of agent $i$, $s^c_t$ entails a further misreport, to offset the impact on the period-$t$ aggregator of the previous misreports. Formally:

**Definition 13** The self-correcting strategy, $s^c_t \in S_i$, is such that for each $t = 1, ..., T$ and public history $h^{t-1} = (\tilde{y}^{t-1}, x^{t-1})$, and for each private history $h^t_i = (h^{t-1}, y^t_i)$,

$$s^c_t(h^t_i) = \arg \min_{m_i, t \in \Theta_{i,t}} \max_{y^t_i, \in Y^t_i} \left\{ \max_{y^t_i, \in Y^t_i} \left| \alpha^t_i(y^t_i, y^t_{-i}) - \alpha^t_i(\tilde{y}^{t-1}, m_i, y^t_{-i}) \right| \right\}.$$  

Clearly, $s^c$ induces truthful reporting (that is: $s^c \in S^*$): if $h^{t-1} = (\tilde{y}^{t-1}, x^{t-1})$ and $y^t_i = (\tilde{y}^{t-1}, \theta_{i,t})$, then $s^c_t(h^{t-1}, y^t_i) = \theta_{i,t}$. Also, notice that $s^c_t(h^t_i)$ only depends on the component of the public history made of $i$’s own reports, $\tilde{y}^{t-1}_i$. Let $\tilde{y}^{t-1}_{-i}$ be such that:

$$\tilde{y}^{t-1}_{-i} \in \arg \max_{y^t_{-i}} \left| \alpha^t_i(y^t_i, y^t_{-i}) - \alpha^t_i(\tilde{y}^{t-1}, s^c_t(h^t_i), y^t_{-i}) \right|.$$  

Then, the definition of $s^c_t$ and the properties of $\alpha^t_i$ (def. 12) imply that three cases are possible:

$$\alpha^t_i(y^t_i, y^t_{-i}) = \alpha^t_i(\tilde{y}^{t-1}, s^c_t(h^t_i), y^t_{-i}) \text{ for all } y^t_{-i} \in Y^t_{-i},$$

$$\alpha^t_i(y^t_i, \tilde{y}^{t-1}_{-i}) > \alpha^t_i(\tilde{y}^{t-1}, s^c_t(h^t_i), \tilde{y}^{t-1}_{-i}) \text{ and } s^c_t(h^t_i) = \theta^h_{i,t},$$

$$\alpha^t_i(y^t_i, \tilde{y}^{t-1}_{-i}) < \alpha^t_i(\tilde{y}^{t-1}, s^c_t(h^t_i), \tilde{y}^{t-1}_{-i}) \text{ and } s^c_t(h^t_i) = \theta^l_{i,t}.$$  

Equation (17) corresponds to the case in which strategy $s^c_t$ can completely offset the previous misreports. But there may exist histories at which no current report can offset the previous misreports. In the example of Section 2, suppose that the first period under- (resp. over-) report is so low (resp. high), that even reporting the highest (lowest) possible message in the second period is not enough to “correct” the implied value of $\varphi$. This was the case corresponding to the possibility of corner solutions, and corresponds cases to (18) and (19) respectively.

---

20 An earlier formulation of the idea of self-correcting strategy can be found in Pavan (2007). I thank Alessandro Pavan for pointing this out.
The Contraction Property. The results on full implementation are based on a \textit{contraction property} that limits the dependence of agents’ aggregator functions on the private signals of the opponents. Before formally introducing the contraction property, some extra notation is needed: for each set of strategy profiles $D = \times_{j \in N} D_j \subseteq S$ and for each private history $h^t_i$, let

$$D_i(h^t_i) := \{m_{i,t} : \exists s_i \in D_i, \text{s.t. } s_i(h^t_i) = m_{i,t}\}$$

and

$$D_i(h^{t-1}) := \bigcup_{y^t_i \in Y^t_i} D_i(h^{t-1}, y^t_i).$$

Define also:

$$s_i[D_i(h^{t-1})] := \{(m_{i,t}, y^t_i) \in M_{i,t} \times Y^t_i : m_{i,t} \in D_i(h^{t-1}, y^t_i)\}$$

and

$$s_i^c[h^{t-1}] := \{(m_{i,t}, y^t_i) \in M_{i,t} \times Y^t_i : m_{i,t} = s_i^c(h^{t-1}, y^t_i)\}$$

\textbf{Definition 14 (Contraction Property)} An environment with monotone aggregators of information satisfies the Contraction Property if, for each $D \subseteq S$ such that $D \neq \{s^c\}$ and for each $h^{t-1} = (y^{t-1}, x^{t-1})$ such that $s[D(h^{t-1})] \neq s^c[h^{t-1}]$, there exists $y^t_i$ and $m^t_{i,t} \in D_i(h^{t-1}, y^t_i)$, $m^t_{i,t} \neq s^c_i(h^{t-1}, y^t_i)$, such that:

$$\text{sign} \left[ s_i^c(h^{t-1}, y^t_i) - \theta^t_{i,t} \right] = \text{sign} \left[ \alpha_i^t(y^t_i, y^t_{-i}) - \alpha_i^t(y^{t-1}, \theta^t_{i,t}, \theta^t_{-i,t}) \right]$$

(20)

for all $y^t_{-i} = (y^{t-1}_{-i}, \theta_{-i,t}) \in Y^t_{-i}$ and $m^t_{-i,t} \in D_{-i}(h^{t-1}, y^t_{-i}).$

To interpret the condition, rewrite the argument of the right-hand side of (20) as follows:

$$\alpha_i^t(y^t_i, y^t_{-i}) - \alpha_i^t(y^{t-1}, m^t_{i,t}, m^t_{-i,t})$$

$$= \left[ \alpha_i^t(y^{t-1}, s_i^c(h^{t-1}, y^t_i), y^t_{-i}) - \alpha_i^t(y^{t-1}, m^t_{i,t}, m^t_{-i,t}) \right]$$

$$+ \kappa_i(h^{t-1}, y^t_i, y^t_{-i})$$

(21)

where

$$\kappa_i(h^{t-1}, y^t_i, y^t_{-i}) = \alpha_i^t(y^t_i, y^t_{-i}) - \alpha_i^t(y^{t-1}, s_i^c(h^{t-1}, y^t_i), y^t_{-i})$$

(22)

The term in square brackets in (21) represents the impact, on the period-$t$ aggregator, of a deviation (in the set $D$) from the self-correcting strategy $s^c_i$ at profile at history $(h^{t-1}, y^t_i)$; the term $\kappa_i(h^{t-1}, y^t_i, y^t_{-i})$ represents the extent by which the self-correcting strategy is incapable of offsetting the previous misreports. Suppose that $\kappa_i(h^{t-1}, y^t_i, y^t_{-i}) = 0$, i.e.
strategy $s_c$ fully offsets the previous misreports (in particular, this is the case if $h^{t-1}$ is a truthful history: $\tilde{y}^{t-1} = y^{t-1}$), then, the contraction property boils down to the following, more directly comparable to Bergemann and Morris’ (2009) static counterpart:

\[(Simple\ CP)\ For\ each\ public\ history\ at\ which\ the\ behavior\ allowed\ by\ the\ set\ of\ deviations\ $D$\ is\ different\ from\ $s_c$,\ there\ exists\ at\ least\ one\ player’s\ “type”\ $y^t_i$\ of\ some\ agent\ $i$,\ for\ which\ for\ some\ $m'_{i,t} \in D_i(h^{t-1}, y^t_i)$, $\alpha^t_i(y^t_i, y^t_{-i}) - \alpha^t_i(\tilde{y}^{t-1}, m'_{i,t}, m'_{-i,t})$\ is\ unilaterally\ signed\ by\ $[s_c(h^{t-1}, y^t_i) - m'_{i,t}]$,\ uniformly\ over\ the\ opponents\ private\ information\ and\ current\ reports.\]

From equations (17)-(19) it is easy to see that $\kappa_i(h^{t-1}, y^t_i, y^t_{-i}) = 0$ whenever $s_c(h^t_i) \in (\theta^t_i, \theta^{h_i}_i)$. Hence, this corresponds precisely to the case considered in the example of Section 2. To account for the possibility that, at some histories, the self-correcting strategy is not sufficient to offset the previous misreports, the simple CP is modified so that the sign of the impact on the aggregator $\alpha^t_i$, of deviations from $s_c$ at $h^{t-1}$, is not offset by the previous misreports, $\kappa$. In Section 8.4 it will be shown how this complexity may be avoided by adopting simple “enlarged” mechanisms, in which agents’ sets of messages are extended at every period so that any possible past misreport can be “corrected”, inducing $\kappa_i(h^{t-1}, y^t_i, y^t_{-i}) = 0$ at all histories. Given the simplicity of their structure, such mechanisms will be called “quasi-direct”.

8.2 Aggregator-Based SCF.

Consider the SCF in the example of Section 2 (equations 2-5): the allocation chosen by the SCF in period $t$, is only a function of the values of the aggregators in period $t$. The notion of aggregator-based SCF generalizes this idea:

**Definition 15** The SCF $f = (f_t)_{t=1}^T$ is aggregator-based if for each $t$, $\alpha^t_i(y^t) = \alpha^t_i(\tilde{y}^t)$ for all $i$ implies $f_t(y^t) = f_t(\tilde{y}^t)$.

The next proposition shows that, if the contraction property is satisfied, an aggregator-based SCF is fully implementable in environments that satisfy a single-crossing condition:

**Definition 16 (SCC-1)** An environment with monotone aggregators of information satisfies SCC-1 if, for each $i$, valuation function $v_i$, is such that: for each $t$, and $\xi, \xi' \in \Xi : \xi_\tau = \xi'_\tau$ for all $\tau \neq t$, then for each $\alpha^*_{i,-t} \in \mathbb{R}^{T-1}$ and for each $\alpha_{i,t} < \alpha'_{i,t} < \alpha''_{i,t}$,

$$v_i(\xi, \alpha_{i,t}, \alpha^*_{i,-t}) > v_i(\xi', \alpha_{i,t}, \alpha^*_{i,-t}) \text{ and } v_i(\xi, \alpha'_{i,t}, \alpha^*_{i,-t}) = v_i(\xi', \alpha'_{i,t}, \alpha^*_{i,-t}) \text{ implies } v_i(\xi, \alpha''_{i,t}, \alpha^*_{i,-t}) < v_i(\xi', \alpha''_{i,t}, \alpha^*_{i,-t})$$

26
In words: For any two allocations \(\xi\) and \(\xi'\) that only differ in their period-\(t\) component, for any \(\alpha_{i,t}^{*} \in \mathbb{R}^{T-1}\), the difference \(\delta_{i,t}(\xi, \xi', \alpha_{i,t}) = v_i(\xi, \alpha_{i,t}, \alpha_{i,t}^{*}) - v_i(\xi', \alpha_{i,t}, \alpha_{i,t}^{*})\) as a function of \(\alpha_{i,t}\) crosses zero (at most) once (see figure 1.a, p. 29). We are now in the position to present the first full-implementation result:

**Proposition 6** In an environment with monotone aggregators (def. 12) satisfying SCC-1 (def. 16) and the contraction property (def. 14), if an aggregator-based social choice function satisfies Strict EPIC (definition 9), then \(\mathcal{BR} = \{s^c\}\).

**Proof.** (See Appendix C.2)

The argument of the proof is analogous to the argument presented in Section 2: For each history of length \(T - 1\), it is proved that the contraction property and SCC-1 imply that agents play according to \(s^c\) in the last stage; then the argument proceeds by induction: given that in periods \(t + 1, \ldots, T\) agents follow \(s^c\), a misreport at period \(t\) only affects the period-\(t\) aggregator (because the SCF is “aggregator-based”). Then, SCC-1 and the contraction property imply that the self-correcting strategy is followed at stage \(t\).

**An Appraisal of the “aggregator-based” assumption.** Consider the important special case of time-separable preferences: Suppose that, for each \(i\) and \(t = 1, \ldots, T\), there exist an “aggregator” function \(\alpha_i^t : Y^t \rightarrow \mathbb{R}\) and a valuation function \(v_i^t : \Xi^t \times \mathbb{R} \rightarrow \mathbb{R}\) such that for each \((\xi^*, \theta^*) \in \Xi \times \Theta\),

\[
u_i(\xi^*, \theta^*) = \sum_{t=1}^{T} v_i^t(\xi_i^*, \alpha_i^t(y^t(\theta^*)))
\]

In this case, the condition that the SCF is aggregator-based (def. 15) can be interpreted as saying that the SCF only responds to changes in preferences: If two distinct payoff states \(\theta\) and \(\theta'\) induce the same preferences over the period-\(t\) allocations, then the SCF chooses the same allocation under \(\theta\) and \(\theta'\) in period \(t\). This is the case of the example in Section 2. These preferences though cannot accommodate phenomena of “path-dependence” such as habit formation or learning-by-doing. For instance, in the context of the example of Section 2, suppose that agents’ preferences are the following:

\[
u_i(q_1, q_2, \pi_{i,1}, \pi_{i,2}, \theta) = \alpha_{i,1}(\theta_1) \cdot q_1 + \pi_{i,1} + [\alpha_{i,2}(\theta_1, \theta_2) \cdot F(q_1) \cdot q_2 + \pi_{j,2}].
\]

\(21\)These preferences are “time separable” in the sense that the marginal utility of period \(t\) allocation \(\xi_t\) does not depend on the allocations chosen in other periods allocations. Nonetheless, it may depend on previous periods’ information.

\(22\)In that example the set of allocations includes the transfers, hence for each \(t\) the social choice function is: \(f_i^{t}(\theta) = (q_{i,t}^{t}(\theta), \pi_{i,t}^{t}(\theta), \pi_{j,t}^{t}(\theta))\). The first component is clearly aggregator-based (see equations 2 and 3); Furthermore, if \(\gamma \in [0, 1]\), the values of the aggregators uniquely determine the size of the transfers (equations 4 and 5). The social choice function is thus “aggregator-based”.

27
The marginal utility of \( q_2 \) now also depends on the amount of public good provided in the first period. Then, the optimal policy for the second period is to set \( q_2^*(\theta) = [\alpha_{i,2}(\theta) + \alpha_{j,2}(\theta)] \cdot F(q_1) \). This rule is not aggregator-based, as the period-2 allocation choice depends on both the period-2 aggregators and the previous period allocation. Thus, to allow the SCF to respond to possible “path-dependencies” in agents’ preferences it is necessary to relax the “aggregator-based” assumption.

In environments with transferable utility (such as the example above) our notion of SCF includes the specification of the transfers scheme: \( f_t(\theta) = (q_t(\theta), \pi_{i,t}(\theta), \pi_{j,t}(\theta)) \) for each \( t \). Since the requirement that the SCF is aggregator-based applies to all its components, it also applies to transfers. In general, it is desirable to allow for arbitrary transfers, not necessarily aggregator-based. The general results of the next section can be easily adapted to accommodate the possibility of arbitrary transfers in environments with transferable utility (Section 8.3.1).

8.3 Relaxing “Aggregator-Based”

In the proof of proposition 6, the problem with relaxing the assumption that the SCF is “aggregator-based” is that a one-shot deviation from \( s^c \) at period-\( t \) may induce different allocations in period-\( t \) and in subsequent periods. Hence, the “within period” single-crossing condition (SCC-1) may not suffice to conclude the inductive step, and guarantee that strategy \( s^c \) is played at period-\( t \): Some bound is needed on the impact that a one-shot deviation has on the outcome of the SCF. The next condition guarantees that the impact of a one-shot deviation is not too strong.

**Definition 17 (SCC-2)** An environment with monotone aggregators of information satisfies SCC-2 if, for each \( i \): for each \( \theta, \theta' \in \Theta \) such that \( \exists t \in \{1, ..., T\} : y^\tau(\theta) = y^\tau(\theta') \) for all \( \tau < t \) and for all \( j \), \( \alpha_j^\tau(\theta) = \alpha_j^\tau(\theta') \) for all \( \tau > t \), then for each \( a_\cdot^\cdot, a_\cdot^\cdot -1 \in \mathbb{R}^{T-1} \) and for each \( \alpha_{i,t} < \alpha'_{i,t} < \alpha''_{i,t} \),

\[
v_i(f(\theta), \alpha_{i,t}, a_{i,-t}) > v_i(f(\theta'), \alpha_{i,t}, a_{i,-t}) \quad \text{and} \quad v_i(f(\theta), \alpha'_{i,t}, a_{i,-t}) = v_i(f(\theta'), \alpha'_{i,t}, a_{i,-t}) \implies v_i(f(\theta'), \alpha''_{i,t}, a_{i,-t}) < v_i(f(\theta'), \alpha''_{i,t}, a_{i,-t})
\]

SCC-2 compares the allocations chosen for any two “similar” states of nature: states \( \theta \) and \( \theta' \) are “similar” in the sense that they are identical up to period \( t - 1 \), and imply the same value of the aggregators at all periods other than \( t \). Since agents’ preferences are uniquely determined by the values of the aggregators (definition 12), the preferences induced by states \( \theta \) and \( \theta' \) only differ along the dimension of the period-\( t \) aggregator. The condition requires a single-crossing condition for the corresponding outcomes to hold along this direction.

28
From a graphical viewpoint, this condition can be interpreted as follows: suppose that $\theta$ and $\theta'$ are as in definition 17. Then, if the SCF is “aggregator-based” and the environment satisfies SCC-1 (definition 16), the difference in payoffs for $f(\theta)$ and $f(\theta')$ as a function of the period-$t$ aggregator crosses zero (at most) once (figure 1.a). If $f$ is not “aggregator based”, allocations at periods $\tau > t$ may differ under $f(\theta)$ and $f(\theta')$, shifting (or changing the shape) of the curve $\delta_{i,t}(f(\theta), f(\theta'), \alpha_{i,t})$. SCC-2 guarantees that such shifting maintains the single-crossing property (figure 1.b).

The economic intuition of SCC-2 is also straightforward. In static environments, single-crossing conditions allow agents’ types to monotonically sort themselves with respect to the incentive compatibility conditions. This monotonicity guarantees the possibility of achieving implementation. By requiring a single-crossing condition to hold within each period, condition SCC-1 (def. 16) is the natural extension of the basic idea to a multi-period setting. With “aggregator based” SCF, such “period by period” single-crossing condition suffices to achieve full implementation in the dynamic mechanism (Proposition 6). When the “aggregator based” assumption is relaxed, the fact that types at time $t$ are sorted relative to the period-$t$ incentives does not guarantee that the required monotonicity also holds when forward looking, intertemporal considerations are taken into account. SCC-2 requires precisely such “intertemporal effects” to be sufficiently well-behaved that they do not upset the single-crossing property at time $t$.

Within the special case of time-separable preferences, SCC-2 is indeed quite permissive. For instance, the preferences in equation (23) satisfy SCC-2 for any choice of $F: \mathbb{R} \to \mathbb{R}$.

**Proposition 7 (Full Implementation)** In an environment with monotone aggregators (def. 12) satisfying the contraction property (def. 14), if a SCF $f$ is Strictly EPIC (definition 9) and satisfies SCC-2 (def. 17), then $\mathcal{B} \mathcal{R} = \{ s^c \}$.

**Proof.** (See Appendix C.3) \[\blacksquare\]
Corollary 1  Since $s^c \in S^*$, if the assumptions of propositions 6 or 7 are satisfied, $f$ is fully robustly implementable by the direct mechanism.

8.3.1 Transferable Utility.

A special case of interest is that of additively separable preferences with transferable utility: For each $t = 1, \ldots, T$, the space of allocations is $\Xi_t = Q_t \times (\times_{i=1}^n \Pi_{i,t})$, where $Q_t$ is the set of “common components” of the allocation and $\Pi_{i,t} \subseteq \mathbb{R}$ is the set of transfers to agent $i$ (i’s “private component”). Maintaining the restriction that the environment admits monotone aggregators, agent $i$’s preferences are as follows: For each $\xi^* = (q_t, \pi_{1,t}, \ldots, \pi_{n,t})_{t=1}^T \in \Xi$ and $\theta^* \in \Theta$,

$$u_i(\xi^*, \theta^*) = \sum_{t=1}^T v^t_i ((q^t_{\tau=1}, \alpha^t_i (y^t(\theta^*)))) + \pi_{i,t},$$

where for each $t = 1, \ldots, T$, $v^t_i : (\times_{\tau=1}^t Q_{\tau}) \times \mathbb{R} \rightarrow \mathbb{R}$ is the period-$t$ valuation of the common component. Notice that functions $v^t_i : (\times_{\tau=1}^t Q_{\tau}) \times \mathbb{R} \rightarrow \mathbb{R}$ are defined over the entire history $(q_1, \ldots, q_t)$: this allows period-$t$ valuation of the current allocation $(q_t)$ to depend on the previous allocative decisions $(q_1, \ldots, q_{t-1})$. This allows us to accommodate the “path dependencies” in preferences discussed above.\(^{23}\)

In environments with transferable utility, it is common to define a social choice rule for the common component, $\chi_t : Y^t \rightarrow Q_t$ ($t = 1, \ldots, T$), while transfer schemes $\pi_{i,t} : Y^t \rightarrow \mathbb{R}$ ($i = 1, \ldots, n$ and $t = 1, \ldots, T$) are specified as part of the mechanism. Not assuming transferable utility, social choice functions above were defined over the entire allocation space ($f_t : Y^t \rightarrow \Xi_t$), they thus include transfers in the case of transferable utility. The transition from one approach to the other is straightforward. Any given pair of choice rule and transfer scheme $(\chi_t, (\pi_{i,t})_{t=1}^n)_{t=1}^T$ trivially induces a social choice function $f^\chi_{\pi} : Y^t \rightarrow \Xi_t$ ($t = 1, \ldots, T$) in the setup above: for each $t$ and $y^t \in Y^t$, $f^\chi_{\pi}(y_t) = (\chi_t(y^t), (\pi_{i,t}(y^t)))_{i=1}^n$.

It is easy to check that, in environments with transferable utility, if agents’ preferences over the common component $Q^* = \times_{t=1}^T Q_t$ satisfy (SCC-1), and $\chi : \Theta \rightarrow Q$ is aggregator-based, then for any transfer scheme $(\pi_{i,t})_{i=1}^n$, the “full” social choice function $f^\chi_{\pi}$ satisfies (SCC-2). More generally, if $\chi$ and agents’ preferences over $Q^*$ satisfy (SCC-2), then $f^\chi_{\pi}$ satisfies (SCC-2) for any transfer scheme $(\pi_{i,t})_{i=1}^n$.

Given this, the following corollary of proposition 7 is immediate:

Corollary 2  In environments with monotone aggregators of information and transferable utility, if agents’ preferences over $Q^*$ and $\chi : \Theta \rightarrow Q^*$ satisfy: (i) the contraction property;\(^{23}\)The special case of “path-independent” preferences corresponding to the example in section 2 is such that period-$t$ valuation are functions $v^t_i : Q_t \times \mathbb{R} \rightarrow \mathbb{R}$.  

30
(ii) the single crossing condition (SCC-2); and (iii) there exist transfers \( \pi \) that make \( \chi \) strictly ex-post incentive compatible. Then: \( f^{x,\pi} \) is fully robustly implemented.

8.4 "Quasi-direct" Mechanisms.

This section shows how simple "enlarged" mechanisms may avoid incurring into the problem of corner solutions, which allows us to modify the contraction property (definition 14) by guaranteeing that the sign condition holds with \( \kappa_i (h^t-1,y^t) = 0 \) at every history (equation 22). This way, the dynamic contraction property is more directly comparable with Bergemann and Morris’ (2009) static counterpart.

Let \( \hat{\alpha}_{i,t} : \mathbb{R}^{nt} \to \mathbb{R} \) be a continuous extension of \( \alpha_{i,t} : Y^t \to \mathbb{R} \) from \( Y^t \) to \( \mathbb{R}^{nt} \), strictly increasing in the component that extends \( \theta_{i,t} \) and constant in all the others on \( \mathbb{R}^{nt} \setminus Y^t \) (recall that from definition 12 \( \alpha_{i,t} \) is only assumed strictly increasing in \( \theta_{i,t} \) on \( Y^t \)). Set \( m_{i,1}^+ = \theta_{i,1}^h \) and \( m_{i,1}^- = \theta_{i,1}^b \), and for each \( t = 1, \ldots, T \), let \( \hat{\Theta}_{i,t} = [m_{i,t}^-, m_{i,t}^+] \), and \( \hat{Y}_i^t = \times_{t=1}^T \hat{\Theta}_{i,t} \) where \( m_{i,t}^- \) and \( m_{i,t}^+ \) are recursively defined so to satisfy:

\[
\begin{align*}
  m_{i,t}^+ &= \max \left\{ m_i \in \mathbb{R} : \max_{(y_i^t,y_i^t'-i)} \left| \hat{\alpha}_{i,t} \left( y_i^t, y_i^t-i \right) - \min_{\hat{y}_i^{t-1} \in \hat{Y}_i^{t-1}} \hat{\alpha}_{i,t} \left( \hat{y}_i^{t-1}, m_i, y_i^t \right) \right| = 0 \right\} \\
  m_{i,t}^- &= \min \left\{ m_i \in \mathbb{R} : \max_{(y_i^t,y_i^t'-i)} \left| \hat{\alpha}_{i,t} \left( y_i^t, y_i^t-i \right) - \max_{\hat{y}_i^{t-1} \in \hat{Y}_i^{t-1}} \hat{\alpha}_{i,t} \left( \hat{y}_i^{t-1}, m_i, y_i^t \right) \right| = 0 \right\}
\end{align*}
\]

Set the message spaces in the mechanism such that \( M_{i,t} = \hat{\Theta}_{i,t} \) for each \( i \) and \( t \). By construction, for any private history \( h_i^t = (h_i^{t-1}, y_i^t) \), the self-correcting report \( s_i^c (h_i^t) \) satisfies equation (17), that is \( s^c \) is capable of fully offsetting previous misreports: messages in \( \hat{\Theta}_{i,t} \setminus \Theta_{i,t} \) are used whenever equations (18) or (19) would be the case in the direct mechanism. (Clearly, such messages never arise if \( s^c \) is played.) To complete the mechanism, we need to extend the domain of the outcome function to account for these "extra" messages. Such extension consists of treating these reports in terms of the implied value of the aggregator: For given sequence of reports \( \hat{y}_i^t \in \hat{Y}_i^t \) such that some message in \( \hat{\Theta}_{i,t} \setminus \Theta_{i,t} \) has been reported at some period \( \tau \leq t \), let \( g_t (\hat{y}_i^t) = f_t (\theta) \) for some \( \theta \) such that \( \alpha_{i,\tau} (\theta) = \alpha_{i,\tau} (\hat{y}_i^\tau) \) for all \( i \) and \( \tau \leq t \), \( f_t (\theta) = f_t (\theta') \).

9 Concluding Remarks.

On the Solution Concepts. IPE is an equilibrium notion. That is, it maintains that agents have correct conjectures about each other’s strategies. "Robust Implementation" was defined with respect to such an equilibrium concept, and required implementation in IPE for all models of beliefs. As an implication of proposition 1, "Robust Implementation" could equivalently be defined in terms of backward rationalizability, which is a non
equilibrium solution concept. In Penta (2011b) I provide an analysis of the epistemic
underpinnings of the solution concepts adopted in this paper, and I argue that $R^\ominus$ (hence,
the set of IPE for all models of beliefs) characterizes the predictions of *backward induction
reasoning* in games with incomplete information. Two alternative epistemic characteri-
zations are provided, in terms of “common belief in future rationality”, and in terms of
“common belief in rationality and in belief persistence.”

**Dynamic Mechanisms in Static Environments.** Consider an environment in which
agents obtain all the relevant information before the planner has to make a decision. Al-
though static mechanisms may be a viable option, the designer may still have reasons
to adopt a dynamic mechanism (e.g. an ascending auction). In an environment with
complete information, Bergemann and Morris (2007) recently argued that dynamic mech-
anisms may improve on static ones by reducing agents’ strategic uncertainty: They showed
how the backward induction outcome of a second-price ascending auction guarantees full
robust implementation of the efficient allocation for a larger set of parameters than the
rationalizable outcomes of a second price sealed-bid auction. The approach of this paper
allows us to extend the analysis to incomplete information settings. It can be shown
that, with incomplete information, the ascending auction does not improve on the static
one (see Penta, 2011b). The reason is that the logic of backward induction loses its bite
when the assumption of complete information is relaxed, and no restrictions on beliefs are
imposed (see discussion in pp. 20-16). In incomplete information environments, the case
for dynamic mechanisms in reducing strategic uncertainty must rely on stronger epistemic
assumptions, that allow agents to draw stronger inferences on their opponents’ private
information from their past moves. Mueller (2011) obtains some results in this sense,
under the assumption of Common Strong Belief in Rationality (CSBR). CSBR is a strong
epistemic assumption, implying sophisticated forward induction reasoning (see Battigalli
and Siniscalchi, 2002 and 2007).

**Non-Direct Mechanisms.** In general, the restriction to direct mechanisms is not with-
out loss of generality for full implementation problems. This remains true here, in that
allowing richer message spaces may facilitate full implementation. Nonetheless, the results
in Section 8 provide sufficient conditions for full implementation in direct mechanisms,
which have the advantage of being simple, an important desideratum from the viewpoint
of the *Wilson doctrine*.

---

24 For the special case of games with complete information, Perea (2011) independently introduced a
procedure that is equivalent to $R^\ominus$. Besides considering games with complete information only, Perea’s
(2011) epistemic characterization differs from Penta’s (2011b) in that Perea does not maintain that agents
are Bayesian.

25 In the present setting, this amounts to assuming $|\Theta_t| = 1$ for all $t > 1$ and $|\Xi_t| = 1$ for all $t < T$. 

32
In the direct mechanisms of definition 1, messages are public. This may appear unnecessarily restrictive at first, because concealing past reports may allow the designer to implement more SCFs. However, this is not the case in the present setting. The reason is the same why, in a static setting, the ascending clock-auction does not improve on its static counterpart: when no restrictions on beliefs are imposed, the logic of backward induction loses its bite, because players cannot “robustly” infer anything about the opponents’ private information from the public histories (see Penta, 2011b). When some restrictions on beliefs are maintained, or under stronger epistemic assumptions (e.g., based on forward induction reasoning), observability of past messages would be important.

 Extensions and Developments. This paper is only a first step towards the developing of a theory of robust dynamic mechanism design. Much remains to be explored. In the following I discuss some possible extensions for future research.

 The results in Section 6 provide foundations to a tractable solution concept that characterizes the robust predictions in a general class of dynamic games. Furthermore, as argued in Penta (2011b), there is a precise sense in which this solution concept captures the logic of backward induction reasoning in games with incomplete information. Backward rationalizability can thus be applied to other settings. Its tractability can prove particularly useful in applied work.

 The belief-free approach to mechanism design is often criticized for imposing an unnecessarily demanding notion of robustness. Yet, belief-free settings are important theoretical benchmarks, useful for developing the necessary conceptual and analytical tools. An important direction for future research, still largely unexplored, is to consider intermediate notions of robustness, where some but not all assumptions on beliefs are relaxed. In static settings, modified versions of rationalizability can be used to this purpose (e.g., Artemov et al., 2011). Similarly, it would be important to extend the methodology developed in this paper identifying the variations of backward rationalizability that allow to study problems of robust dynamic mechanism design when some assumptions on agents’ beliefs are maintained.

 This paper puts forward a “backward induction” approach to robust dynamic mechanism design. While backward induction is analytically convenient and a natural starting point, it would be important to study problems of robust dynamic mechanism design based on a forward induction logic. Under stronger restrictions on the solution concepts, stronger implementation results can be obtained.
Appendix

A Topological structures and Conditional Probability Systems.

A.1 Topological structures.

Sets $\Theta_{i,t} \subseteq \mathbb{R}^{n_i,t}$, $\Xi_t \subseteq \mathbb{R}^{l_t}$ and $M_{i,t} \subseteq \mathbb{R}^{\nu_{i,t}}$ are non-empty and compact, for each $i$ and $t$ (Sections 3 and 4). Let $n_t = \sum_{i \in N} n_{i,t}$ and $\nu_t = \sum_{i \in N} \nu_{i,t}$. For each $h_i^t = \tau < t$, let $\alpha_\tau (h_i^t)$ denote the triple $\left( \theta_{i,\tau}, m_\tau, \xi_\tau \right)$ consisting of $i$'s private signal at period $\tau$, the message profile and allocation chosen at stage $\tau$ along history $h_i^t$. For each $k \in \mathbb{N}$, let $d(k)$ denote the Euclidean metric on $\mathbb{R}^k$. We endow the sets $\mathcal{H}_i$ with the following metrics, $d^i(i \in N)$, defined as: For each $h_i^t, h_i^t \in \mathcal{H}_i$ (w.l.o.g.: let $\tau \geq t$)

$$d^i(h_i^t, h_i^t) = \sum_{k=1}^{t-1} d(n_{i,k} + \nu_{i,k} + l_k) \left( \alpha_k \left( h_i^t \right), \alpha_k \left( h_i^t \right) \right) + d_{n_{i,t}} \left( \theta_{i,t}, \theta_{i,t} \right) + \sum_{k=t+1}^{\tau} 1.$$  

It can be checked that $(\mathcal{H}_i, d^i)$ are complete, separable metric spaces.

Sets of strategies are endowed with the supmetrics $d_{S_i}$ defined as:

$$d_{S_i} (s_i, s_i') = \sum_{t=1}^{T} \left( \sup_{h_i^t \in H^{t-1} \times \mathcal{Y}_i^t} d_{n_{i,t}} \left( s_i \left( h_i^t \right), s_i' \left( h_i^t \right) \right) \right)$$

Under these topological structures, the following lemma implies that CPSs introduced in Section A.2 are well-defined.

**Lemma 1** For all public histories $h \in \mathcal{H}$, $S_i(h)$ is closed.  

**Proof.** See lemma 2.1 in Battigalli (2003).  

A.2 Conditional Probability Systems

Let $\Omega$ be a metric space and $\mathcal{A}$ its Borel sigma-algebra. Fix a non-empty collection of subsets $\mathcal{C} \subseteq \mathcal{A} \setminus \emptyset$, to be interpreted as “relevant hypothesis”. A conditional probability system (CPS hereafter) on $(\Omega, \mathcal{A}, \mathcal{C})$ is a mapping $\mu : \mathcal{A} \times \mathcal{C} \to [0, 1]$ such that:

**Axiom 1** For all $B \in \mathcal{C}$, $\mu \left( B \right) \left[ B \right] = 1$  

**Axiom 2** For all $B \in \mathcal{C}$, $\mu \left( B \right)$ is a probability measure on $(\Omega, \mathcal{A})$.  

**Axiom 3** For all $A \in \mathcal{A}$, $B, C \in \mathcal{C}$, if $A \subseteq B \subseteq C$ then $\mu \left( B \right) \left[ A \right] \cdot \mu \left( C \right) \left[ B \right] = \mu \left( C \right) \left[ A \right]$. 

34
The set of CPS on \((\Omega, \mathcal{A}, \mathcal{C})\), denoted by \(\Delta^C(\Omega)\), can be seen as a subset of \([\Delta(\Omega)]^C\) (i.e. mappings from \(\mathcal{C}\) to probability measures over \((\Omega, \mathcal{A})\)). CPS’s will be written as \(\mu = (\mu(B))_{B \in \mathcal{C}} \in \Delta^C(\Omega)\). The subsets of \(\Omega\) in \(\mathcal{C}\) are the conditioning events, each inducing beliefs over \(\Omega\); \(\Delta(\Omega)\) is endowed with the topology of weak convergence of measures and \([\Delta(\Omega)]^C\) is endowed with the product topology. Below, for each player \(i\), we will set \(\Omega = \Theta \times S\) in games with payoff uncertainty (or \(\Omega = \Theta \times \Sigma\) if the game is appended with a model of beliefs). The set of conditioning events is naturally provided by the set of private histories \(\mathcal{H}_i\): for each private history \(h^i_t = (h^{i-1}, y^i_t) \in \mathcal{H}_i\), the corresponding event \([h^i_t]\) is defined as:

\[
[h^i_t] = \{y^i_t\} \times (\times_{\tau=t+1}^T \Theta_{i, \tau}) \times \Theta_{-i} \times S(h^{i-1}) .
\]

Under the maintained assumptions and topological structures, sets \([h^i_t]\) are compact for each \(h^i_t\), thus \(\Delta^{\mathcal{H}_i}(\Omega)\) is a well-defined space of conditional probability systems. With a slight abuse of notation, we will write \(\mu^i(h^i_t)\) instead of \(\mu^i([h^i_t])\)

**B Proof Proposition 1**

**Proposition 2.** Fix a game \((\mathcal{E}, \mathcal{M})\). For each \(i\): \(s^i \in \mathcal{BR}_i\) if and only if \(\exists B, b^i \in B_i\) and \((\hat{s}, \hat{p})\) such that: (i) \((\hat{s}, \hat{p})\) is an IPE of \((\mathcal{E}, \mathcal{M}, B)\) and (ii) \(s^i = \hat{s}_i(b^i)\).

**Proof:**

**Step 1: \((\Leftarrow)\).** Fix \(B, (\hat{s}, \hat{p})\) and \(b^i\). For each \(h^i_t \in \mathcal{H}_i\), let \(P_{\hat{s}, \hat{p}}(b^i_t, h^i_t) \in \Delta(\Theta \times B_{-i} \times S_{-i})\) denote the probability measure on \(\Theta \times B_{-i} \times S_{-i}\) induced by \(\hat{p}_i(b^i_t, h^i_t)\) and \(\hat{s}_{-i}\). For each \(j\), let

\[
\bar{S}_j = \{s^j \in S_j : \exists b^j \in B_j \text{ s.t. } s^j = \hat{s}_j(b^j)\},
\]

and for each \(h^i_t\),

\[
\bar{S}^h_j = \{s^j_h \in S^h_j : \exists s^j \in \bar{S}_j \text{ s.t. } s^j | h = s^j_h\}
\]

We will prove that \(\bar{S}_j \subseteq \mathcal{BR}_j\) for every \(j\). For each \(h^i_t = (y^i_t, h^{i-1}) \in \mathcal{H}_i\), let \(\varphi_{j}^{h^i_t} : \bar{S}_j \rightarrow S_j(h^i_t)\) be a measurable function such that

\[
\varphi_{j}^{h^i_t}(s^j) = \begin{cases} 
  s^j(h^i_t) & \text{if } \tau \geq t \\
  m^j & \text{otherwise}
\end{cases}
\]

where \(m^j\) is the message (action) played by \(j\) at period \(\tau < t\) in the public history \(h^{i-1}\). Thus, \(\varphi_{j}^{h^i_t}\) transforms any strategy in \(\bar{S}_j\) into one that has the same continuation from \(h^i_t\), and that agrees with \(h^i_t\) for the previous periods. Define the mapping \(L_{h^i_t} : \Theta \times B_{-i} \times S_{-i} \rightarrow \Theta \times B \times S\) such that

\[
L_{h^i_t}(\theta, b_{-i}, s_{-i}) = (\theta, \hat{b^i_t}, b_{-i}, \varphi_{j}^{h^i_t}(\hat{s}_i(b^i_t)), \varphi_{-i}^{h^i_t}(s_{-i}))
\]

35
(In particular, \( L_\phi (\theta, b_{-i}, s_{-i}) = (\theta, \hat{b}_i, b_{-i}, \hat{\sigma}_i (\hat{b}_i), s_{-i}) \).)

Define the CPS \( \lambda_i \in \Delta^{H_i} (\Theta \times B \times S) \) such that, for any measurable \( E \subseteq \Theta \times B \times S \),

\[
\lambda_i (\phi) [E] = P_i (\hat{\sigma}, \hat{\nu}) \left( \hat{b}_i, \phi \right) \left[ L_\phi^{-1} (E) \right]
\]

and for all \( h_i^t \in \mathcal{H}_i \) s.t. \( \lambda_i (h_i^{t-1}) [h_i^t] = 0 \), let

\[
\lambda_i (h_i^t) [E] = P_i (\hat{\sigma}, \hat{\nu}) \left( \hat{b}_i, h_i^t \right) \left[ L_{h_i^t}^{-1} (E) \right].
\]

(Conditional beliefs \( \lambda_i (h_i^t) \) at histories \( h_i^t \) s.t. \( \lambda_i (h_i^{t-1}) [h_i^t] > 0 \) are determined via Bayesian updating, from the definition of CPS. See appendix A.2)

Define the CPS \( \mu^i \in \Delta^{H_i} (\Theta \times S) \) s.t. \( \forall h_i^t \in \mathcal{H}_i, \mu^i (h_i^t) = \text{marg}_{\Theta \times S} \lambda_i (h_i^t) \). By construction, \( \hat{s}_i \in r_i (\mu^i) \). We only need to show that conditions (2) and (3) in the definition of \( \mathcal{BR} \) are satisfied by \( \mu^i \). This part proceeds by induction: The initial step, for \( k = 1 \), is trivial. Hence, \( \tilde{S}_j \subseteq \mathcal{BR}^j \) for every \( j \). To complete the proof, let (as inductive hypothesis) \( \tilde{S}_j \subseteq \mathcal{BR}_j^k \) for every \( j \). Then \( \mu^i \) constructed above satisfies \( \mu^i (\phi) \subseteq \Theta \times \{ \hat{s}_i \} \times \mathcal{BR}_i^{-1} \) and

\[
\begin{align*}
supp (\text{marg}_{S^{t-1}} \mu^i (\phi)) \\
= supp (\text{marg}_{S^{t-1}} \mu^i (h_i^t)) \\
\subseteq \tilde{S}^{t-1}
\end{align*}
\]

thus \( \hat{s}_i \in \mathcal{BR}_i^{k+1} \). This concludes the first part of the proof.

**Step 2:** \((\Rightarrow)\). Let \( \mathcal{B} \) be such that for each \( i, B_i = \mathcal{BR}_i \) and let strategy \( \hat{\sigma}_i : B_i \to S_i \) be the identity map. Define the map \( M_{i,\phi} : \Theta \times S \to \Theta \times B_{-i} \) s.t.

\[
M_{i,\phi} (\theta, s_i, s_{-i}) = (\theta, \hat{\sigma}_i^{-1} (s_{-i}))
\]

Notice that, for each \( i \) and \( s_i \in \mathcal{BR}_i \), \( \exists \mu^{s_i} \in \Delta^{H_i} (\Theta \times S) \) s.t.

1. \( \hat{s}_i \in r_i (\mu^{s_i}) \)
2. for all \( h_i^t \in \mathcal{H}_i \): \( s_j \in supp (\text{marg}_{S_j} \mu^{s_i} (h_i^t)) \)

\( \Rightarrow \exists s'_j \in \mathcal{BR}_j : s_j | h_i^{t-1} = s'_j | h_i^{t-1} \).

Hence, for each \( h_i^t \neq \phi \), we can define the map \( \rho_{s_i, h_i^t} : supp (\text{marg}_{S_{-i}} \mu^{s_i} (h_i^t)) \to \mathcal{BR}_{-i} \) that satisfies \( \rho_{s_i, h_i^t} (s_{-i}) | h_i^{t-1} = s_{-i} | h_i^{t-1} \). Let \( m_{s_i, h_i^t} \equiv \hat{\sigma}_i^{-1} \circ \rho_{s_i, h_i^t} \). Define maps \( M_{\hat{s}_i, h_i^t} : \Theta \times supp (\mu^{s_i} (h_i^t)) \to \Theta \times B_{-i} \)

\[
M_{\hat{s}_i, h_i^t} (\theta, s_i, s_{-i}) = (\theta, m_{h_i^t} (s_{-i}))
\]

Let beliefs \( \beta_i : B_i \to \Delta (\Theta \times B_{-i}) \) be s.t. for every measurable \( E \subseteq \Theta \times B_{-i} \)

\[
\beta_i (b_i) [E] = \mu^{\hat{\nu} (b_i)} (\phi) \left[ M_{i,\phi}^{-1} (E) \right]
\]

36
Let beliefs $\hat{\beta}_i$ be derived from $\hat{\sigma}$ and the initial beliefs $\beta_i$ via Bayesian updating whenever possible. At all other histories $h^t_i \in \mathcal{H}_i$, for every measurable $E \subseteq \Theta \times B$, set

$$\hat{\beta}_i (h^t_i) [E] = \mu^{\beta_i(h^t_i)}(h^t_i) \left[ \frac{M_{\hat{\sigma}_i(h^t_i), \hat{h}^t_i}(E)}{M_{\beta_i(h^t_i), h^t_i}(E)} \right].$$

By construction, $(\hat{\sigma}, \hat{\beta})$ is an IPE of $(\mathcal{E}, \mathcal{M}, \mathcal{B})$.

**Proposition 3.** $\mathcal{BR}_i = R^\phi_i$ for each $i$.

**Proof:**

**Step 1** ($R^\phi_i \subseteq \mathcal{BR}_i$): let $s_i \in R^\phi_i$. Then, for each $h^t_i = (h^{t-1}, y^t_i)$, $s_i|h^t_i \in R_i (h^{t-1}, y^t_i)$ (equivalently: $s^t_i \in R_i (h^{t-1})$). Notice that for each $h^{t-1}$ and $s^t_i \in R_i (h^{t-1})$, there exists $s_i \in R^\phi_i$ such that $s_i|_{h^{t-1}} = s^t_i$. Thus, for each $j$ and $h^{t-1}$, we can define measurable functions $\beta_j^{h^{t-1}} : R_j (h^{t-1}) \to R^\phi_j$ such that: $\forall s^t_j \in R_j (h^{t-1})$,

$$\beta_j^{h^{t-1}} \left( s^t_j \right) |_{h^{t-1}} = s^t_i.$$

(Function $\beta_j^{h^{t-1}}$ assign to strategies in $R_j (h^{t-1})$, strategies in $R^\phi_j$ with the same continuation from $h^{t-1}$.) As usual, denote by $\beta_j^{h^{t-1}}$ the product $\times_{j \neq i} \beta_j^{h^{t-1}}$.

For each $h^{t-1}$, let $\varphi_j^{h^{t-1}} : S_j \to S_j (h^{t-1})$ be a measurable function such that

$$\varphi_j^{h^{t-1}} (s_j) (h^t_j) = \begin{cases} s_j (h^t_j) & \text{if } \tau > t \\ m^\tau_j & \text{otherwise} \end{cases}$$

where $m^\tau_j$ is the message (action) played by $j$ at period $\tau < t$ in the public history $h^{t-1}$. (As usual, denote by $\varphi_j^{h^{t-1}}$ the product $\times_{j \neq i} \varphi_j^{h^{t-1}}$.)

For each $h^{t-1}$, define the measurable mapping $\varphi^{h^{t-1}}_{-i} : R_{-i} (h^{t-1}) \to S_{-i} (h^{t-1})$ such that $\forall s^t_{-i} \in R_{-i} (h^{t-1})$,

$$\varphi^{h^{t-1}}_{-i} \left( s^t_{-i} \right) = \varphi^{h^{t-1}}_{-i} \circ \beta^{h^{t-1}}_{-i} \left( s^t_{-i} \right).$$

It will be shown that: for each $k = 0, 1, \ldots$, $s_i \in R^k_i (\phi)$ implies $s_i \in \mathcal{BR}_i$.

The initial step is trivially satisfied ($\mathcal{BR}^0_i = S_i = R^0_i (\phi)$).

For the inductive step, suppose that the statement is true for $n = 0, \ldots, k - 1$: Since $s_i \in R^k_i (\phi)$, for each $h^t_i = (h^{t-1}, y^t_i)$ there exists $\pi^{h^t_i} \in \Delta \left( \Theta \times S_{-i}^{h^{t-1}} \right)$ that satisfies

$$s_i|_{h^t_i} \in \operatorname{arg \ max} \int_{\Theta \times S_{-i}^{h^{t-1}}} U_i (s^t_i, s_{-i}, \theta; h^{t-1}) \cdot d\pi^{h^t_i},$$

and such that $\pi^\phi (\Theta \times R^{k-1}_{-i} (\phi)) = 1$ and for all $h^t_i \neq \phi$,

$$\pi^{h^t_i} \left( \{y^t_i\} \times (\times_{\tau=t+1}^T \Theta_{i, \tau}) \times \Theta_{-i} \times R_{-i} (h^{t-1}) \right) = 1.$$
Now, consider the CPS \( \mu^i \in \Delta^{\mathcal{H}_i} (\Theta \times S) \) such that, for all measurable \( E \subseteq \Theta \times S_{-i} \),

\[
\mu^i (\phi) [\{ \hat{s}_i \} \times E] = \pi^\phi (E).
\]

By definition of CPS, \( \mu^i (\phi) \) defines \( \mu (h^t_i) \) for all \( h^t_i \) s.t. \( \mu^i (\phi) [h^t_i] > 0 \). Let \( h^t_i \) be such that \( \mu^i (\phi) [h^{t-1}_i] > 0 \) and \( \mu^i (\phi) [h^t_i] = 0 \). Define the measurable mapping \( M_{h^t_i} : \Theta \times R^{h^{t-1}_i} \to \Theta \times S (h^{t-1}) \) such that for all \( (\theta, s^{h^{t-1}_i}) \in \Theta \times S (h^{t-1}) \),

\[
M_{h^t_i} (\theta, s^{h^{t-1}_i}) = (\theta, \varphi^{h^{t-1}_i}_i (\hat{s}_i), \varphi^{h^{t-1}_i}_j (s^{h^{t-1}_i}_j))
\]

and set \( \mu^i (h^t_i) \) equal to the pushforward of \( \pi^{h^t_i} \) under \( M_{h^t_i} \), i.e. such that for every measurable \( E \subseteq \Theta \times S \)

\[
\mu^i (h^t_i) [E] = \pi^{h^t_i} [M_{h^t_i}^{-1} (E)].
\]

Again, by definition of CPS, \( \mu^i (h^t_i) \) defines \( \mu^i (h^{t}_j) \) for all \( h^t_j \) that receive positive probability under \( \mu^i (h^t_i) \). For other histories, proceeds as above, setting \( \mu^i (h^t_j) \) equal to the pushforward of \( \pi^{h^t_j} \) under \( M_{h^t_j} \), and so on.

By construction, \( \hat{s}_i \in r_i (\mu^i) \) (condition 1 in the definition of \( \mathcal{BR}^k_i \)). Since by construction \( \mu^i (\phi) [\Theta \times \{ \hat{s}_i \} \times R^{h^{t-1}_i} (\phi)] = 1 \), under the inductive hypothesis \( \mu^i (\phi) [\Theta \times \{ \hat{s}_i \} \times \mathcal{BR}^{k-1}_i] = 1 \) (condition 2 in the definition of \( \mathcal{BR}^k_i \)). From the definition of \( \varphi^{h^{t-1}_i}_i (\hat{s}_i) \), CPS \( \mu^i \) satisfies condition (3.1) at each \( h^t_i \). From the definition of \( \varphi^{h^t_i}_j \), under the inductive hypothesis, \( \mu^i \) satisfies condition (3.2).

**Step 2 (\( \mathcal{BR}_i \subseteq R^k_i \))**: let \( \hat{s}_i \in R^k_i \) and \( \mu^i \in \Delta^{\mathcal{H}_i} (\Theta \times S) \) be such that \( \hat{s}_i \in r_i (\mu^i) \). For each \( h^t_i = (h^{t-1}_i, y^t_i) \), define the mapping \( \psi^{h^t_i}_i : S_{-i} \to S^{h^t_i}_{-i} \) s.t. \( \psi^{h^t_i}_i (s_{-i}) |h^{t-1}_{-i} = s_{-i}|h^{t-1} \) for all \( s_{-i} \in S_{-i} \). (Function \( \psi^{h^t_i}_i \) transforms each strategy profile of the opponents into its continuation from \( h^{t-1} \).) Define also \( \Psi^{h^t_i}_i : \Theta \times S \to \Theta \times S^{h^t_i}_{-i} \) such that

\[
\Psi^{h^t_i}_i (\theta, s_i, s_{-i}) = (\theta, \psi^{h^t_i}_i (s_{-i}))
\]

For each \( h^t_i \in \mathcal{H}_i \), let \( \pi^{h^t_i} \in \Delta \left( \Theta \times S^{h^t_i}_{-i} \right) \) be such that for every measurable \( E \subseteq \Theta \times S^{h^t_i}_{-i} \)

\[
\pi^{h^t_i} [E] = \mu^i (h^t_i) \left[ \Psi^{h^t_i}_i^{-1} (E) \right].
\]

so that the implied joint distribution over payoff states and continuation strategies \( s_{-i}|h^{t-1} \) is the same under \( \mu^i (\cdot, h^t_i) \) and \( \pi^{h^t_i} \). We will show that \( \hat{s}_i |h^t_i \in R_i (h^{t-1}, y^t_i) \) for each \( h^t_i = (h^{t-1}, y^t_i) \). Notice that, by construction,

\[
\hat{s}_i |h^t_i \in \arg \max_{\hat{s}_i \in \mathcal{S}_i ^{h^t_i}} \int U_i (s_i, s_{-i}; h^t_i) \cdot d\pi^{h^t_i}.
\]
The argument proceeds by induction on the length of histories.

**Initial Step** \((T - 1)\). Fix history \(h^T\) \(= (h^{T-1}, y^T)\): for each \(k\), if \(\hat{s}^k \in \mathcal{BR}^k_i\), then \(\hat{s}^k| h^T \in \mathcal{R}^k_i(h^{T-1}, y^T)\). For \(k = 0\), it is trivial. For the inductive step, let \(\pi^{h_T}\) be defined as above: under the inductive hypothesis, \(\pi^{h_T}(\Theta_i \times R_{-i}^{k-1}(h^{T-1})) = 1\) (condition 1), while \(\hat{s}^k \in r_i(\mu^i)\) implies that condition (2) is satisfied.

**Inductive Step:** suppose that for each \(\tau = t + 1, \ldots, T\), \(\hat{s}^k \in \mathcal{BR}^k_i\), implies \(\hat{s}^k| h^\tau \in \mathcal{R}^k_i(h^{\tau-1}, y^\tau)\) for each \(h^\tau = (h^{\tau-1}, y^\tau)\). We will show that for each \(k\), \(h^T_i = (h^{T-1}, y^T_i)\), \(\hat{s}^k_i| h^T_i \in \mathcal{R}^k_i(h^{T-1}, y^T_i)\). We proceed by induction on \(k\): under the inductive hypothesis on \(\tau\), \(\hat{s}^k_i| h^\tau_i \in \mathcal{R}^0_i(h^{\tau-1}, y^\tau)\). For the inductive step on \(k\), suppose that \(\hat{s}^k_i \in \mathcal{BR}^k_i\), implies \(\hat{s}^k_i| h^T_i \in \mathcal{R}^k_i(h^{T-1}, y^T_i)\) for \(n = 0, \ldots, k - 1\), and suppose (as contrapositive) that \(\hat{s}^k_i| h^T_i \notin \mathcal{R}^k_i(h^{T-1}, y^T_i)\). Then, for \(\pi^{h^T_i}\) defined as above, it must be that \(\text{supp}(\pi^{h^T_i}) \notin \Theta \times R_{-i}^{k-1}(h^{T-1})\), which, under the inductive hypothesis on \(n\), implies that \(\exists s_{-i} \in \text{supp}(\text{marg}_{s_{-i}} \mu^i(h^T_i))\) s.t. \(\forall s'_{-i} \in \mathcal{BR}_{-i}: s'_{-i}| h^{T-1} = s_{-i}| h^{T-1}\), which contradicts that \(\mu^i\) justifies \(\hat{s}^k_i\) in \(\mathcal{BR}^k_i\).

Proposition 1 follows from propositions 2 and 3.

**C Proofs of results from Sections 7 and 8.**

**C.1 Proof of Proposition 4**

**Step 1 (If):** For the if part, suppose that \(f\) is ex-post incentive compatible, and let \(B\) be an arbitrary type space \(\mathcal{B}\). It will be shown that there exists an IPE \((p, \hat{\sigma}^*)\) of \((\mathcal{E}, \mathcal{M}(f), \mathcal{B})\) such that \(\hat{\sigma}^* \in \hat{\Sigma}^*\). Fix some arbitrary truthtelling strategy profile \(\sigma^* \in \hat{\Sigma}^*\), and let beliefs system \((p^i)_{i \in \mathcal{N}}\) be such that: (A) \(\forall i \in \mathcal{N}, \forall b_i \in B_i, p^i(b_i, \phi) = \beta_i(b_i)\) (cf. condition (9) in def. 2); (B) for each \(h^i \in \mathcal{H}_i\) such that \(P^{\sigma^*}(b_i, h_i^{T-1})[h^i_1] > 0\), \(p^i(b_i, h^i_1)\) is obtained via Bayesian updating; (C) for histories \(h^i\) such that \(P^{\sigma^*}(b_i, h_i^{T-1})[h^i_1] = 0\), let \(h^i_1 = (h^{T-1}, y^T_i)\) be s.t. \(h^{T-1} = (y^T_i, y^{T-1}_i)\), then let conditional beliefs be such that

\[
\text{supp}\left(\text{marg}_{t \neq i} p^i \left( b_i, h^i_1 \right) \right) \subseteq \{ y^T_i \} \times \left( \times_{\tau = t+1}^{T_i} \Theta_{-i, \tau} \right)
\]

That is, at unexpected histories, each \(i\) believes that the opponents have not deviated from the truthtelling strategy: If “unexpected reports” were observed, player \(i\) rather revises his beliefs about the opponents’ types, not their behavior.

Notice that if \(U_i(s^i, \theta) \geq U_i(s_{-i}^i, s_{-i}^*, \theta)\) for all \(\theta\) (cf. def. 9), then for any \(\pi^i \in \Delta(\Theta \times B_{-i})\),

\[
\int_{\Theta \times B_{-i}} U_i(s^i, \theta, b_{-i}; b_i, \phi) \cdot d\pi^i \geq \int_{\Theta \times B_{-i}} U_i(s_{-i}^*, \theta, b_{-i}; b_i, \phi) \cdot d\pi^i.
\]

Hence, the incentive compatibility constraints are certainly satisfied at the beginning of the game, i.e. for \(\pi^i = p^i(b_i, \phi)\), and so at all histories reached with positive probability
according to the initial conjectures and strategy profile. At zero probability histories, the belief system satisfies (24) by construction, which implies that for every \( h_t^i \) consistent with player \( i \)'s truthtelling strategy,

\[
\int_{\Theta \times B_{-i}} U_i (\sigma^*, \theta, b_{-i}; b_i, h_t^i) \cdot dp^i (b_i, h_t^i) = \int_{\Theta \times B_{-i}} U_i (\sigma^*, \theta, b_{-i}; b_i, \phi) \cdot dp^i (b_i, h_t^i) \\
\int_{\Theta \times B_{-i}} U_i (s_t^i, \sigma^*_{-i}, \theta, b_{-i}; h_t^i) \cdot dp^i (b_i, h_t^i) = \int_{\Theta \times B_{-i}} U_i (s_t^i, \sigma^*_{-i}, \theta, b_{-i}; b_i, \phi) \cdot dp^i (b_i, h_t^i).
\]

But then, equation (25) implies that for all histories \( h_t^i \) consistent with \( i \)'s truthtelling, we have

\[
\int_{\Theta \times B_{-i}} U_i (\sigma^*, \theta, b_{-i}; b_i, h_t^i) \cdot dp^i (b_i, h_t^i) \geq \int_{\Theta \times B_{-i}} U_i (s_t^i, \sigma^*_{-i}, \theta, b_{-i}; b_i, h_t^i) \cdot dp^i (b_i, h_t^i).
\]

Now, let \( \hat{\sigma}_t^* (b_i, h_t^i) = \sigma_t^* (b_i, h_t^i) \) for all histories consistent with \( i \)'s truthtelling; at all histories \( h_t^i \), choose \( \hat{\sigma}_t^* (h_t^i) \) so that it is a best response to \( \sigma^*_{-i} \), given beliefs \( p^i (b_i, h_t^i) \). Since \( \hat{\sigma}_t^* \) only differs from \( \sigma_t^* \) at non-truthful histories, \( \hat{\sigma}_t^* \) is a truthful strategy as well. Also, since beliefs \( p^i \) satisfy (24), only opponents’ truthful histories receive positive probability under \( p^i (b_i, h_t^i) \), hence the behavior of \( \sigma^*_{-i} \) at non-truthful histories is irrelevant from agent \( i \)'s viewpoint. For this reason, we can replace \( \sigma^*_{-i} \) with similarly constructed \( \hat{\sigma}^*_{-i} \), and still have that \( \hat{\sigma}_t^* \) is a sequential best response to \( \hat{\sigma}^*_{-i} \), given \( p^i \). By construction \((p, \hat{\sigma}^*)\) is an IPE of the Bayesian game.

**Step 2 (only if):** Since perfect implementability implies interim implementability, the “only if” immediately follows the results by Bergemann and Morris (2005), who showed that if a SCF is interim implementable on all type spaces, then it is ex-post implementable.■

### C.2 Proof of Proposition 6.

By contradiction, suppose \( \mathcal{BR} = D \neq \{s^c\} \). By continuity of \( u_i \) and compactness of \( \Theta \), \( D(h_t^t) \) is compact for each \( h_t^t \). (Because if \( D = \mathcal{BR} \), strategies in \( D \) must be best responses to conjectures concentrated on \( D \), see definition of \( \mathcal{BR} \)).

It will be shown that for each \( t \) and for each public history \( h_t^{t-1} \), \( s[D(h_t^{t-1})] = s^c[h_t^{t-1}] \), contradicting the absurd hypothesis. The proof proceeds by induction on the length of the history, proceeding backwards from public histories \( h_T^{-1} \) to the empty history \( h_0^0 \).

**Initial Step:** It will be proven that \( s[D(h_T^{-1})] = s^c[h_T^{-1}] \) for each \( h_T^{-1} \).

Suppose, by contradiction, that \( \exists h_T^{-1} = (y_T^{-1}, x_T^{-1}) : s[D(h_T^{-1})] \neq s^c(h_T^{-1}) \).

40
Then, by the contraction property,
\[ \exists y_i^T \text{ and } \theta'_{i,T} \in D_i \left( h^{T-1}, y_i^T \right) : \theta'_{i,T} \neq s_i^c \left( h^{T-1}, y_i^T \right) \text{ such that:} \]
\[ \text{sign} \left[ s_i^c \left( h^{T-1}, y_i^T \right) - \theta'_{i,T} \right] = \text{sign} \left[ \alpha_i^T \left( y_i^T, y_{-i}^T \right) - \alpha_i^T \left( y_i^T, \theta'_{i,T} \right) \right] \]
\[ \text{for all } y_{-i}^T = \left( y_{-i}^{T-1}, \theta_{-i,T} \right) \text{ and } \theta'_{-i,T} \in D_{-i} \left( h^{T-1}, y_{-i}^T \right). \]

Fix such \( y_i^T \) and \( \theta'_{i,T} \neq s_i^c \left( h^{T-1}, y_i^T \right) \), and suppose (w.l.o.g.) that \( s_i^c \left( h^{T-1}, y_i^T \right) > \theta'_{i,T} \).

Define:
\[ \delta \left( h^{T-1}, y_i^T \right) := \min_{\substack{y_{-i}^T \in Y_{-i}^T \text{ and} \\
\theta'_{-i,T} \in D_{-i} \left( h^{T-1}, y_{-i}^T \right)}} \left[ \alpha_i^T \left( y_i^T, y_{-i}^T \right) - \alpha_i^T \left( y_i^T, \theta'_{i,T} \right) \right] \]
(by compactness of \( Y^T \) and \( D \left( h^T \right) \), \( \delta \left( h^{T-1}, y_i^T \right) \) is well-defined). Also, from \( s_i^c \left( h^{T-1}, y_i^T \right) > \theta'_{i,T} \) and the Contraction Property, \( \delta \left( h^{T-1}, y_i^T \right) > 0. \)

For any \( \varepsilon > 0 \), let
\[ \psi \left( h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon \right) = \max_{\theta_{-i,T} \in \Theta_{-i,T}} \left\{ \alpha_i^T \left( y_i^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T} \right) - \alpha_i^T \left( y_i^{T-1}, \theta_{i,T}, \theta_{-i,T} \right) \right\} \]
(again, compactness of \( \Theta_{-i,T} \) guarantees that \( \psi \left( h^T, \varepsilon \right) \) is well-defined). Since \( \alpha_i^T \) is strictly increasing in \( \theta_{i,T} \), \( \psi \left( h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon \right) \) is increasing in \( \varepsilon \) and \( \psi \left( h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon \right) \to 0 \) as \( \varepsilon \to 0. \)

Let \( \left( f_t \left( \tilde{y}_i^T \right) \right)_{t=1}^{T-1} = x_{T-1} \) (def. 1). From strict EPIC, we have that for each \( \varepsilon, \)
\[ v_i \left( x^{T-1}, f_T \left( y_i^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T} \right), \alpha_i^T \left( y_i^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T} \right), \alpha^{-T} \left( y_i^{T-1} \right) \right) \]
\[ > v_i \left( x^{T-1}, f_T \left( y_i^{T-1}, \theta'_{i,T}, \theta_{-i,T} \right), \alpha_i^T \left( y_i^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T} \right), \alpha^{-T} \left( y_i^{T-1} \right) \right) \]
and
\[ v_i \left( x^{T-1}, f_T \left( y_i^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T} \right), \alpha_i^T \left( y_i^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T} \right), \alpha^{-T} \left( y_i^{T-1} \right) \right) \]
\[ < v_i \left( x^{T-1}, f_T \left( y_i^{T-1}, \theta'_{i,T}, \theta_{-i,T} \right), \alpha_i^T \left( y_i^{T-1}, \theta'_{i,T}, \theta_{-i,T} \right), \alpha^{-T} \left( y_i^{T-1} \right) \right) \]
Thus, by continuity, there exists \( a^T \left( \varepsilon \right) \) such that
\[ \alpha_i^T \left( y_i^{T-1}, \theta'_{i,T}, \theta_{-i,T} \right) < a^T \left( \varepsilon \right) < \alpha_i^T \left( y_i^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T} \right) \]
\[ \text{such that} \]
\[ v_i \left( x^{T-1}, f_T \left( y_i^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T} \right), a^T \left( \varepsilon \right), \alpha^{-T} \left( y_i^{T-1} \right) \right) \]
\[ = v_i \left( x^{T-1}, f_T \left( y_i^{T-1}, \theta'_{i,T}, \theta_{-i,T} \right), a^T \left( \varepsilon \right), \alpha^{-T} \left( y_i^{T-1} \right) \right) \]
From single-crossing condition SCC-1 (def. 16),
\[ v_i \left( x^{T-1}, f_T \left( y_i^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T} \right), a^*, \alpha^{-T} \left( y_i^{T-1} \right) \right) \]
\[ > v_i \left( x^{T-1}, f_T \left( y_i^{T-1}, \theta'_{i,T}, \theta_{-i,T} \right), a^*, \alpha^{-T} \left( y_i^{T-1} \right) \right) \]
whenever \( a^* > a^T \left( \varepsilon \right) \)

41
Thus, to reach the contradiction, it suffices to show that for any $y^T_{i,t} \in Y_{i,T}^T$, $\alpha^T_i (y^T_i, y^T_{-i}) > a^T (\varepsilon)$: If this is the case, reporting $\theta^\prime_{i,T}$ is (conditionally) strictly dominated by reporting $\theta^\prime_{i,T} + \varepsilon$ at $h^T_{i,T} = (h^T_{i,T-1}, y^T_i)$, hence it cannot be that $D_i = BR_i$ and $\theta^\prime_{i,T} \in D_i (h^T_{i,T-1}, y^T_i)$. To this end, it suffices to choose $\varepsilon$ sufficiently small that

$$\psi (h^T_{i,T-1}, y^T_i, \theta^\prime_{i,T}, \varepsilon) < \delta (h^T_{i,T-1}, y^T_i)$$

(29)

and operate the substitutions as follows

$$\alpha^T_i (y^T_i, y^T_{-i}) \geq \alpha^T_i (y^T_{i,T}, \theta^\prime_{i,T}, \theta^\prime_{-i,T}) + \delta (h^T_{i,T-1}, y^T_i)$$

$$\geq \alpha^T_i (y^T_{i,T}, \theta^\prime_{i,T} + \varepsilon, \theta^\prime_{-i,T}) + \delta (h^T_{i,T-1}, y^T_i) - \psi (h^T_{i,T-1}, y^T_i, \theta^\prime_{i,T}, \varepsilon)$$

$$> \alpha^T_i (y^T_{i,T}, \theta^\prime_{i,T} + \varepsilon, \theta^\prime_{-i,T})$$

$$> a^T (\varepsilon)$$

Thus: $\alpha^T_i (y^T_i, y^T_{-i}) > a^T (\varepsilon)$ for any $y^T_{i,t}$. This concludes the initial step.

**Inductive Step:** [for $t = 1, ..., T$, it will be shown that if $s [D (h^T)] = s^c [h^T]$ for all $h^T$ s.t. $\tau \geq t$, then $s [D (h^T_{i,t})] = s^c [h^T_{i,t}]$ for all $h^T_{i,t}$]

By contradiction, suppose that there exists $h^T_{i,t} = (\bar{y}^T_{i,t}, x^T_{-i,t})$ such that:

$$s^c (h^T_{i,t}) \neq s^c (h^T_{i,t})$$

for all $h^T_{i,t} = (y^T_{i,t}, \theta^T_{i,t})$ and $\theta^T_{-i,t} \in D_{-i} (h^T_{i,t}, y^T_{i,t})$.

Fix such $y^T_{i,t}$ and $\theta^T_{i,t} \neq s^c_i (h^T_{i,t}, y^T_{i,t})$, and suppose (w.l.o.g.) that $s^c_i (h^T_{i,t}, y^T_{i,t}) > \theta^T_{i,t}$. Similar to the initial step, it will be shown that there exists $\theta^T_{i,t} = \theta^T_{i,t} + \varepsilon$ for some $\varepsilon > 0$ such that for any conjecture consistent with $D_{-i}$, playing $\theta^T_{i,t}$ is strictly better than playing $\theta^T_{i,t}$ at history $(h^T_{i,t}, y^T_{i,t})$, contradicting the hypothesis that $BR = D$.

For any $\varepsilon > 0$, set $\theta^T_{i,t} = \theta^T_{i,t} + \varepsilon$; for each realization of signals $\tilde{\theta}_i = \left( \tilde{\theta}_{i,k} \right)_{k=1}^T$ and opponents’ reports $\tilde{m}_{-i} = \left( \tilde{m}_{-i,k} \right)_{k=1}^T$, for each $\tau > t$, denote by $s^e_i (\theta^T_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i)$ the action taken at period $\tau$ if $\theta^T_{i,t}$ is played at $t$, $s^c_i$ is followed in the following stages, and the realized payoff type and opponents’ messages are $\tilde{\theta}_i$ and $\tilde{m}_{-i}$, respectively. (By continuity properties 1 and 2 of the aggregators functions (def. 12) and definition of $s^c$ (def. 13), $s^c_i (\theta^T_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i)$ is continuous in $\varepsilon$, and converges to $s^c_i (\theta^T_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i)$ as $\varepsilon \to 0$.)

For each realization $\tilde{\theta}_i = \left( \tilde{\theta}_{i,k} \right)_{k=1}^T$ and reports $\tilde{m}_{-i} = \left( \tilde{m}_{-i,k} \right)_{k=1}^T$ and for each $\tau > t$, $s^c_i (\theta^T_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i)$ may be one of five cases (cf. equations 17-19):
1. $s^c_{i,T} \left( \theta'_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \in (\theta^l_{i,T}, \theta^h_{i,T})$, then

$$\alpha^\tau_i \left( y^c_i, y^\tau_{-i} \right) = \alpha^\tau_i \left( \tilde{y}^c_{i-1}, \theta'_{i,t}, \left( s^c_{i,k} \left( \theta'_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \right)^\tau_{k=t+1}, y^\tau_{-i} \right)$$

for all $y^\tau_{-i}$, and we can choose $\varepsilon$ sufficiently small that $s^c_{i,T} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \in (\theta^l_{i,T}, \theta^h_{i,T})$, i.e.

$$\alpha^\tau_i \left( y^c_i, y^\tau_{-i} \right) = \alpha^\tau_i \left( \tilde{y}^c_{i-1}, \theta'_{i,t}, \left( s^c_{i,k} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \right)^\tau_{k=t+1}, y^\tau_{-i} \right)$$

for all $y^\tau_{-i}$.

2. $s^c_{i,T} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) = \theta^h_{i,T}$ and

$$\alpha^\tau_i \left( y^c_i, y^\tau_{-i} \right) > \alpha^\tau_i \left( \tilde{y}^c_{i-1}, \theta'_{i,t}, \left( s^c_{i,k} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \right)^\tau_{k=t+1}, \bar{y}^\tau_{-i} \right)$$

at $\bar{y}^\tau_{-i}$ defined as in equation (16), then we can choose $\varepsilon$ sufficiently small that $s^c_{i,T} \left( \theta^l_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) = \theta^h_{i,T}$ as well.

3. $s^c_{i,T} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) = \theta^h_{i,T}$ and

$$\alpha^\tau_i \left( y^c_i, y^\tau_{-i} \right) = \alpha^\tau_i \left( \tilde{y}^c_{i-1}, \theta'_{i,t}, \left( s^c_{i,k} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \right)^\tau_{k=t+1}, y^\tau_{-i} \right)$$

for all $y^\tau_{-i}$.

Then, either $s^c_{i,T} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) = \theta^h_{i,T}$ as well, or $s^c_{i,T} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \in (\theta^l_{i,T}, \theta^h_{i,T})$, i.e.

$$\alpha^\tau_i \left( y^c_i, y^\tau_{-i} \right) = \alpha^\tau_i \left( \tilde{y}^c_{i-1}, \theta'_{i,t}, \left( s^c_{i,k} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \right)^\tau_{k=t+1}, y^\tau_{-i} \right)$$

for all $y^\tau_{-i}$.

In either case,

$$\alpha^\tau_i \left( \tilde{y}^c_{i-1}, \theta'_{i,t}, \left( s^c_{i,k} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \right)^\tau_{k=t+1}, y^\tau_{-i} \right)$$

$$= \alpha^\tau_i \left( \tilde{y}^c_{i-1}, \theta'_{i,t}, \left( s^c_{i,k} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \right)^\tau_{k=t+1}, y^\tau_{-i} \right)$$

for all $y^\tau_{-i}$.

4. $s^c_{i,T} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) = \theta^l_{i,T}$ and

$$\alpha^\tau_i \left( y^c_i, y^\tau_{-i} \right) < \alpha^\tau_i \left( \tilde{y}^c_{i-1}, \theta'_{i,t}, \left( s^c_{i,k} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) \right)^\tau_{k=t+1}, \bar{y}^\tau_{-i} \right)$$

at $\bar{y}^\tau_{-i}$ defined as in equation (16), then we can choose $\varepsilon$ sufficiently small that $s^c_{i,T} \left( \theta^c_{i,t}, \bar{m}_{-i}, \tilde{\theta}_i \right) = \theta^l_{i,T}$ as well.
5. \( s^e_{i,\tau} \left( \theta^e_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta^e_{i,T} \) and
\[
\alpha^e_i \left( y_i^\tau, y_{-i}^\tau \right) = \alpha^e_i \left( \tilde{y}_{i}^{t-1}, \theta^e_{i,t}, \left( s^e_{i,k} \left( \theta^e_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^{\tau}, y_{-i}^\tau \right) \text{ for all } y_{-i}^\tau.
\]

Then, either \( s^e_{i,\tau} \left( \theta^e_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta^l_{i,T} \) as well, or \( s^e_{i,\tau} \left( \theta^e_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \in \left( \theta^l_{i,T}, \theta^h_{i,T} \right) \), i.e.
\[
\alpha^e_i \left( y_i^\tau, y_{-i}^\tau \right) = \alpha^e_i \left( \tilde{y}_{i}^{t-1}, \theta^l_{i,t}, \left( s^e_{i,k} \left( \theta^e_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^{\tau}, y_{-i}^\tau \right) \text{ for all } y_{-i}^\tau.
\]

In either case,
\[
\alpha^e_i \left( \tilde{y}_{i}^{t-1}, \theta^l_{i,t}, \left( s^e_{i,k} \left( \theta^e_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^{\tau}, y_{-i}^\tau \right) = \alpha^e_i \left( \tilde{y}_{i}^{t-1}, \theta^l_{i,t}, \left( s^e_{i,k} \left( \theta^e_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^{\tau}, y_{-i}^\tau \right) \text{ for all } y_{-i}^\tau.
\]

That is, for each \( \tau > t \), and for each \( \left( \tilde{\theta}_i, \tilde{m}_{-i} \right) \), in all five cases there exists \( \bar{\varepsilon} \left( \tilde{\theta}_i, \tilde{m}_{-i}, \tau \right) > 0 \) such that:
\[
\text{for all } \varepsilon \in \left( 0, \bar{\varepsilon} \left( \tilde{\theta}_i, \tilde{m}_{-i}, \tau \right) \right), \text{ for all } y_{-i}^\tau
\]
\[
\alpha^e_i \left( \tilde{y}_{i}^{t-1}, \theta^l_{i,t}, \left( s^e_{i,k} \left( \theta^e_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^{\tau}, y_{-i}^\tau \right) = \alpha^e_i \left( \tilde{y}_{i}^{t-1}, \theta^l_{i,t}, \left( s^e_{i,k} \left( \theta^e_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^{\tau}, y_{-i}^\tau \right).
\]

Let \( \bar{\varepsilon} = \min \varepsilon \left( \tilde{\theta}_i, \tilde{m}_{-i}, \tau \right) \) (by compactness, this is well-defined and such that \( \bar{\varepsilon} > 0 \)). Hence, if the continuation strategies are self-correcting, if \( f \) is aggregator-based, for any \( \varepsilon \in \left( 0, \bar{\varepsilon} \right) \), reporting \( \theta^e_{i,t} \) or \( \theta^l_{i,t} \) at period \( t \) does not affect the allocation chosen at periods \( \tau > t \) (recall that the opponents’ self-correcting report cannot be affected by \( i \)-th components of the public history). Hence, for \( \varepsilon \in \left( 0, \bar{\varepsilon} \right) \), for each \( \theta_{-i} \in \Theta_{-i} \), the allocations induced by following \( s^e_{i} \) at periods \( \tau > t \) and playing \( \theta^l_{i,t} \) or \( \theta^e_{i,t} \) at history \( h^i_t \), respectively \( \xi^l \) and \( \xi^e \), are such that \( \xi^l_{\tau} = \xi^e_{\tau} \) for all \( \tau \neq t \).

Consider types of player \( i \), \( \theta^l_{i}, \theta^e_{i} \in \Theta_i \) such that for each \( \tau < t \), \( \theta^e_{i,\tau} = \theta^l_{i,\tau} = \tilde{\theta}_{i,\tau} \) (where \( \tilde{\theta}_{i,\tau} \) are the types actually reported on the path), for all \( \tau > t \) and \( \theta^l_{i,\tau} = s^e_{i,\tau} \) as above, while at \( t \) respectively equal to \( \theta^e_{i,t} \) and \( \theta^l_{i,t} \). Thus, the induced allocations are \( \xi^e \) and \( \xi^l \) discussed above, and for each \( \tau \neq t \), \( \alpha^e_i \left( \theta^l \right) = \alpha^e_i \left( \theta^e \right) = \tilde{a}^e_i \).
From strict EPIC, we have that for any \( i \)
\[
v_i \left( \xi^\varepsilon, \alpha^t (\theta^\varepsilon), \{\tilde{a}_i^\tau \}_{\tau \neq t} \right) > v_i \left( \xi^t, \alpha^t (\theta^t), \{\tilde{a}_i^\tau \}_{\tau \neq t} \right)
\]
and
\[
v_i \left( \xi^\varepsilon, \alpha^t (\theta^t), \{\tilde{a}_i^\tau \}_{\tau \neq t} \right) < v_i \left( \xi^t, \alpha^t (\theta^t), \{\tilde{a}_i^\tau \}_{\tau \neq t} \right)
\]
Thus, by continuity, there exists \( a^t (\varepsilon) \)
\[
\alpha_i^t \left( \tilde{y}^{t-1}, \theta^t_{i,t}, \theta_{-i,t} \right) < a^t (\varepsilon) < \alpha_i^t \left( \tilde{y}^{t-1}, \theta^\varepsilon_{i,t}, \theta_{-i,t} \right)
\]
such that
\[
v_i \left( \xi^\varepsilon, a^t (\varepsilon), \{\tilde{a}_i^\tau \}_{\tau \neq t} \right) = v_i \left( \xi^t, a^t (\varepsilon), \{\tilde{a}_i^\tau \}_{\tau \neq t} \right)
\]
From the Single Crossing Condition,
\[
v_i \left( \xi^\varepsilon, a^*, \{\tilde{a}_i^\tau \}_{\tau \neq t} \right) > v_i \left( \xi^t, a^*, \{\tilde{a}_i^\tau \}_{\tau \neq t} \right)
\]
whenever \( a^* > a^t (\varepsilon) \).

Thus, since the continuations in periods \( \tau > t \) are the same under both \( \theta^t_{i,t} \) and \( \theta^\varepsilon_{i,t} \), to reach the desired contradiction it suffices to show that for any \( y_{-i}^t \in Y_{-i}^t, \alpha_i^t (y_i^t, y_{-i}^t) > a^t (\varepsilon) \).
(This, for any realization of \( \theta_{-i} \)).

As in the initial step, define:
\[
\delta := \min_{\begin{align*}
y_{-i}^t & \in Y_{-i}^t \text{ and } \\
\theta^t_{-i,t} & \in B_{-i}(h^{t-1}, \tilde{y}_{-i}^t)
\end{align*}} \left[ \alpha_i^t \left( y_i^t, y_{-i}^t \right) - \alpha_i^t \left( \tilde{y}^{t-1}, \theta^t_{i,t}, \theta^t_{-i,t} \right) \right] \quad (31)
\]
For any \( \varepsilon > 0 \), let
\[
\psi (\varepsilon) = \max_{\theta^t_{-i,t} \in \Theta_{-i,t}} \left\{ \alpha_i^t \left( \tilde{y}^{t-1}, \theta^\varepsilon_{i,t}, \theta_{-i,t} \right) - \alpha_i^t \left( \tilde{y}^{t-1}, \theta^t_{i,t}, \theta^t_{-i,t} \right) \right\} \quad (32)
\]
Since \( \alpha_i^t \) is strictly increasing in \( \theta_{i,t}, \psi (\varepsilon) \) is increasing in \( \varepsilon \) and \( \psi (\varepsilon) \to 0 \) as \( \varepsilon \to 0 \).
To obtain the desired contradiction, it suffices to choose \( \varepsilon \) sufficiently small that
\[
\psi (\varepsilon) < \delta \quad (33)
\]
and operate the substitutions as follows:
\[
\alpha_i^t \left( y_i^t, y_{-i}^t \right) \geq \alpha_i^t \left( \tilde{y}^{t-1}, \theta^t_{i,t}, \theta^t_{-i,t} \right) + \delta \\
\geq \alpha_i^t \left( \tilde{y}^{t-1}, \theta^\varepsilon_{i,t}, \theta^t_{-i,t} \right) + \delta - \psi (\varepsilon) \\
> \alpha_i^t \left( \tilde{y}^{t-1}, \theta^\varepsilon_{i,t}, \theta^t_{-i,t} \right) \\
> a^t (\varepsilon).
\]
C.3 Proof of Proposition 7.

The proof is very similar to that of proposition 6.

**Initial Step:** \([s[D^T(h^\tau)]] = s^c [h^T] \text{ for each } h^T\).

The initial step is the same, to conclude (in analogy with equation 28), that there exists \(a^T(\varepsilon)\) such that

\[ \alpha^T_i (\widetilde{y}^T, \theta'_i, \theta_{-i}) < a^T(\varepsilon) < \alpha^T_i (\widetilde{y}^T + \varepsilon, \theta_{-i}) \]  

such that

\[ v_i (x^T, f_T (\widetilde{y}^T, \theta'_i, \theta_{-i}), a^T(\varepsilon), \alpha^T_i (\widetilde{y}^T)) \]

\[ = v_i (x^T, f_T (\widetilde{y}^T, \theta'_i, \theta_{-i}), a^T(\varepsilon), \alpha^T_i (\widetilde{y}^T)) \]

Then, SCC-2 (def. 17) implies that

\[ v_i (x^T, f_T (\widetilde{y}^T, \theta'_i, \theta_{-i}), a^*, \alpha^T_i (\widetilde{y}^T)) \]

\[ > v_i (x^T, f_T (\widetilde{y}^T, \theta'_i, \theta_{-i}), a^*, \alpha^T_i (\widetilde{y}^T)) \]

whenever \(a^* > a^T(\varepsilon)\).

From this point, the argument proceeds unchanged, concluding the initial step.

**Inductive Step:** [for \(t = 1, ..., T - 1\): if \(s[D^T(h^\tau)] = s^c [h^\tau]\) for all \(h^\tau\) and all \(\tau > t\) then \(s[D^T(h^\tau)] = s^c [h^\tau]\) for all \(h^\tau\)]

The argument proceeds as in proposition 6, to show that for each \(\tau > t\), and for each \((\tilde{\theta}, \tilde{m}_{-i})\), if continuation strategies are self-correcting, there exists \(\varepsilon(\tilde{\theta}, \tilde{m}_{-i}, \tau) > 0\) such that:

for all \(\varepsilon \in \left(0, \varepsilon(\tilde{\theta}, \tilde{m}_{-i}, \tau)\right)\),

\[ \alpha^\tau_i (\widetilde{y}^T, \theta'_i, \theta_{-i}, \theta_k, \tilde{m}_{-i}, \tilde{\theta}_i, k = t+1, y^T) \]

\[ = \alpha^\tau_i (\widetilde{y}^T, \theta'_i, \theta_{-i}, \theta_k, \tilde{m}_{-i}, \tilde{\theta}_i, k = t+1, y^T) \]

for all \(y^T\).

Consider types of player \(i, \theta'_i, \theta^\varepsilon_i \in \Theta_i\) such that for each \(\tau < t\), \(\theta'_{i,\tau} = \theta^\varepsilon_{i,\tau} = \tilde{\theta}_{i,\tau}\) (the one actually reported on the path), for all \(\tau > t\) and \(\theta_{i,\tau} = s^c_{i,\tau}\) as above, while at \(t\) respectively equal to \(\theta'_{i,t}\) and \(\theta'_{i.t}\). By construction, such types are such that for any \(\tau \neq t\),

\[ \alpha^\tau_i (\theta^\varepsilon_i) = \alpha^\tau_i (\theta') \]

From strict EPIC, we have that for any \(\theta_{-i}\)

\[ v_i (f (\theta^\varepsilon_i), \alpha^t (\theta^\varepsilon_i), \{\tilde{a}^t_i\}_{\tau \neq t}) > v_i (f (\theta^\varepsilon_i), \alpha^t (\theta^\varepsilon_i), \{\tilde{a}^t_i\}_{\tau \neq t}) \]

and

\[ v_i (f (\theta^\varepsilon_i), \alpha^t (\theta^\varepsilon_i), \{\tilde{a}^t_i\}_{\tau \neq t}) < v_i (f (\theta^\varepsilon_i), \alpha^t (\theta^\varepsilon_i), \{\tilde{a}^t_i\}_{\tau \neq t}) \]

46
Thus, by continuity, there exists $a^t(\varepsilon)\quad (36)$

$$\alpha_i^t(\tilde{y}^{t-1}, \theta_i^t, \theta_{-i,t}) < a^t(\varepsilon) < \alpha_i^t(\tilde{y}^{t-1}, \theta_i^t, \theta_{-i,t})$$

such that

$$v_i(\xi^\varepsilon, a^t(\varepsilon); \{\hat{a}_i^t\}_{\tau \neq t}) = v_i(\xi', a^t(\varepsilon); \{\hat{a}_i^t\}_{\tau \neq t})$$

From SCC-2 (def. 17),

$$v_i\left(f(\theta^\varepsilon), a^*, \{\hat{a}_i^t\}_{\tau \neq t}\right) > v_i\left(f(\theta'), a^*, \{\hat{a}_i^t\}_{\tau \neq t}\right)$$

whenever $a^* > a^t(\varepsilon)$. The remaining part of the proof is identical to proposition 6.$\square$

References


