

# Robust Dynamic Mechanism Design\*

Antonio Penta<sup>†</sup>

University of Pennsylvania,  
Dept. of Economics

This version: November 11, 2009

## 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, the limit being tighter the stronger the “intertemporal effects”.

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.

**JEL Codes:** C72; C73; D82.

---

\*I thank my advisor, George Mailath, for his unique dedication, and committee members Andrew Postlewaite and Qingmin Liu. I'm indebted to all of them for constantly challenging me throughout this project, as well as for their encouragement. I also thank participants to the Workshop on “Information and Dynamic Mechanism Design” (HIM, Bonn) and seminar audiences at Penn. I am grateful to Pierpaolo Battigalli, Dirk Bergemann, Stephen Morris and Alessandro Pavan for their helpful comments.

<sup>†</sup> *email:* penta@sas.upenn.edu.

# 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.<sup>1</sup> 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 in which, *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’ entire hierarchies of beliefs is too strong. In dynamic settings in particular, the assumption of a commonly known prior entails the planner’s knowledge of significantly more complex objects, such as agents’ *hierarchies of conditional beliefs*: For instance, in the example above, it means that the designer knows the firms’ conditional beliefs (conditional on all possible realizations of private signals) over own future productivity and the other firms’ current and future productivities *and* their beliefs, conditional on all realization of *their* signals. Not only are these assumptions strong, but the sensitivity of game

---

<sup>1</sup>Among others, see Bergemann and Valimaki (2008), Athey and Segal (2007), Pavan, Segal and Toikka (2009). Gershov and Moldovanu (2009a,b) depart from the “standard” approach described above in that the designer does not know the “true” distribution, combining implementation problems with learning.

theoretic results to the fine details of agents' higher order beliefs is also well documented. Weakening the reliance of game theoretic analysis on common knowledge assumptions seems thus crucial to enable us "to conduct useful analysis of practical problems" (Wilson, 1987, p.34).<sup>2</sup>

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*. The analysis of the full implementation question instead raises novel problems: in order to achieve robust full implementation we need a mechanism such that, for any model of beliefs, all the PBE induce outcomes consistent with the SCF. The direct approach to the question is to compute the set of PBE for each model of beliefs; but the obvious difficulties that this task presents have been a major impediment to the development of a *robust* approach to *dynamic mechanism design*.

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 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.)

The results are 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

---

<sup>2</sup>In the context of mechanism design, this research agenda (sometimes referred to as *Wilson's doctrine*) has been put forward in a series of papers by Bergemann and Morris, who developed a belief-free approach to classical implementation questions, known as "*robust*" *mechanism design*.

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, such bounds being more stringent the stronger the “intertemporal effects”.

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 introducing 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 chooses some quantity  $q_t$  of 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 the public goods  $q_1$  and  $q_2$  are, respectively,

$$\begin{aligned} \alpha_{i,1}(\theta_1) &= \theta_{i,1} + \gamma\theta_{j,1} \\ &\text{and} \\ \alpha_{i,2}(\theta_1, \theta_2) &= \varphi(\theta_{i,1}, \theta_{i,2}) + \gamma\varphi(\theta_{j,1}, \theta_{j,2}) \end{aligned}$$

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

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

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

$$q_1^*(\theta_1) = \alpha_{i,1}(\theta_1) + \alpha_{j,1}(\theta_1) \quad \text{and} \quad (2)$$

$$q_2^*(\theta) = \alpha_{i,2}(\theta_1, \theta_2) + \alpha_{j,2}(\theta_1, \theta_2). \quad (3)$$

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  $m = (m_{i,1}, m_{j,1}, m_{i,2}, m_{j,2})$ , agent  $i$  receives *generalized Vickrey-Clarke-Groves* transfers

$$\pi_{i,1}^*(m_{i,1}, m_{j,1}) = - (1 + \gamma) \left[ \gamma \cdot m_{i,1} \cdot m_{j,1} + \frac{1}{2} m_{j,1}^2 \right] \quad \text{and} \quad (4)$$

$$\pi_{i,2}^*(m_{i,1}, m_{j,1}) = - (1 + \gamma) \left[ \gamma \cdot \varphi(m_{i,1}, m_{i,2}) \cdot \varphi(m_{j,1}, m_{j,2}) + \frac{1}{2} \varphi(m_{j,1}, m_{j,2})^2 \right], \quad (5)$$

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 game with incomplete information. The solution concept that will be used for this kind of environments is "*interim perfect equilibrium*" (IPE), a weaker version of Perfect Equilibrium 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  $f = (q_t^*, \pi_{i,t}^*, \pi_{j,t}^*)_{t=1,2}$  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

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

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  $(\hat{\theta}_{i,1}, \hat{\theta}_{i,2})$ , 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:<sup>3</sup>

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

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

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

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 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 across models of beliefs. In the procedure, for each public history  $\hat{m}_1$  (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

---

<sup>3</sup>We ignore here the possibility of corner solutions, which do not affect the fundamental insights. Corner solutions will be discussed in Section 8.

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 ( $m_{j,t} = \theta_{j,t}$  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.:  $\Delta(\theta, \hat{m}_1, m_2) = 0$ .) This is the notion of *self-correcting strategy*,  $s_i^c$ : 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  $s_i^c$  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  $\hat{m}_1$  and  $\theta_i = (\theta_{i,1}, \theta_{i,2})$ , let  $x_i(\theta_i) = [\varphi(\hat{m}_{i,1}, m_{i,2}) - \varphi(\theta_{i,1}, \theta_{i,2})]$  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  $\underline{x}_j^0$  and  $\bar{x}_j^0$  denote the minimum and maximum possible values of  $x_j$ . Then, if  $i$  is rational, his over-reports are bounded by  $x_i(\theta_i) \leq \bar{x}_i^1 \equiv \gamma \cdot \underline{x}_j^0$  and  $x_i(\theta_i) \geq \underline{x}_i^1 \equiv -\gamma \cdot \bar{x}_j^0$ . Recursively, define  $\bar{x}_i^k = -\gamma \cdot \underline{x}_j^{k-1}$  and  $\underline{x}_i^k = -\gamma \sum_{j \neq i} \bar{x}_j^{k-1}$ . Also, for each  $k$  and  $i$ , let  $y_i^k \equiv [\bar{x}_i^k - \underline{x}_i^k]$  denote the distance between the maximum and lowest possible over-report at step  $k$ . Then, substituting, we obtain the following system of difference equations:

$$\mathbf{y}^k = \mathbf{\Gamma} \cdot \mathbf{y}^{k-1} \quad \text{where} \quad (8)$$

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

Notice that the continuation game from  $\hat{m}_1$  is dominance solvable if and only if  $\mathbf{y}^k \rightarrow 0$  as  $k \rightarrow \infty$ . In that case, for each  $\theta_i$ ,  $x_i(\theta_i) \rightarrow 0$ , and so truthtelling is the unique rationalizable strategy. 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 “backward 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 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, allowing us to construct the recursive system (8). Several versions of single-crossing conditions generalize this aspect in Section 8. Finally, the generalization of the idea that  $\gamma < 1$  takes the form of a “contraction property”.<sup>4</sup> 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  $i \in \{1, 2\}$  which can unilaterally sign  $[\alpha_{i,1}(m'_{i,1}, m_{j,1}) - \alpha_{i,1}(\theta_{i,1}, \theta_{j,1})]$  by reporting some message  $m'_{i,1} \neq \theta_{i,1}$ , irrespective of  $\theta_{j,1}$  and  $m_{j,1}$ . That is, for all  $\theta_{j,1}$  and  $m_{j,1}$  in  $D$ :

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

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  $[\alpha_{i,2}(\hat{m}_1, m'_{i,2}, m_{j,2}) - \alpha_{i,2}(\theta_i, \theta_j)]$ , (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,  $s_{i,2}^c$ :

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

This property will be required to hold at all histories.<sup>5</sup>

### 3 Environments.

Consider an environment with  $n$  agents and  $T$  periods. In each period  $t = 1, \dots, T$ , each agent  $i = 1, \dots, n$  observes a signal  $\theta_{i,t} \in \Theta_{i,t}$ . For each  $t$ ,  $\Theta_t := \Theta_{1,t} \times \dots \times \Theta_{n,t}$  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

<sup>4</sup>The name, borrowed from Bergemann and Morris (2009), is evocative of the logic behind equation 8.

<sup>5</sup>The general formulation (Section 8) allows to accommodate the case analogous to the possibility of corner solutions in the example above.

every period. A *state of nature* is a profile of agents' payoff types, and the set of states of nature is defined as  $\Theta^* := \Theta_1^* \times \dots \times \Theta_n^*$ .

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  $i = 1, \dots, n$ , preferences are represented by utility functions  $u_i : \Xi^* \times \Theta^* \rightarrow \mathbb{R}$ . 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' *knowledge* 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 knowledge of everyone's preferences about the feasible allocations.

For each  $t$ , let  $Y_i^t := \times_{\tau=1}^t \Theta_i^\tau$  denote the set of possible histories of player  $i$ 's signals up to period  $t$ . For each  $t$  and private signals  $y_i^t = (\theta_{i,1}, \dots, \theta_{i,t}) \in Y_i^t$ , agent  $i$  *knows* that the "true" state of nature  $\theta^* \in \Theta^*$  belongs to the set  $\{y_i^t\} \times (\times_{\tau=t+1}^T \Theta_{i,\tau}) \times \Theta_{-i}^*$ .

At any point in time, agent form *beliefs* about the features of the environment they don't *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. That is:

**Remark 1** *A necessary condition for a SCF  $f : \Theta^* \rightarrow \Xi^*$  to be implementable is that there exist functions  $f_t : Y^t \rightarrow \Xi_t$ ,  $t = 1, \dots, T$ , such that for each  $\theta = (\theta_1, \dots, \theta_T)$ ,  $f(\theta) = (f_t(\theta_1, \dots, \theta_t))_{t=1}^T$ .*

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

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

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

Formally, let  $H^0 := \{\phi\}$  ( $\phi$  denotes the empty history). For each  $t = 1, \dots, T$ , 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:

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

The set of *public histories* is defined as  $\mathcal{H} = \cup_{\tau=0}^T H^\tau$ . 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 Games with Payoff Uncertainty.

An environment  $\mathcal{E}$  and a mechanism  $\mathcal{M}$  determine a *dynamic game with payoff uncertainty*, that is a tuple

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

Sets  $N$ ,  $\Theta_i^*$  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  $Y_i^t = (\times_{\tau=1}^t \Theta_{i,\tau})$ ; sets  $H^t$  ( $t = 0, 1, \dots, T$ ) are defined as in  $\mathcal{M}$ ; player  $i$ 's set of *private histories of length  $t$*  ( $t = 1, \dots, T$ ) is defined as  $H_i^t := H^{t-1} \times Y_i^t$ , and finally  $\mathcal{H}_i := \{\phi\} \cup (\cup_{\tau=1}^T H_i^\tau)$  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:  $(h^{\tau-1}, y_i^\tau) \prec (h^{t-1}, y_i^t)$  if and only if  $h^\tau \prec h^t$  and  $y_i^\tau \prec y_i^t$ .)

**Remark 2** *The tuple  $(\mathcal{E}, \mathcal{M})$  is not a standard incomplete information game (Harsanyi, 1967-68), because it does not encompass a specification of agents' interactive beliefs. A standard incomplete information 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  $s_i(h_i^t) \in M_{i,t}$  for each  $h_i^t \in \mathcal{H}_i$ . The set of player  $i$ 's 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  $\mathbf{g}^s(\theta) \in \Xi^*$ . Hence, we can define strategic-form payoff functions  $U_i : S \times \Theta^* \rightarrow \mathbb{R}$  as  $U_i(s, \theta) = u_i(\mathbf{g}^s(\theta), \theta)$  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  $h^t$  and player  $i$ , let  $S_i(h^t)$  denote the set of player  $i$ 's strategies that are consistent with history  $h^t$  being observed. Clearly, since  $i$ 's private histories are only informative about the opponents' behavior through the public history, for each  $i$ ,  $h_i^t = (h^{t-1}, y_i^t) \in \mathcal{H}_i$  and  $j \neq i$ ,  $S_j(h_i^t) = S_j(h^{t-1})$ .

For each  $h^t$ ,  $S_i^{h^t}$  denotes the set of strategies in the subform starting from  $h^t$ , and for each  $s_i \in S_i$ ,  $s_i|_{h^t} \in S_i^{h^t}$  denotes the continuation  $s_i$  starting from  $h^t$ . The notation  $\mathbf{g}^{s|h^t}(\theta)$  refers to the terminal history induced by strategy profile  $s$  from the public history  $h^t$  if the realized state of nature is  $\theta$ . Strategic-form payoff functions can be defined for continuations from a given public history: for each  $h^t \in \mathcal{H}$  and each  $(s, \theta) \in S \times \Theta^*$ ,  $U_i(s, \theta; h^t) = u_i(\mathbf{g}^{s|h^t}(\theta), \theta)$ . For the initial history  $\phi$ , it will be written  $U_i(s, \theta)$  instead of  $U_i(s, \theta; \phi)$ . Sets  $\mathcal{H}_i$  and  $S_i$  are endowed with the standard metrics derived from the Euclidean metric on  $H^T \times \Theta^*$ .<sup>6</sup>

## 4.2 Direct Mechanisms.

The notion of *direct mechanism* is based on the observation made in remark 1:

**Definition 1** *A mechanism is direct, denoted by  $\mathcal{M}^*$ , if for each  $i$  and for each  $t = 1, \dots, T$ ,  $M_{i,t} = \Theta_{i,t}$ , 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  $f_t : Y^t \rightarrow \Xi_t$ , 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  $S_i^*$ , with typical element  $s_i^*$ .

---

<sup>6</sup>See Appendix A.1 for details.

## 5 Models of Beliefs.

A *model of beliefs* for an environment  $\mathcal{E}$  is a tuple

$$\mathcal{B} = \langle N, \Theta^*, (B_i, \beta_i)_{i \in N} \rangle$$

such that for each  $i$ ,  $B_i$  is measurable space (the set of *types*), and  $\beta_i : B_i \rightarrow \Delta(\Theta^* \times B_{-i})$  is a measurable function.<sup>7</sup>

At period 0 agents have no information about the environment. Their (subjective) “priors” about the payoff state and the opponents’ beliefs is implicitly represented by means of types  $b_i$ , with beliefs given by  $\beta_i(b_i) \in \Delta(\Theta^* \times B_{-i})$ . In periods  $t = 1, \dots, T$ , 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 don’t *know* their own *payoff-types* 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.

In standard models of dynamic mechanism design (e.g. Bergemann and Valimaki, 2008, and Athey and Segal, 2007, Pavan et al., 2009), the history of payoff types completely determines players’ beliefs about the payoff states and the opponents’ beliefs at each point of the process.<sup>8</sup> In the present setting this corresponds to the case where, for each  $i$ ,  $B_i$  is a singleton and  $\text{supp}\left(\text{marg}_{\Theta_i^*} \beta_i(b_i)\right) = \Theta_i^*$ : A unique “prior” describes the beliefs (of any order) for each player, so that conditional beliefs are uniquely determined for all possible realizations of the payoff types.

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  $t = 1, \dots, T$ , when a type’s beliefs after a history of signals  $y_i^t$  would be concentrated on the set

$$\{y_i^t\} \times \left(\times_{\tau=t+1}^T \Theta_{i,\tau}\right) \times \Theta_{-i}^* \times B_{-i}.$$

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

---

<sup>7</sup>For a measurable space  $X$ ,  $\Delta(X)$  denotes the set of probability measures on  $X$ , endowed with the topology of weak convergence and the corresponding Borel sigma-algebra.

<sup>8</sup>Classical mechanism design focuses almost exclusively on the case of *payoff type spaces*. Neeman (2004) shows how relaxing this assumption may crucially affect the results.

to period  $t$  are revealed. By “*ex-post*” stage we refer to the final realization, when *payoff-states* are fully revealed (or the *period- $T$  ex-post* stage).

## 6 Solution Concepts.

This section is organized in two parts: the first, introduces the main solution concept for environments with a model of beliefs; the second introduces the solution concept for environments without beliefs, which will be used in the analysis of the *full-implementation* problem in Section 8.

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

A tuple  $(\mathcal{E}, \mathcal{M}, \mathcal{B})$  determines a dynamic incomplete information game in the sense of Harsanyi. Strategies are thus measurable mappings  $\sigma_i : B_i \rightarrow S_i$ , and the set of strategies is denoted by  $\Sigma_i$ . At period 0, agents only know their own type. Hence, the set of agent  $i$ 's private histories of length 0 coincides with the set of his types. It is therefore convenient to identify types with such histories, and write  $h_i^0 \in B_i$ .

A *system of beliefs* consists of collections  $(p_i(h_i^t))_{h_i^t \in \mathcal{H}_i \setminus \{\phi\}}$  for each agent  $i$ , such that  $p_i(h_i^t) \in \Delta(\Theta^* \times B_{-i})$ : a *belief system* represents agents' conditional beliefs about the payoff state and the opponents' *types* at each private history. 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 private history  $h_i^{t-1}$ , a probability measure  $P^{\sigma, p_i}(h_i^{t-1})$  over the histories of length  $t$ .

**Definition 2** Fix a strategy profile  $\sigma \in \Sigma$ . A beliefs system  $p$  is consistent with  $\sigma$  if for each  $i \in N$  :

$$\forall h_i^0 \in B_i, p_i(h_i^0) = \beta_i(h_i^0) \tag{9}$$

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

and for each  $h_i^t$  such that  $h_i^{t-1} \prec h_i^t$ ,  $p_i(h_i^t)$  is obtained from  $p_i(h_i^{t-1})$  and  $P^{\sigma, p_i}(h_i^{t-1})$  via Bayesian updating (whenever possible).

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  $\mathcal{B}$ ; condition (10) requires conditional beliefs at each private history to be consistent with the information about the payoff state contained in the history itself; finally, the belief system  $p_i$  is consistent with Bayesian updating whenever possible.

From the point of view of each  $i$ , for each  $h_i^t \in \mathcal{H}_i \setminus \{\phi\}$  and strategy profile  $\sigma$ , the induced terminal history is a random variable that depends on the realization of the state of nature and opponents' type profile (agent  $i$ 's type  $h_i^0$  is known to agent  $i$  at  $h_i^t$ ,  $h_i^0 \prec h_i^t$ ). This is denoted by  $\mathbf{g}^{\sigma|h_i^t}(\theta, b_{-i})$ . As done for games without a model of beliefs (Section 4.1), we can define strategic-form payoff functions as follows:

$$U_i(\sigma, \theta, b_{-i}; h_i^t) = u_i\left(\mathbf{g}^{\sigma|h_i^t}(\theta, b_{-i}), \theta\right).$$

**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  $h_i^t \in \mathcal{H}_i \setminus \{\phi\}$ , the following inequality is satisfied for every  $\sigma'_i \in \Sigma_i$ :

$$\begin{aligned} & \int_{\Theta^* \times B_{-i}} U_i(\sigma, \theta, b_{-i}; h_i^t) \cdot dp_i(h_i^t) \\ & \geq \int_{\Theta^* \times B_{-i}} U_i(\sigma'_i, \sigma_{-i}, \theta, b_{-i}; h_i^t) \cdot dp_i(h_i^t). \end{aligned} \tag{11}$$

**Definition 4** An assessment  $(\sigma, p)$  is an Interim Perfect Equilibrium (IPE) if:<sup>9</sup>

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

If inequality (11) 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, consistency of the belief system.

The notion of consistency adopted here imposes no restrictions on the beliefs held at histories that receive zero probability at the preceding node.<sup>10</sup> 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  $h_i^t$  is precluded by  $\sigma_i(h_i^{t-1})$  alone,

---

<sup>9</sup>I avoid the adjective "Bayesian" (preferring the terminology "*interim*" *perfect equilibrium*) because the models of beliefs under consideration are not necessarily consistent with a common prior. For the same reason, Bergemann and Morris (2005) preferred the terminology "*interim*" to that of *Bayes-Nash equilibrium*.

<sup>10</sup>IPE is consistent with a "trembling-hand" view of unexpected moves, in which no restrictions on the possible correlations between "trembles" and other elements of uncertainty are imposed. 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.

$h_i^t \notin \text{supp} P^{\sigma, p_i}(h_i^{t-1})$ , and agent  $i$ 's beliefs at  $h_i^t$  are unrestricted the same way they would be after an unexpected move of the opponents. This feature of IPE is not standard, but it is key to the result that the set of IPE-strategies across models of beliefs can be computed by means of a convenient “backwards procedure”: Treating own deviations the same as the opponents’ is key to the possibility of considering continuation games “in isolation”, necessary for the result. 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 that follows adapts to that case: The “backwards procedure” to compute the set of equilibria across models of beliefs must be modified, so to keep track of the restrictions the extensive form imposes on the agents’ beliefs at unexpected nodes. Losing the possibility of envisioning continuation games “in isolation”, the modified procedure is more complicated, essentially undoing the advantages of the *indirect approach*.

Furthermore, for the sake of the full-implementation analysis, it can be shown that in the framework considered in Section 8, the set of IPE-strategies across models of beliefs coincides with the set of *strong IPE*-strategies across models of beliefs. Thus, from the point of view of the full-implementation results of Section 8, this point is not critical.

## 6.2 Mechanisms in Environments without Beliefs: $(\mathcal{E}, \mathcal{M})$ .

This section introduces a solution concept for dynamic games without a model of beliefs, *backward rationalizability* ( $\mathcal{BR}$ ), and shows that it characterizes the set of IPE-strategies across models of beliefs (proposition 1). It is also shown that  $\mathcal{BR}$  can be conveniently solved by a “backwards procedure” that extends the logic of backward induction to environments with incomplete information (proposition 2).

In environments without a model of beliefs we will not follow a classical equilibrium approach: 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 represented in the models of Section 5.

**Conjectures.** 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  $h_i^t = (h^{t-1}, y_i^t) \in \mathcal{H}_i$ , define the event  $[h_i^t] \subseteq \Theta^* \times S$  as:

$$[h_i^t] = \{y_i^t\} \times (\times_{\tau=t+1}^T \Theta_{i,\tau}) \times \Theta_{-i}^* \times S (h^{t-1}).$$

(Notice that, by definition,  $[h_i^t] \subseteq [h_i^{t-1}]$  whenever  $h_i^{t-1} \preceq h_i^t$ .)

**Definition 5** A conjecture for agent  $i$  is a conditional probability system (CPS hereafter), that is a collection  $\mu^i = (\mu^i(h_i^t))_{h_i^t \in \mathcal{H}_i}$  of conditional distributions  $\mu^i(h_i^t) \in \Delta(\Theta^* \times S)$

that satisfy the following conditions:

C.1 For all  $h_i^t \in \mathcal{H}_i$ ,  $\mu^i(h_i^t) \in \Delta([h_i^t])$ ;

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

The set of agent  $i$ 's conjectures is denoted by  $\Delta^{\mathcal{H}_i}(\Theta^* \times S)$ .<sup>11</sup>

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 in Section 6.3.

**Sequential Rationality.** A strategy  $s_i$  is *sequentially rational with respect to conjectures*  $\mu^i$  if, at each history  $h_i^t \in \mathcal{H}_i$ , it prescribes optimal behavior with respect to  $\mu^i(\cdot; h_i^t)$  in the continuation of the game. Formally: Given a CPS  $\mu^i \in \Delta^{\mathcal{H}_i}(\Theta^* \times S)$  and a history  $h_i^t = (h^{t-1}, y_i^t)$ , strategy  $s_i$  expected payoff at  $h_i^t$ , given  $\mu^i$ , is defined as:

$$U_i(s_i, \mu^i; h_i^t) = \int_{\Theta^* \times S_{-i}} U_i(s_i, s_{-i}, \theta; h^{t-1}) \cdot dmarg_{\Theta^* \times S_{-i}} \mu^i(h_i^t). \quad (12)$$

**Definition 6** A strategy  $s_i$  is sequentially rational with respect to  $\mu^i \in \Delta^{\mathcal{H}_i}(\Theta^* \times S)$ , written  $s_i \in r_i(\mu^i)$ , if and only if for each  $h_i^t \in \mathcal{H}_i$  and each  $s'_i \in S_i$  the following inequality is satisfied:

$$U_i(s_i, \mu^i; h_i^t) \geq U_i(s'_i, \mu^i; h_i^t). \quad (13)$$

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

### 6.2.1 Backward Rationalizability.

We introduce now the solution concept that will be shown (proposition 1) to characterize the set of IPE-strategies across models of beliefs, *Backwards Rationalizability* ( $\mathcal{BR}$ ). The name is justified by proposition 2, which shows that  $\mathcal{BR}$  can be computed by means of a “backwards procedure” that combines the logic of rationalizability and backwards induction.

---

<sup>11</sup>The general definition of a CPS is in appendix A.2.

**Definition 7** For each  $i \in N$ , let  $\mathcal{BR}_i^0 = S_i$ . Define recursively, for  $k = 1, 2, \dots$

$$\mathcal{BR}_i^k = \left\{ \begin{array}{l} \exists \mu^i \in \Delta^{\mathcal{H}_i}(\Theta^* \times S) \text{ s.t.} \\ (1) \hat{s}_i \in r_i(\mu^i) \\ (2) \text{supp}(\mu^i(\phi)) \subseteq \Theta^* \times \{\hat{s}_i\} \times \mathcal{BR}_{-i}^{k-1} \\ \hat{s}_i \in \mathcal{BR}_i^{k-1} : (3) \text{ for each } h_i^t = (h^{t-1}, y_i^t) \in \mathcal{H}_i: \\ \quad s \in \text{supp}(\text{marg}_S \mu^i(h_i^t)) \text{ implies:} \\ (3.1) s_i | h_i^t = \hat{s}_i | h_i^t, \text{ and} \\ (3.2) \exists s'_{-i} \in \mathcal{BR}_{-i}^{k-1} : s'_{-i} | h^{t-1} = s_{-i} | h^{t-1} \end{array} \right\}$$

Finally,  $\mathcal{BR} := \bigcap_{k \geq 0} \mathcal{BR}^k$ .<sup>12</sup>

$\mathcal{BR}$  consists of an iterated deletion procedure. At each round, strategy  $s_i$  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  $s_i$  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 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  $\text{marg}_{\Theta^*} \mu^i(\cdot | h_i^t)$  is concentrated on a "type"  $y_{-i}^t$  for which some of the previous moves in  $h^{t-1}$  are irrational. Nonetheless, condition (3.2) implies that it is believed that  $y_{-i}^t$  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*.

---

<sup>12</sup>It goes without saying that whenever we write a condition like  $\mu^i(X|h_i^t) \geq \kappa$  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.

Thus, CCFR can be interpreted as a condition of *belief persistence* on the continuation strategies.<sup>13</sup>

### 6.3 Results.

We discuss now the two main results which are useful to tackle the problem of full implementation in Section 8. The first result shows that  $\mathcal{BR}$  characterizes the set of IPE-strategies across models of beliefs; the second result shows that this set can be computed by means of a convenient “backwards procedure”.

**Characterization of the set of IPE.** As emphasized above, in  $\mathcal{BR}$  agents hold conjectures about both the opponents’ and their own strategies. First, notice that conditions (2) and (3.2) in the definition of  $\mathcal{BR}$  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  $\mathcal{BR}$  does not affect the standard notion of rationality. The fact that conjectures are elements of  $\Delta^{\mathcal{H}_i}(\Theta^* \times S)$  rather than  $\Delta^{\mathcal{H}_i}(\Theta^* \times S_{-i})$  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  $\Theta^* \times S_{-i}$ , 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.1):

**Proposition 1 (Characterization)** *Fix a game  $(\mathcal{E}, \mathcal{M})$ . For each  $i$ :  $\hat{s}_i \in \mathcal{BR}_i$  if and only if  $\exists \mathcal{B}, \hat{b}_i \in B_i$  and  $(\hat{\sigma}, \hat{p})$  such that: (i)  $(\hat{\sigma}, \hat{p})$  is an IPE of  $(\mathcal{E}, \mathcal{M}, \mathcal{B})$  and (ii)  $\hat{s}_i \in \text{supp } \hat{\sigma}_i(\hat{b}_i)$ .*

An analogous result can be obtained for the more standard refinement of IPE, in which unilateral own deviations leave an agents’ beliefs unchanged, applying to a modified version of  $\mathcal{BR}$ : such modification entails assuming that agents only form conjectures about  $\Theta^* \times S_{-i}$  (that is:  $\mu^i \in \Delta^{\mathcal{H}_i}(\Theta^* \times S_{-i})$ ) and by consequently adapting conditions (2) and (3) in the definition of  $\mathcal{BR}$ . (See Penta, 2009.) Hence, the assumption that IPE treats anyone’s deviation the same (and, correspondingly, that in  $\mathcal{BR}$  agents hold conjectures

---

<sup>13</sup>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”. (cf. Perea, 2009, and discussion below, Section 9.)

about their own strategy as well) is not crucial to characterize the set of equilibrium strategies across models of beliefs. As already discussed in Section 6.1, it is crucial instead for the next result, which shows that such set can be computed applying a procedure that extends the logic of backward induction to environments with incomplete information (proposition 2 below).

**The “Backwards Procedure”.** The backwards procedure is described as follows: Fix a public history  $h^{T-1}$  of length  $T - 1$ . For each payoff-type  $y_i^T \in \Theta_i^*$  of each agent, the continuation game is a static game, to which we can apply the standard notion of  $\Delta$ -rationalizability (Battigalli and Siniscalchi, 2003). For each  $h^{T-1}$ , let  $\mathcal{R}_i^{h^{T-1}}$  denote the set of pairs  $(y_i^T, s_i|h^{T-1})$  such that continuation strategy  $s_i|h^{T-1}$  is rationalizable in the continuation game from  $h^{T-1}$  for type  $y_i^T$ . We now proceed backwards: for each public history  $h^{T-2}$  of length  $T - 2$ , we apply again  $\Delta$ -rationalizability to the continuation game from  $h^{T-2}$  (in normal form), restricting continuation strategies  $s_i|h^{T-2} \in S_i^{h^{T-2}}$  to be  $\Delta$ -rationalizable in the continuation games from histories of length  $h^{T-1}$ .  $\mathcal{R}_i^{h^{T-2}}$  denotes the set of pairs  $(y_i^{T-1}, s_i|h^{T-2})$  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  $\mathcal{R}_i^\phi$  is computed. We can now introduce the “backwards procedure” result (the proof and the formal definition of  $\mathcal{R}^\phi$  are in appendix B.2):

**Proposition 2 (Computation)**  $\mathcal{BR}_i = \mathcal{R}_i^\phi$  for each  $i$ .

Properties UIA and CCFR provide the basic insight behind this result. 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  $h^{t-1}$ . 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  $\mathcal{BR}$ , 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”.<sup>14</sup>

---

<sup>14</sup>Hence,  $\mathcal{BR}$  satisfies a property that generalizes the notion of “subgame consistency”, according to which ‘the behavior prescribed on a subgame is nothing else than the solution of the subgame itself’ (Harsanyi and Selten, 1988, p.90).

## 7 Partial Implementation.

Under our assumption that the designer can commit to the mechanism, it is easy to show that a *revelation principle* holds for dynamic environments, so that restricting attention to *direct mechanisms* (definition 1) entails no loss of generality for the analysis of the *partial implementation* problem.

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

**Definition 8** A SCF is interim implementable (or 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  $(\mathcal{E}, \mathcal{M}^*, \mathcal{B})$ . That is,  $\exists \sigma^* \in \Sigma^*$  such that for each  $i \in N$  and  $b_i \in B_i$ , for all  $\sigma_i \in \Sigma_i$ ,

$$\begin{aligned} & \int_{\Theta^* \times B_{-i}} U_i(\sigma^*, \theta, b_{-i}; b_i) \cdot d\beta(b_i) \\ & \geq \int_{\Theta^* \times B_{-i}} U_i(\sigma_i, \sigma_{-i}^*, \theta, b_{-i}; b_i) \cdot d\beta(b_i). \end{aligned}$$

(Recall that  $\Sigma^*$  denotes the set of truthtelling 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** A SCF  $f$  is ex post implementable (or ex post incentive compatible) if for each  $i$ , for each  $\theta \in \Theta^*$  and  $s'_i \in S_i$

$$U_i(s^*, \theta) \geq U_i(s'_i, s_{-i}^*, \theta).$$

We say that a SCF is *Strictly Ex-Post Incentive Compatible* if for any  $s'_i \notin S_i^*$ , 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 implementable* if the truthtelling strategy is an IPE of the direct mechanism.

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

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 3 (Partial Implementation)** *A SCF is perfect implementable on all models of beliefs if and only if it is ex post implementable.*

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.

On the other hand, in environments with dynamic revelation of information, agents’ signals are intrinsically multidimensional. Hence, given proposition 3, the negative result on ex-post implementation by Jehiel et al. (2006) can be interpreted as setting tight limits for the *Wilson’s doctrine* applied to dynamic mechanism design problems. However, the literature provides examples of environments of economic interest where ex-post implementation with multidimensional signals is possible (e.g. Picketty, 1999; Bikhchandani, 2006; Eso and Maskin, 2002).

## 8 Full Implementation in Direct Mechanisms.

We begin by focusing on *direct mechanisms*. Unlike the static case of Bergemann and Morris (2009), in environments with dynamic revelation of information direct mechanisms may not suffice to achieve full robust implementation: Section 8.4 shows how simple “enlarged” mechanisms, can improve on the direct ones, yet avoiding the intricacies of the “augmented mechanisms” required for classical *Bayesian Implementation*.<sup>15</sup>

**Definition 11** *SCF  $f$  is fully perfectly implementable in the direct mechanism if for every  $\mathcal{B}$ , the set of IPE-strategies of  $(\mathcal{E}, \mathcal{M}^*, \mathcal{B})$  is included in  $\Sigma^*$ .*

The following proposition follows immediately from proposition 1.

**Proposition 4** *SCF  $f$  is (fully) robustly perfect-implementable in the direct mechanism if and only if  $\mathcal{BR} \subseteq S^*$ .*

---

<sup>15</sup>Classical references are Postlewaite and Schmeidler (1988), Palfrey and Srivastava (1989) and Jackson (1991).

## 8.1 Environments with Monotone Aggregators of Information.

In this Section it is maintained that each set  $\Theta_{i,t} = [\theta_{i,t}^l, \theta_{i,t}^h] \subseteq \mathbb{R}$ , so that, for each  $t = 1, \dots, T$ ,  $Y^t \subseteq \mathbb{R}^{nt}$ . 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 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  $t = 1, \dots, T$ , there exists an aggregator function  $\alpha_i^t : Y^t \rightarrow \mathbb{R}$  and a valuation function  $v_i : \Xi^* \times \mathbb{R}^T \rightarrow \mathbb{R}$  such that  $\alpha_i^t$  and  $v_i$  are continuous,  $\alpha_i^t$  is strictly increasing in  $\theta_{i,t}$  and for each  $(\xi^*, \theta^*) \in \Xi^* \times \Theta^*$ ,*

$$u_i(\xi^*, \theta^*) = v_i\left(\xi^*, (\alpha_i^\tau(y^\tau(\theta^*)))_{\tau=1}^T\right).$$

Assuming the existence of the aggregators and the valuation functions, per se, entails no loss of generality: the bite of the representation derives from the continuity assumptions and the further restrictions on the aggregator functions that will be imposed in the following.

**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,  $s_i^c$  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,  $s_i^c$  truthfully reports  $i$ 's period- $t$  signal; at histories that come after previous misreports of agent  $i$ ,  $s_i^c$  entails a further misreport, to offset the impact on the period- $t$  aggregator of the previous misreports.<sup>16</sup> Formally:

**Definition 13** *The self-correcting strategy,  $s_i^c \in S_i$ , is such that for each  $t = 1, \dots, T$  and public history  $h^{t-1} = (\tilde{y}^{t-1}, x^{t-1})$ , and for each private history  $h_i^t = (h^{t-1}, y_i^t)$ ,*

$$s_i^c(h_i^t) = \arg \min_{m_{i,t} \in \Theta_{i,t}} \left\{ \max_{y_{-i}^t \in Y_{-i}^t} |\alpha_i^t(y_i^t, y_{-i}^t) - \alpha_i^t(\tilde{y}_i^{t-1}, m_{i,t}, y_{-i}^t)| \right\}. \quad (14)$$

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

$$\tilde{y}_{-i}^t \in \arg \max_{y_{-i}^t \in Y_{-i}^t} |\alpha_i^t(y_i^t, y_{-i}^t) - \alpha_i^t(\tilde{y}_i^{t-1}, s_i^c(h_i^t), y_{-i}^t)|.$$

---

<sup>16</sup>An earlier formulation of the idea of *self-correcting strategy* can be found in Pavan (2007). I thank Alessandro Pavan for pointing this out.

Then, by definition of  $s_i^c$  and the fact that  $\alpha_i^t$  is strictly increasing in  $\theta_{i,t}$ , we may have three cases:

$$\alpha_i^t(y_i^t, y_{-i}^t) = \alpha_i^t(\tilde{y}_i^{t-1}, s_i^c(h_i^t), y_{-i}^t) \text{ for all } y_{-i}^t \in Y_{-i}^t, \quad (15)$$

$$\alpha_i^t(y_i^t, \tilde{y}_{-i}^t) > \alpha_i^t(\tilde{y}_i^{t-1}, s_i^c(h_i^t), \tilde{y}_{-i}^t) \text{ and } s_i^c(h_i^t) = \theta_{i,t}^+, \quad (16)$$

$$\alpha_i^t(y_i^t, \tilde{y}_{-i}^t) < \alpha_i^t(\tilde{y}_i^{t-1}, s_i^c(h_i^t), \tilde{y}_{-i}^t) \text{ and } s_i^c(h_i^t) = \theta_{i,t}^-. \quad (17)$$

Equation (15) corresponds to the case in which strategy  $s_i^c$  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 (16) and (17) respectively.

**The Contraction Property.** The results on full implementation are based on a *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  $D = \times_{j \in I} D_j \subseteq S$  and for each private history  $h_i^t$ , let

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

and

$$D_i(h^{t-1}) := \bigcup_{y_i^t \in Y_i^t} D_i(h^{t-1}, y_i^t).$$

Define also:

$$\mathbf{s}_i[D_i(h^{t-1})] := \{(m_{i,t}, y_i^t) \in M_{i,t} \times Y_i^t : m_{i,t} \in D_i(h^{t-1}, y_i^t)\}$$

and

$$\mathbf{s}_i^c[h^{t-1}] := \{(m_{i,t}, y_i^t) \in M_{i,t} \times Y_i^t : m_{i,t} = s_i^c(h^{t-1}, y_i^t)\}$$

**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  $h^{t-1} = (\tilde{y}^{t-1}, x^{t-1})$  such that  $\mathbf{s}[D(h^{t-1})] \neq \mathbf{s}^c[h^{t-1}]$ , there exists  $y_i^t$  and  $m'_{i,t} \in D_i(h^{t-1}, y_i^t)$ ,  $m'_{i,t} \neq s_i^c(h^{t-1}, y_i^t)$ , such that:*

$$\text{sign}[s_i^c(h^{t-1}, y_i^t) - \theta'_{i,t}] = \text{sign}[\alpha_i^t(y_i^t, y_{-i}^t) - \alpha_i^t(\tilde{y}_i^{t-1}, \theta'_{i,t}, \theta'_{-i,t})] \quad (18)$$

for all  $y_{-i}^t = (y_{-i}^{t-1}, \theta_{-i,t}) \in Y_{-i}^t$  and  $m'_{-i,t} \in D_{-i}(h^{t-1}, y_{-i}^t)$ .

To interpret the condition, rewrite the argument of the right-hand side of (18) as follows:

$$\begin{aligned} & \alpha_i^t(y_i^t, y_{-i}^t) - \alpha_i^t(\tilde{y}^{t-1}, m'_{i,t}, m'_{-i,t}) \\ = & \left[ \alpha_i^t(\tilde{y}^{T-1}, s_i^c(h^{T-1}, y_i^T), s_{-i}^c(h^{T-1}, y_{-i}^T)) - \alpha_i^t(\tilde{y}^{t-1}, m'_{i,t}, m'_{-i,t}) \right] \\ & + \kappa(h^{t-1}, y_i^t, y_{-i}^t) \end{aligned} \quad (19)$$

where

$$\kappa(h^{t-1}, y_i^t, y_{-i}^t) = \alpha_i^t(y_i^t, y_{-i}^t) - \alpha_i^t(\tilde{y}^{t-1}, s^c(h^{t-1}, y_i^t, y_{-i}^t)) \quad (20)$$

The term in the first square bracket in (19) represents the impact, on the period- $t$  aggregator, of a deviation (in the set  $D$ ) from the self-correcting profile at history  $h^{t-1}$ ; the term  $\kappa(h^{t-1}, y_i^t, y_{-i}^t)$  represents the extent by which the self-correcting profile is incapable of offsetting the previous misreports. Suppose that  $\kappa(h^{t-1}, y_i^t, y_{-i}^t) = 0$ , i.e. strategy profile  $s^c$  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:

(*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"  $y_i^t$  of some agent  $i$ , for which for some  $m'_{i,t} \in D_i(h^{t-1}, y_i^t)$ ,  $\alpha_i^t(y_i^t, y_{-i}^t) - \alpha_i^t(\tilde{y}^{t-1}, m'_{i,t}, m'_{-i,t})$  is unilaterally signed by  $[s_i^c(h^{t-1}, y_i^t) - m'_{i,t}]$ , uniformly over the opponents private information and current reports.

From equations (15)-(17) it is easy to see that  $\kappa(h^{t-1}, y_i^t, y_{-i}^t) = 0$  whenever  $s_i^c(h_i^t) \in (\theta_{i,t}^-, \theta_{i,t}^+)$ . Hence, this corresponds precisely to the case considered in the example of Section 2. For histories such that the self-correcting strategy is not sufficient to offset the previous misreports, then the *simple CP* must be strengthened so that the sign of the impact of deviations from  $s^c$  at  $h^{t-1}$  on the aggregator  $\alpha_i^t$  is not upset by the previous misreports,  $\kappa$ . So, in principle, the bound on the interdependence in agents' valuations may depend on the histories of payoff signals. Section 8.4 though will show how simple "enlarged" mechanisms, in which agents' sets of messages are extended at every period so that any possible past misreport can be "corrected", eliminate this problem, inducing  $\kappa(h^{t-1}, y_i^t, y_{-i}^t) = 0$  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  $t$ ,  $\alpha_i^t(y^t) = \alpha_i^t(\tilde{y}^t)$  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  $a_{i,-t}^* \in \mathbb{R}^{T-1}$  and for each  $\alpha_{i,t} < \alpha'_{i,t} < \alpha''_{i,t}$ ,

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

In words: For any two allocations  $\xi$  and  $\xi'$  that only differ in their period- $t$  component, for any  $a_{i,-t}^* \in \mathbb{R}^{T-1}$ , the difference  $\delta_{i,t}(\xi, \xi', \alpha_{i,t}) = v_i(\xi, \alpha_{i,t}, a_{i,-t}^*) - v_i(\xi', \alpha_{i,t}, a_{i,-t}^*)$  as a function of  $\alpha_{i,t}$  crosses zero (at most) once (see figure 1.a, p. 27). We are now in the position to present the first full-implementation result:

**Proposition 5** 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  $\mathcal{BR} = \{s^c\}$ .

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  $t + 1, \dots, T$  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  $t = 1, \dots, T$ , 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  $(\xi^*, \theta^*) \in \Xi^* \times \Theta^*$ ,

$$u_i(\xi^*, \theta^*) = \sum_{t=1}^T v_i^t(\xi_t^*, \alpha_i^t(y^t(\theta^*))).$$

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.<sup>17</sup> These preferences though cannot accommodate phenomena of “path-dependence” such as “learning-by-doing”. For instance, in the context of the example of Section 2, suppose that agents’ preferences are the following:

$$u_i(q_1, q_2, \pi_{i,1}, \pi_{i,2}, \theta) = \alpha_{i,1}(\theta_1) \cdot q_1 + \pi_{i,1} + [\alpha_{i,2}(\theta_1, \theta_2) \cdot F(q_1) \cdot q_2 + \pi_{j,2}]. \quad (21)$$

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 (such as “learning-by-doing” effects) 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).

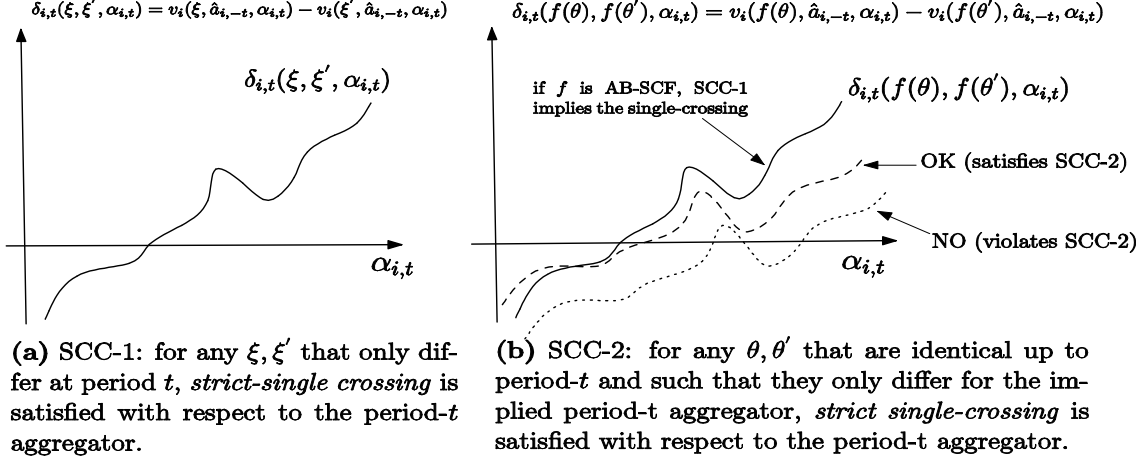
### 8.3 Relaxing “Aggregator-Based”.

In the proof of proposition 5, 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  $\exists t \in \{1, \dots, T\} : y^\tau(\theta) = y^\tau(\theta')$  for all  $\tau < t$  and for all  $j$ ,  $\alpha_j^\tau(\theta) = \alpha_j^\tau(\theta')$  for all  $\tau > t$ , then for each  $a_{i,-t}^* \in \mathbb{R}^{T-1}$  and*

---

<sup>17</sup>In that example the set of allocations includes the transfers, hence for each  $t$  the social choice function is:  $f_t(\theta) = (q_t^*(\theta), \pi_{i,t}^*(\theta), \pi_{j,t}^*(\theta))$ . 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”.



for each  $\alpha_{i,t} < \alpha'_{i,t} < \alpha''_{i,t}$ ,

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

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. The condition is easily interpretable from a graphical viewpoint: 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 “path-dependent” preferences in equation (21) satisfy SCC-2 for any choice of  $F : \mathbb{R} \rightarrow \mathbb{R}$ .)

**Proposition 6 (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  $\mathcal{BR} = \{s^c\}$ .*

**Corollary 1** *Since  $s^c \in S^*$ , if the assumptions of propositions 5 or 6, then  $f$  is fully robustly implementable.*

### 8.3.1 Transferable Utility.

A special case of interest is that of additively separable preferences with transferable utility: For each  $t = 1, \dots, T$ , 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  $\xi^* = (q_t, \pi_{1,t}, \dots, \pi_{n,t})_{t=1}^T \in \Xi^*$  and  $\theta^* \in \Theta^*$ ,

$$u_i(\xi^*, \theta^*) = \sum_{t=1}^T v_i^t((q_\tau)_{\tau=1}^t, \alpha_i^t(y^t(\theta^*))) + \pi_{i,t},$$

where for each  $t = 1, \dots, T$ ,  $v_i^t : (\times_{\tau=1}^t Q_\tau) \times \mathbb{R} \rightarrow \mathbb{R}$  is the period- $t$  valuation of the common component. Notice that functions  $v_i^t : (\times_{\tau=1}^t Q_\tau) \times \mathbb{R} \rightarrow \mathbb{R}$  are defined over the entire history  $(q_1, \dots, q_t)$ : this allows period- $t$  valuation of the current allocation  $(q_t)$  to depend on the previous allocative decisions  $(q_1, \dots, q_{t-1})$ . This allows us to accommodate the “path dependencies” in preferences discussed above.<sup>18</sup>

In environments with transferable utility, it is common to define a social choice function only for the common component,  $\chi_t : Y^t \rightarrow Q_t$  ( $t = 1, \dots, T$ ), while transfer schemes  $\pi_{i,t} : Y^t \rightarrow \mathbb{R}$  ( $i = 1, \dots, n$  and  $t = 1, \dots, T$ ) 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 function and transfer scheme  $(\chi_t, (\pi_{i,t})_{i=1}^n)_{t=1}^T$  trivially induces a social choice function  $f_t^{\chi, \pi} : Y^t \rightarrow \Xi_t$  ( $t = 1, \dots, T$ ) in the setup above: for each  $t$  and  $y^t \in Y^t$ ,  $f_t^{\chi, \pi}(y^t) = (\chi_t(y^t), (\pi_{i,t}(y^t))_{i=1}^n)$ .

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  $(\pi_{i,t}(y^t))_{i=1}^n$ , 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  $(\pi_{i,t}(y^t))_{i=1}^n$ .

Given this, the following corollary of proposition 6 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; (ii) the single crossing condition (SCC-2); and (iii) there exist transfers  $\pi$  that make  $\chi$  strictly ex-post incentive compatible; then  $f^{\chi, \pi}$  is fully robustly implemented.*

<sup>18</sup>The special case of “path-independent” preferences corresponding to the example in section 2 is such that period- $t$  valuation are functions  $v_i^t : Q_t \times \mathbb{R} \rightarrow \mathbb{R}$ .

## 8.4 “Quasi-direct” Mechanisms.

This section shows how simple “enlarged” mechanisms may avoid incurring into corner solutions, which allows us to relax the bite of the contraction property (definition 14) by guaranteeing that the sign condition holds with  $\kappa(h^{t-1}, y^t) = 0$  at every history (equation 20), thus weakening the sufficient condition for full implementation.

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

$$m_{i,t}^+ = \max \left\{ m_i \in \mathbb{R} : \max_{(y_i^t, y_{-i}^t) \in Y^t} \left| \hat{\alpha}_{i,t}(y_i^t, y_{-i}^t) - \min_{\hat{y}^{t-1} \in \hat{Y}_i^{t-1}} \hat{\alpha}_{i,t}(\hat{y}^{t-1}, m_i, y_{-i}^t) \right| = 0 \right\}$$

$$m_{i,t}^- = \min \left\{ m_i \in \mathbb{R} : \max_{(y_i^t, y_{-i}^t) \in Y^t} \left| \hat{\alpha}_{i,t}(y_i^t, y_{-i}^t) - \max_{\hat{y}^{t-1} \in \hat{Y}_i^{t-1}} \hat{\alpha}_{i,t}(\hat{y}^{t-1}, m_i, y_{-i}^t) \right| = 0 \right\}$$

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  $h_i^t = (h^{t-1}, y_i^t)$ , the self-correcting report  $s_i^c(h_i^t)$  satisfies equation (15), that is  $s^c$  is capable of fully offset previous misreports: messages in  $\hat{\Theta}_{i,t} \setminus \Theta_{i,t}$  are used whenever equations (16) or (17) 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  $\hat{y}^t \in \hat{Y}^t$  such that some message in  $\hat{\Theta}_{i,t} \setminus \Theta_{i,t}$  has been reported at some period  $\tau \leq t$ , let  $g_t(\hat{y}^t) = f_t(\theta)$  for some  $\theta$  such that  $\alpha_{i,\tau}(\theta) = \alpha_{i,\tau}(\hat{y}^\tau)$  for all  $i$  and  $\tau \leq t$ ,  $f_t(\theta) = f_t(\theta')$ .

## 9 Further Remarks on the Solution Concepts.

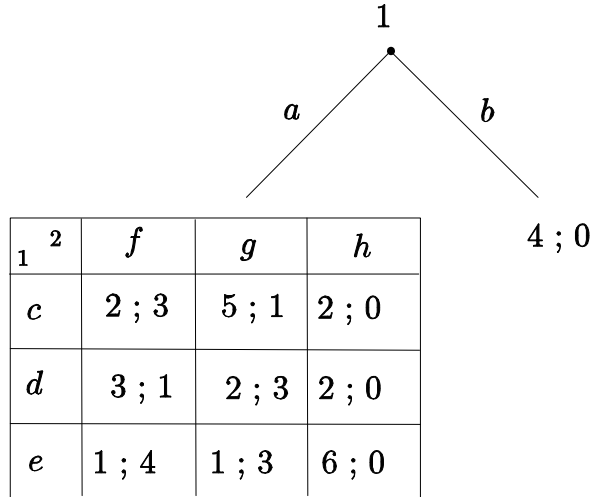
**Backwards procedure, Subgame-Perfect Equilibrium and IPE.** In games with complete and perfect information, the “backwards procedure”  $\mathcal{R}^\phi$  coincides with the backward induction solution, hence with subgame perfection.<sup>19</sup> The next example (borrowed from Perea, 2009) shows that if the game has complete but imperfect information, strategies played in Subgame-Perfect Equilibrium (SPE) may be a strict subset of  $\mathcal{R}^\phi$ :

### Example 1

---

<sup>19</sup>For the special case of games with complete information, Perea (2009) independently introduced a procedure that is equivalent to  $\mathcal{R}^\phi$ , and showed that  $\mathcal{R}^\phi$  coincides with the backward induction solution if the game has perfect information.

Consider the game in the following figure:



$\mathcal{R}_1^\phi = \{bc, bd, ac\}$  and  $\mathcal{R}_2^\phi = \{f, g\}$ . The game though has only one SPE, in which player 1 chooses  $b$ : in the proper subgame, the only Nash equilibrium entails the mixed (continuation) strategies  $\frac{1}{2}c + \frac{1}{2}d$  and  $\frac{3}{4}f + \frac{1}{4}g$ , yielding a continuation payoff of  $\frac{11}{4}$  for player 1. Hence, player 1 chooses  $b$  at the first node.  $\square$

In games with complete information, IPE coincides with SPE, but  $\mathcal{R}^\phi$  in general is weaker than subgame perfection. At first glance, this may appear in contradiction with propositions 1 and 2, which imply that  $\mathcal{R}^\phi$  characterizes the set of strategies played in IPE across models of beliefs. The reason is that even if the environment has no payoff uncertainty ( $\Theta^*$  is a singleton), the complete information model in which  $B_i$  is a singleton for every  $i$  is not the only possible: models with *redundant types* may exist, for which IPE strategies differ from the SPE-strategies played in the complete information model. The source of the discrepancy is analogous to the one between Nash equilibrium and subjective correlated equilibrium (Aumann, 1974). We illustrate the point constructing a model of beliefs and an IPE in which strategy  $(ac)$  is played with positive probability by some type of player 1.<sup>20</sup> Let payoffs be the same as in example 1, and consider the model  $\mathcal{B}$  such that  $B_1 = \{b_1^{bc}, b_1^{bd}, b_1^{ac}\}$  and  $B_2 = \{b_2^f, b_2^g\}$ , with the following beliefs:

$$\beta_1(b_1) \left[ b_2^f \right] = \begin{cases} 1 & \text{if } b_1 = b_1^{bc}, b_1^{ac} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\beta_2(b_2^g) \left[ b_1^{ad} \right] = 1, \beta_2(b_2^f) \left[ b_1^{bc} \right] = 1$$

---

<sup>20</sup>It is easy to see that such difference is not merely due to the possibility of zero-probability types. Also the relaxation of the common prior assumption is not crucial.

The equilibrium strategy profile  $\sigma$  is such that  $\forall i, \forall b_i, \sigma_i(b_i^{s_i}) = s_i$ . The system of beliefs agrees with the model's beliefs at the initial history, hence the beliefs of types  $b_2^g$  and  $b_1^{ac}$  are uniquely determined by Bayesian updating. For types  $b_i^{s_i} \neq b_2^g, b_1^{ac}$ , it is sufficient to set  $p_i(b_i^{s_i}, a_i) = \beta_i(b_i^{s_i})$  (i.e. maintain whatever the beliefs at the beginning of the game were) Then, it is easy to verify that  $(\theta, p)$  is an IPE.

On the other hand, if  $|\Theta^*| = 1$  and the game has *perfect* information (no stage with simultaneous moves), then  $\mathcal{R}^\phi$  coincides with the set of SPE-strategies. Hence, in environments with no payoff uncertainty and with perfect information, only SPE-strategies are played in IPE for any model of beliefs.

## 10 Concluding Remarks.

**On the Solution Concepts.** Proposition 1 can be seen as the dynamic counterpart of Brandenburger and Dekel's (1987) characterization of correlated equilibrium. As discussed above, the adoption of IPE in this paper is motivated by the result in proposition 2, which makes the full implementation problem of Section 8 tractable. The weakness of IPE (relative to other notions of perfect Bayesian equilibrium) is key to that result: the heart of proposition 2 is  $\mathcal{BR}$ 's property of "subgame consistency" (cf. footnote 14), which allows us to analyze continuation games "in isolation", in analogy with the logic of backward induction. The CCFR and UIA assumptions (p. 18) provide the epistemic underpinnings of the argument. To understand this point, it is instructive to compare  $\mathcal{BR}$  with Battigalli and Sinicalchi's (2007) *weak* and *strong* versions of *extensive form rationalizability* (EFR), which correspond respectively to the epistemic assumptions of *(initial) common certainty of rationality* (CCR) and *common strong belief in rationality* (CSBR):  $\mathcal{BR}$  is stronger than the first, and weaker than the latter. The strong version of EFR fails the property of "subgame consistency" because it is based on a forward induction logic, which inherently precludes the possibility of envisioning continuations "in isolation": by taking into account the possibility of counterfactual moves, agents may draw inferences from their opponents' past moves and refine their conjectures on the behavior in the continuation. The weak version of EFR fails "subgame consistency" for opposite reasons: an agent can make weaker predictions on the opponents' behavior in the continuation than he would make if he envisioned the continuation game "in isolation", because no restrictions on the agents' beliefs about their opponents' rationality are imposed after an unexpected history. Thus, the form of "backward induction reasoning" implicit in IPE (which generalizes the idea of subgame perfection) is based on stronger (respectively, weaker) epistemic assumptions than CCR (respectively, CSBR).

**Dynamic Mechanisms in Static Environments.** Consider an environment in which agents obtain all the relevant information before the planner has to make a decision. The designer may still have reasons to adopt a dynamic mechanism (e.g. an ascending auction).<sup>21</sup> In the context of an environment with complete information, Bergemann and Morris (2007) recently argued that dynamic mechanisms may improve on static ones by reducing agents' strategic uncertainty: They showed how the backward induction outcome of a second-price clock-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 clock-auction does not improve on the static one. The reason is that the logic of *backward induction* loses its bite when the assumption of complete information is relaxed.<sup>22</sup> In incomplete information environments, the case for the role of dynamic mechanisms in reducing strategic uncertainty must rely on stronger solution concepts (e.g. based on *forward induction* reasoning), that allow agents to draw stronger inferences on their opponents' private information from their past moves (see Mueller, 2009).

## 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  $h_i^t = (\theta_{i,t}, m_t, \xi_t)$ ,  $\tau < t$ , let  $\alpha_\tau(h_i^t)$  denote the triple  $(\theta_{i,\tau}, m_\tau, \xi_\tau)$  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  $h_i^t, h_i^\tau \in \mathcal{H}_i$  (w.l.o.g.: let  $\tau \geq t$ )

$$d^i(h_i^t, h_i^\tau) = \sum_{k=1}^{t-1} d_{(n_{i,k} + \nu_k + l_k)}(\alpha_k(h_i^t), \alpha_k(h_i^\tau)) + d_{n_{i,t}}(\theta_{i,t}, \theta'_{i,t}) + \sum_{k=t+1}^{\tau} 1.$$

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:

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

<sup>21</sup>In the formal setup of the paper, this amounts to a situation in which  $|\Theta_t| = 1$  for all  $t > 1$  and  $|\Xi_t| = 1$  for all  $t < T$ .

<sup>22</sup>See Kunamoto and Tercieux (2009) for a related negative result.

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  $\mu : \mathcal{A} \times \mathcal{C} \rightarrow [0, 1]$  such that:

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

**Axiom 2** *For all  $B \in \mathcal{C}$ ,  $\mu(B)$  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  $\mu(B)[A] \cdot \mu(C)[B] = \mu(C)[A]$ .*

The set of CPS on  $(\Omega, \mathcal{A}, \mathcal{C})$ , denoted by  $\Delta^{\mathcal{C}}(\Omega)$ , can be seen as a subset of  $[\Delta(\Omega)]^{\mathcal{C}}$  (i.e. mappings from  $\mathcal{C}$  to probability measures over  $(\Omega, \mathcal{A})$ ). CPS's will be written as  $\mu = (\mu(B))_{B \in \mathcal{C}} \in \Delta^{\mathcal{C}}(\Omega)$ . 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  $[\Delta(\Omega)]^{\mathcal{C}}$  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  $h_i^t = (h^{t-1}, y_i^t) \in \mathcal{H}_i$ , the corresponding event  $[h_i^t]$  is defined as:

$$[h_i^t] = \{y_i^t\} \times \left(\times_{\tau=t+1}^T \Theta_{i,\tau}\right) \times \Theta_{-i}^* \times S(h^{t-1}).$$

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 Proofs of results from Section 6.

### B.1 Proof of Proposition 1.

**Proof:**

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

$$\bar{S}_j = \{s_j \in S_j : \exists b_j \in B_j \text{ s.t. } s_j \in \text{supp}(\hat{\sigma}_j(b_j))\}.$$

We will prove that  $\bar{S}_j \subseteq \mathcal{BR}_j^M$  for every  $j$ . For each  $h_i^t = (y_i^t, h^{t-1}) \in \mathcal{H}_i$ , let  $\varphi_j^{h_i^t} : \bar{S}_j \rightarrow S_j(h_i^t)$  be a measurable function such that

$$\varphi_j^{h_i^t}(s_j)(h_j^\tau) = \begin{cases} s_j(h_j^\tau) & \text{if } \tau \geq t \\ m_j^\tau & \text{otherwise} \end{cases}$$

where  $m_j^\tau$  is the message (action) played by  $j$  at period  $\tau < t$  in the public history  $h^{t-1}$ . 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

$$L_{h_i^t}(\theta, b_{-i}, s_{-i}) = \left( \theta, \hat{b}_i, b_{-i}, \varphi_i^{h_i^t}(\hat{\sigma}_i(\hat{b}_i)), \varphi_{-i}^{h_i^t}(s_{-i}) \right).$$

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

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

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

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

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

(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, appendix A.2)

Define the CPS  $\mu^i \in \Delta^{\mathcal{H}_i}(\Theta^* \times S)$  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,  $\bar{S}_j \subseteq \mathcal{BR}_j^1$  for every  $j$ . To complete the proof, let (as inductive hypothesis)  $\bar{S}_j \subseteq \mathcal{BR}_j^k$  for every  $j$ . Then  $\mu^i$  constructed above satisfies  $\mu^i(\phi) \subseteq \Theta^* \times \{\hat{s}_i\} \times \mathcal{BR}_{-i}^k$  and

$$\begin{aligned} & \text{supp}(\text{marg}_{S|h^{t-1}} \mu^i(\phi)) \\ &= \text{supp}(\text{marg}_{S|h^{t-1}} \mu^i(h_i^t)) \\ &\subseteq \bar{S}|h^{t-1}. \end{aligned}$$

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  $\hat{\sigma}_i : B_i \rightarrow S_i$  be the identity map. Define the map  $M_{i,\phi} : \Theta^* \times S \rightarrow \Theta^* \times B_{-i}$  s.t.

$$M_{i,\phi}(\theta, s_i, s_{-i}) = (\theta, \hat{\sigma}_{-i}^{-1}(s_{-i}))$$

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

1.  $\hat{s}_i \in r_i(\mu^{\hat{s}_i})$
2. for all  $h_i^t \in \mathcal{H}_i$ :  $s_j \in \text{supp}\left(\text{marg}_{S_j} \mu^{\hat{s}_i}(h_i^t)\right)$   
 $\Rightarrow \exists s'_j \in \mathcal{BR}_j : s_j | h^{t-1} = s'_j | h^{t-1}$ .

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

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

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

$$\beta_i(b_i)[E] = \mu^{\hat{\sigma}_i(b_i)}(\phi) [M_{i, \phi}^{-1}(E)]$$

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

$$\hat{p}_i(h_i^t)[E] = \mu^{\hat{\sigma}_i(b_i)}(h_i^t) \left[ M_{\hat{\sigma}_i(b_i), h_i^t}^{-1}(E) \right].$$

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

## B.2 The backwards procedure.

Fix a public history of length  $T-1$ ,  $h^{T-1}$ . For each  $k = 0, 1, \dots$ , let  $\mathcal{R}_i^{k, h^{T-1}} \subseteq S_i^{h^{T-1}}$  be such that  $(y_i^T, s_i^{h_i^T}) \in \mathcal{R}_i^{k, h^{T-1}}$  if and only if  $s_i^{h_i^T} \in \mathcal{R}_i^{k, h^{T-1}}(y_i^T)$ ,  $\mathcal{R}^{k, h^{T-1}} = \times_{i \in N} \mathcal{R}_i^{k, h^{T-1}}$  and  $\mathcal{R}_{-i}^{k, h^{T-1}} = \times_{j \neq i} \mathcal{R}_j^{k, h^{T-1}}$ . For each  $h_i^T = (h^{T-1}, y_i^T) \in Y_i^T$ , let  $\mathcal{R}_i^{0, h^{T-1}}(y_i^T) = S_i^{h_i^T}$  and for  $k = 1, 2, \dots$ , for each  $\tilde{y}_i^T \in Y_i^T$  let

$$\begin{aligned} \mathcal{R}_i^{k, h^{T-1}}(\tilde{y}_i^T) &= \left\{ s_i \in \mathcal{R}_i^{k-1, h^{T-1}}(\tilde{y}_i^T) : \exists \pi^{h_i^T} \in \Delta(\Theta^* \times S_{-i}^{h^{T-1}}) \right. \\ &\quad 1. \pi^{h_i^T}(\{\tilde{y}_i^T\} \times \Theta_{-i}^* \times \mathcal{R}_{-i}^{k-1, h^{T-1}}) = 1 \\ &\quad 2. \text{ for all } s' \in S_i^{h^{T-1}} : \\ &\quad \int_{(\theta, s_{-i}) \in \Theta^* \times S_{-i}^{h^{T-1}}} U_i(s_i, s_{-i}, \theta; h^{T-1}) \cdot d\pi^{h_i^T} \\ &\quad \left. \geq \int_{(\theta, s_{-i}) \in \Theta^* \times S_{-i}^{h^{T-1}}} U_i(s'_i, s_{-i}, \theta; h^{T-1}) \cdot d\pi^{h_i^T} \right\} \end{aligned}$$

and  $\mathcal{R}_i^{h^{T-1}}(\tilde{y}_i^t) = \bigcap_{k=1}^{\infty} \mathcal{R}_i^{k, h^{T-1}}(\tilde{y}_i^t)$ .

Notice that  $\mathcal{R}_i^{h^{T-1}}$  consists of pairs of types  $y_i^T$  and continuation strategies  $s_i \in S_i^{(h^{T-1}, y_i^T)}$ . Hence, each  $\mathcal{R}_i^{h^{T-1}}$  can be seen as a subset of  $S_i^{h^{T-1}}$ .

For each  $t = 1, \dots, T-1$ , for each  $h_i^t = (h^{t-1}, y_i^t)$  let:

$$\mathcal{R}_i^{0, h^{t-1}}(y_i^t) = \left\{ s_i \in S_i^{h_i^t} : \forall h^t \text{ s.t. } h^{t-1} \prec h^t, \right. \\ \left. \forall y_i^{t+1} \text{ s.t. } \succ y_i^t \prec y_i^{t+1}, s_i | (h^t, y_i^{t+1}) \in \mathcal{R}_i^{h^t}(y_i^{t+1}) \right\}$$

and for each  $k$ ,  $(y_i^t, s_i^{h_i^t}) \in \mathcal{R}_i^{k, h^{t-1}}$  if and only if  $s_i^{h_i^t} \in \mathcal{R}_i^{k, h^{t-1}}(y_i^t)$ . For each  $k = 1, 2, \dots$  and for each  $k = 1, 2, \dots$

$$\mathcal{R}_i^{k, h^{t-1}}(\tilde{y}_i^t) = \left\{ s_i \in \mathcal{R}_i^{k-1, h^{t-1}}(\tilde{y}_i^t) : \exists \pi^{h_i^t} \in \Delta(\Theta^* \times S_{-i}^{h^{t-1}}) \right. \\ \left. \begin{aligned} &1. \pi \left( \{\tilde{y}_i^t\} \times \left( \times_{\tau=t+1}^T \Theta_\tau \right) \times \Theta_{-i}^* \times \mathcal{R}_{-i}^{k-1, h^{T-1}} \right) = 1 \\ &2. \text{ for all } s' \in S_i^{h_i^t} : \\ &\int_{(\theta, s_{-i}) \in \Theta^* \times S_{-i}^{h^{t-1}}} U_i(s_i, s_{-i}, \theta; h^{t-1}) \cdot d\pi^{h_i^t} \\ &\geq \int_{(\theta, s_{-i}) \in \Theta^* \times S_{-i}^{h^{t-1}}} U_i(s'_i, s_{-i}, \theta; h^{t-1}) \cdot d\pi^{h_i^t} \end{aligned} \right\}$$

$$\text{and } \mathcal{R}_i^{h^{t-1}}(\tilde{y}_i^t) = \bigcap_{k=1}^{\infty} \mathcal{R}_i^{k, h^{t-1}}(\tilde{y}_i^t).$$

Finally:  $\mathcal{R}_i^\phi = \left\{ s_i \in S_i : s_i | y_i^1 \in \mathcal{R}_i^\phi(y_i^1) \text{ for each } y_i^1 \in Y_i^1 \right\}$ .

**Proposition 2.**  $\mathcal{BR}_i = \mathcal{R}_i^\phi$  for each  $i$ .

**Proof:**

**Step 1** ( $\mathcal{R}_i^\phi \subseteq \mathcal{BR}_i$ ): let  $\hat{s}_i \in \mathcal{R}_i^\phi$ . Then, for each  $h_i^t = (h^{t-1}, y_i^t)$ ,  $s_i | h_i^t \in \mathcal{R}_i^{h^{t-1}}(y_i^t)$  (equivalently:  $s_i^{h^{t-1}} \in \mathcal{R}^{h^{t-1}}$ ). Notice that for each  $h^{t-1}$  and  $s_i^{h^{t-1}} \in \mathcal{R}_i^{h^{t-1}}$ , there exists  $s_i \in \mathcal{R}_i^\phi$  such that  $s_i | h^{t-1} = s_i^{h^{t-1}}$ . Thus, for each  $j$  and  $h^{t-1}$ , we can define measurable functions  $\rho_j^{h^{t-1}} : \mathcal{R}_j^{h^{t-1}} \rightarrow \mathcal{R}_j^\phi$  such that:  $\forall s_j^{h^{t-1}} \in \mathcal{R}_j^{h^{t-1}}$

$$\rho_j^{h^{t-1}}(s_j^{h^{t-1}}) | h^{t-1} = s_j^{h^{t-1}}.$$

(Functions  $\rho_j^{h^{t-1}}$  assign to strategies in  $\mathcal{R}_j^{h^{t-1}}$ , strategies in  $\mathcal{R}_j^\phi$  with the same continuation from  $h^{t-1}$ .) As usual, denote by  $\rho_{-i}^{h^{t-1}}$  the product  $\times_{j \neq i} \rho_j^{h^{t-1}}$ .

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

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

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

For each  $h^{t-1}$ , define the measurable mapping  $\varrho_{-i}^{h^{t-1}} : \mathcal{R}_{-i}^{h^{t-1}} \rightarrow S_{-i}(h^{t-1})$  such that  $\forall s_{-i}^{h^{t-1}} \in \mathcal{R}_{-i}^{h^{t-1}}$ ,

$$\varrho_{-i}^{h^{t-1}} \left( s_{-i}^{h^{t-1}} \right) = \varphi_{-i}^{h^{t-1}} \circ \rho_{-i}^{h^{t-1}} \left( s_{-i}^{h^{t-1}} \right).$$

It will be shown that: for each  $k = 0, 1, \dots$ ,  $\hat{s}_i \in \mathcal{R}_i^{\phi, k}$  implies  $\hat{s}_i \in \mathcal{BR}_i^k$ .

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

For the inductive step, suppose that the statement is true for  $n = 0, \dots, k-1$ : Since  $\hat{s}_i \in \mathcal{R}_i^{\phi, k}$ , for each  $h_i^t = (h^{t-1}, y_i^t)$  there exists  $\pi^{h_i^t} \in \Delta \left( \Theta^* \times S_{-i}^{h^{t-1}} \right)$  that satisfy

$$\hat{s}_i | h_i^t \in \arg \max_{s'_i \in S_{-i}^{h_i^t}} \int_{\Theta^* \times S_{-i}^{h^{t-1}}} U_i \left( s'_i, s_{-i}, \theta; h^{t-1} \right) \cdot d\pi^{h_i^t},$$

and such that  $\pi^\phi \left( \Theta^* \times \mathcal{R}_{-i}^{\phi, k-1} \right) = 1$  and for all  $h_i^t \neq \phi$ ,  $\pi^{h_i^t} \left( \{y_i^t\} \times \left( \times_{\tau=t+1}^T \Theta_{i,\tau} \right) \times \Theta_{-i}^* \times \mathcal{R}_{-i}^{h^{t-1}} \right) = 1$ .

Now, consider the CPS  $\mu^i \in \Delta^{\mathcal{H}_i} \left( \Theta^* \times S \right)$  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  $\mu(h_i^t)$  for all  $h_i^t$  s.t.  $\mu^i(\phi) [h_i^t] > 0$ . Let  $h_i^t$  be such that  $\mu^i(\phi) [h_i^{t-1}] > 0$  and  $\mu^i(\phi) [h_i^t] = 0$ . Define the measurable mapping  $M_{h_i^t} : \Theta^* \times \mathcal{R}_{-i}^{h^{t-1}} \rightarrow \Theta^* \times S(h^{t-1})$  such that for all  $(\theta, s_{-i}^{h^{t-1}}) \in \Theta^* \times S(h^{t-1})$ ,

$$M_{h_i^t} \left( \theta, s_{-i}^{h^{t-1}} \right) = \left( \theta, \varphi_i^{h^{t-1}}(\hat{s}_i), \varrho_{-i}^{h^{t-1}} \left( s_{-i}^{h^{t-1}} \right) \right)$$

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

$$\mu^i(h_i^t) [E] = \pi^{h_i^t} \left[ M_{h_i^t}^{-1}(E) \right].$$

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

By construction,  $\hat{s}_i \in r_i(\mu^i)$  (condition 1 in the definition of  $\mathcal{BR}_i^k$ ). Since by construction  $\mu^i \left( \Theta^* \times \{\hat{s}_i\} \times \mathcal{R}_{-i}^{\phi, k-1}; \phi \right) = 1$ , under the inductive hypothesis  $\mu^i \left( \Theta^* \times \{\hat{s}_i\} \times \mathcal{BR}_{-i}^{k-1}; \phi \right) = 1$  (condition 2 in the definition of  $\mathcal{BR}_i^k$ ). From the definition of  $\varphi_i^{h^{t-1}}(\hat{s}_i)$ , CPS  $\mu^i$  satisfies condition (3.1) at each  $h_i^t$ . From the definition of  $\varrho_{-i}^{h^{t-1}}$ , under the inductive hypothesis,  $\mu^i$  satisfies condition (3.2).

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

$$\Psi_{h_i^t}(\theta, s_i, s_{-i}) = \left( \theta, \psi_{h_i^t}(s_{-i}) \right)$$

For each  $h_i^t \in \mathcal{H}_i$ , let  $\pi^{h_i^t} \in \Delta(\Theta^* \times S_{-i}^{h^{t-1}})$  be such that for every measurable  $E \subseteq \Theta^* \times S_{-i}^{h^{t-1}}$

$$\pi^{h_i^t}[E] = \mu^i(h_i^t) \left[ \Psi_{h_i^t}^{-1}(E) \right].$$

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

$$\hat{s}_i|_{h_i^t} \in \arg \max_{s_i \in S_i^{h_i^t}} \int U_i(s_i, s_{-i}; h_i^t) \cdot d\pi^{h_i^t}.$$

The argument proceeds by induction on the length of histories.

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

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

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

### C.1 Proof of Proposition 3

**Step 1 (If):** For the *if* part, fix an arbitrary type space  $\mathcal{B}$ , and consider a direct mechanism  $\mathcal{M}$ . Let  $(p^i)_{i \in N}$  be any beliefs system such that,  $\forall i \in N, \forall (\theta, b_{-i}) \in \Theta^* \times B_{-i}, p^i(h_i^0) = \beta_i(h_i^0)$  and for each  $h_i^t \in \mathcal{H}_i$  such that  $P^{\sigma^*, p}(h_i^{t-1})[h_i^t] > 0$ ,  $p^i(h_i^t)$  is obtained via Bayesian

updating. If instead  $h_i^t$  is such that  $P^{\sigma^*, p}(h_i^{t-1})[h_i^t] = 0$ , and  $h_i^t = (h^{t-1}, y_i^t)$  s.t.  $h^{t-1} = (\tilde{y}_i^t, \tilde{y}_{-i}^t)$ , then let beliefs be such that

$$\text{supp} \left( \text{marg}_{\Theta_{-i}^*} p^i(h_i^t) \right) \subseteq \{\tilde{y}_{-i}^t\} \times \left( \times_{\tau=t+1}^T \Theta_{-i, \tau} \right) \quad (22)$$

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  $U_i(s^*, \theta) \geq U_i(s'_i, s_{-i}^*, \theta)$  for all  $\theta$ , then for any  $p^i(\phi) \in \Delta(\Theta^* \times B_{-i})$ ,

$$\begin{aligned} & \int_{\Theta^* \times B_{-i}} u_i(\mathbf{g}^{s^*}(\theta), \theta) \cdot dp^i(\phi) \\ & \geq \int_{\Theta^* \times B_{-i}} u_i\left(\mathbf{g}^{(s'_i, s_{-i}^*)|h_i^t}(\theta), \theta\right) \cdot dp^i(\phi). \end{aligned}$$

Hence, the incentive compatibility constraints are satisfied at the beginning of the game, and so at all histories reached with positive probability according to the initial conjectures and strategy profile. Being  $\sigma^* \in \Sigma^*$ , only truthful histories receive positive probability. At zero probability histories, we maintain that the belief system satisfies (22). With these beliefs, the only payoff-relevant component of the opponents’ strategies at history  $h_i^t$  is the “truthful report”: from the point of view of player  $i$ , what  $\sigma_{-i}^*$  specifies at non-truthful histories is irrelevant. Let  $\sigma_i^*(h_i^t)$  be a best response to such beliefs and  $\sigma_{-i}^*$  in the continuation game: Notice that under these beliefs, any  $\sigma_{-i} \in \Sigma^*$  determines the same  $\sigma_i^*(h_i^t)$ . Hence, for any  $i$  we can chose  $\sigma_i^* \in \Sigma^*$  so that the strategy profile thus constructed 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 5.

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

It will be shown that for each  $t$  and for each public history  $h^{t-1}$ ,  $\mathbf{s}[B(h^{t-1})] = \mathbf{s}^c[h^{t-1}]$ , 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  $h^0$ .

**Initial Step:**  $\mathbf{s}[B(h^{T-1})] = \mathbf{s}^c[h^{T-1}]$  for each  $h^{T-1}$ .

Suppose, by contradiction, that  $\exists h^{T-1} = (\tilde{y}^{T-1}, x^{T-1}) : \mathbf{s}[B(h^{T-1})] \neq \mathbf{s}^c(h^{T-1})$ . Then, by the contraction property,

$$\begin{aligned} & \exists y_i^T \text{ and } \theta'_{i,T} \in B_i(h^{T-1}, y_i^T) : \theta'_{i,T} \neq s_i^c(h^{T-1}, y_i^T) \text{ such that:} \\ & \text{sign}[s_i^c(h^{T-1}, y_i^T) - \theta'_{i,T}] = \text{sign}[\alpha_i^T(y_i^T, y_{-i}^T) - \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,t}, \theta'_{-i,t})] \\ & \text{for all } y_{-i}^T = (y_{-i}^{T-1}, \theta_{-i,T}) \text{ and } \theta'_{-i,T} \in B_{-i}(h^{T-1}, y_{-i}^T). \end{aligned}$$

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

$$\delta(h^{T-1}, y_i^T) := \min_{\substack{y_{-i}^T \in Y_{-i}^T \text{ and} \\ \theta'_{-i,T} \in B_{-i}(h^{T-1}, y_{-i}^T)}} [\alpha_i^T(y_i^T, y_{-i}^T) - \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,t}, \theta'_{-i,t})] \quad (23)$$

(by compactness of  $Y^T$  and  $B(h^T)$ ,  $\delta(h^T, y_i^T)$  is well-defined). Also, from  $\theta'_{i,T} \neq s_i^c(h^{T-1}, y_i^T)$  and the Contraction Property,  $\delta(h^{T-1}, y_i^T) > 0$ .

For any  $\varepsilon > 0$ , let

$$\psi(h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon) = \max_{\theta_{-i,T} \in \Theta_{-i,T}} \{ \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}) - \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}) \} \quad (24)$$

(again, compactness of  $\Theta_{-i,T}$  guarantees that  $\psi(h^T, \varepsilon)$  is well-defined). Since  $\alpha_i^T$  is strictly increasing in  $\theta_{i,T}$ ,  $\psi(h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon)$  is increasing in  $\varepsilon$  and  $\psi(h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ .

Let  $(f_t(\tilde{y}^t))_{t=1}^{T-1} = x^{T-1}$ . From strict EPIC, we have that for each  $\varepsilon$ ,

$$\begin{aligned} & v_i(x^{T-1}, f_T(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), \alpha^{-T}(\tilde{y}^{T-1})) \\ & > v_i(x^{T-1}, f_T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), \alpha^{-T}(\tilde{y}^{T-1})) \end{aligned}$$

and

$$\begin{aligned} & v_i(x^{T-1}, f_T(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), \alpha^{-T}(\tilde{y}^{T-1})) \\ & < v_i(x^{T-1}, f_T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), \alpha^{-T}(\tilde{y}^{T-1})) \end{aligned}$$

Thus, by continuity, there exists  $a^T(\varepsilon)$  such that

$$\alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}) < a^T(\varepsilon) < \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}) \quad (25)$$

such that

$$\begin{aligned} & v_i(x^{T-1}, f_T(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), a^T(\varepsilon), \alpha^{-T}(\tilde{y}^{T-1})) \\ & = v_i(x^{T-1}, f_T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), a^T(\varepsilon), \alpha^{-T}(\tilde{y}^{T-1})) \end{aligned}$$

From the ‘‘within-period SCC’’ (def. 16),

$$\begin{aligned} & v_i(x^{T-1}, f_T(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), a^*, \alpha^{-T}(\tilde{y}^{T-1})) \\ & > v_i(x^{T-1}, f_T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), a^*, \alpha^{-T}(\tilde{y}^{T-1})) \end{aligned}$$

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

Thus, to reach the contradiction, it suffices to show that for any  $y_{-i}^T \in Y_{-i}^T$ ,  $\alpha_i^T(y_i^T, y_{-i}^T) > a^T(\varepsilon)$ : If this is the case, reporting  $\theta'_{i,T}$  is (conditionally) strictly dominated by reporting  $\theta'_{i,T} + \varepsilon$  at  $h_i^T = (h^{T-1}, y_i^T)$ , hence it cannot be that  $B_i = \mathcal{BR}_i$  and  $\theta'_{i,T} \in B_i(h^{T-1}, y_i^T)$ . To this end, it suffices to choose  $\varepsilon$  sufficiently small that

$$\psi(h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon) < \delta \quad (26)$$

and operate the substitutions as follows

$$\begin{aligned} \alpha_i^T(y_i^T, y_{-i}^T) &\geq \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T}, \theta'_{-i,T}) + \delta(h^{T-1}, y_i^T) \\ &\geq \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta'_{-i,T}) + \delta(h^{T-1}, y_i^T) - \psi(h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon) \\ &> \alpha_i^T(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta'_{-i,T}) \\ &> a^T(\varepsilon) \end{aligned}$$

Thus:  $\alpha_i^T(y_i^T, y_{-i}^T) > a^T(\varepsilon)$  for any  $y_{-i}^T$ . This concludes the initial step.

**Inductive Step:** [for  $t = 1, \dots, T-1$ : if  $\mathbf{s}[B(h^\tau)] = \mathbf{s}^c[h^\tau]$  for all  $h^\tau$  and all  $\tau > t$  then  $\mathbf{s}[B(h^t)] = \mathbf{s}^c[h^t]$  for all  $h^t$ ]

Suppose, by contradiction, that  $\exists h^{t-1} = (\tilde{y}^{t-1}, x^{t-1}) : \mathbf{s}[B(h^{t-1})] \neq \mathbf{s}^c(h^{t-1})$ . Then, by the contraction property,

$$\begin{aligned} &\exists y_i^t \text{ and } \theta'_{i,t} \in B_i(h^{t-1}, y_i^t) : \theta'_{i,t} \neq s_i^c(h^{t-1}, y_i^t) \text{ such that:} \\ &\text{sign}[s_i^c(h^{t-1}, y_i^t) - \theta'_{i,t}] = \text{sign}[\alpha_i^t(y_i^t, y_{-i}^t) - \alpha_i^t(\tilde{y}^{t-1}, \theta'_{i,t}, \theta'_{-i,t})] \\ &\text{for all } y_{-i}^t = (y_{-i}^{t-1}, \theta_{-i,t}) \text{ and } \theta'_{-i,t} \in B_{-i}(h^{t-1}, y_{-i}^t). \end{aligned}$$

Fix such  $y_i^t$  and  $\theta'_{i,t} \neq s_i^c(h^{t-1}, y_i^t)$ , and suppose (w.l.o.g.) that  $s_i^c(h^{t-1}, y_i^t) > \theta'_{i,t}$ . Similar to the initial step, it will be shown that there exists  $\theta_{i,t}^\varepsilon = \theta'_{i,t} + \varepsilon$  for some  $\varepsilon > 0$  such that for any conjecture consistent with  $B_{-i}$ , playing  $\theta_{i,t}^\varepsilon$  is strictly better than playing  $\theta'_{i,t}$  at history  $(h^{t-1}, y_i^t)$ , contradicting the hypothesis that  $\mathcal{BR} = B$ .

For any  $\varepsilon > 0$ , set  $\theta_{i,t}^\varepsilon = \theta'_{i,t} + \varepsilon$ ; for each realization of signals  $\tilde{\theta}_i = (\tilde{\theta}_{i,k})_{k=1}^T$  and opponents' reports  $\tilde{m}_{-i} = (\tilde{m}_{-i,k})_{k=t}^T$ , for each  $\tau > t$ , denote by  $s_{i,\tau}^c(\theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i)$  the action taken at period  $\tau$  if  $\theta_{i,t}^\varepsilon$  is played at  $t$ ,  $s_i^c$  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 of the aggregators functions,  $s_{i,\tau}^c(\theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i)$  is continuous in  $\varepsilon$ , and converges to  $s_{i,\tau}^c(\theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i)$  as  $\varepsilon \rightarrow 0$ .)

For each realization  $\tilde{\theta}_i = (\tilde{\theta}_{i,k})_{k=1}^T$  and reports  $\tilde{m}_{-i} = (\tilde{m}_{-i,k})_{k=t}^T$  and for each  $\tau > t$ ,  $s_{i,\tau}^c(\theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i)$  may be one of five cases:

1.  $s_{i,\tau}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \in (\theta_{i,T}^-, \theta_{i,T}^+)$ , then

$$\alpha_i^\tau (y_i^\tau, y_{-i}^\tau) = \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)$$

for all  $y_{-i}^\tau$ , and we can choose  $\varepsilon$  sufficiently small that  $s_{i,\tau}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \in (\theta_{i,T}^-, \theta_{i,T}^+)$ , i.e.

$$\alpha_i^\tau (y_i^\tau, y_{-i}^\tau) = \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)$$

for all  $y_{-i}^\tau$

2.  $s_{i,\tau}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^+$  and

$$\alpha_i^\tau (y_i^\tau, y_{-i}^\tau) > \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)$$

at the argmax over  $y_{-i}^\tau$ , then we can choose  $\varepsilon$  sufficiently small that  $s_{i,\tau}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^+$  as well.

3.  $s_{i,\tau}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^+$  and

$$\alpha_i^\tau (y_i^\tau, y_{-i}^\tau) = \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)$$

for all  $y_{-i}^\tau$ . Then, either  $s_{i,\tau}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^+$  as well, or  $s_{i,\tau}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \in (\theta_{i,T}^-, \theta_{i,T}^+)$ , i.e.

$$\alpha_i^\tau (y_i^\tau, y_{-i}^\tau) = \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)$$

for all  $y_{-i}^\tau$ . In either case,

$$\alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right) = \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)$$

for all  $y_{-i}^\tau$

4.  $s_{i,\tau}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^-$  and

$$\alpha_i^\tau (y_i^\tau, y_{-i}^\tau) < \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)$$

at the argmax over  $y_{-i}^\tau$ , Then we can choose  $\varepsilon$  sufficiently small that  $s_{i,\tau}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^-$  as well.

5.  $s_{i,\tau}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^-$  and

$$\alpha_i^\tau (y_i^\tau, y_{-i}^\tau) = \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)$$

for all  $y_{-i}^\tau$ . Then, either  $s_{i,\tau}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^-$  as well, or  $s_{i,\tau}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \in \left( \theta_{i,T}^-, \theta_{i,T}^+ \right)$ , i.e.

$$\alpha_i^\tau (y_i^\tau, y_{-i}^\tau) = \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right)$$

for all  $y_{-i}^\tau$ . In either case,

$$\begin{aligned} & \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right) \\ &= \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right) \end{aligned}$$

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:

$$\begin{aligned} & \text{for all } \varepsilon \in \left( 0, \bar{\varepsilon} \left( \tilde{\theta}_i, \tilde{m}_{-i}, \tau \right) \right), \text{ for all } y_{-i}^\tau \\ & \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right) \\ &= \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right). \end{aligned}$$

Let  $\bar{\varepsilon} = \min_{\tilde{\theta}_i, \tilde{m}_{-i}, \tau} \bar{\varepsilon} \left( \tilde{\theta}_i, \tilde{m}_{-i}, \tau \right)$  (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  $\varepsilon \in (0, \bar{\varepsilon})$ , reporting  $\theta_{i,t}^\varepsilon$  or  $\theta'_{i,t}$  at period  $t$  does not affect the allocation chosen at periods  $\tau > t$  (the opponents' self-correcting report cannot be affected by  $i$ -th components of the public history). Hence, for  $\varepsilon \in (0, \bar{\varepsilon})$ , for each  $\theta_{-i} \in \Theta_{-i}^*$ , the allocations induced following  $s_i^c$  at periods  $\tau > t$  and playing  $\theta'_{i,t}$  or  $\theta_{i,t}^\varepsilon$  at history  $h_{i,t}^t$ , respectively  $\xi'$  and  $\xi^\varepsilon$ , are such that  $\xi'_\tau = \xi^\varepsilon_\tau$  for all  $\tau \neq t$ .

Consider types of player  $i$ ,  $\theta'_i, \theta_i^\varepsilon \in \Theta_i^*$  such that for each  $\tau < t$ ,  $\theta'_{i,\tau} = \theta_{i,\tau}^\varepsilon = \hat{\theta}_{i,\tau}$  (the one actually reported on the path), for all  $\tau > t$  and  $\theta_{i,\tau} = s_{i,\tau}^c$  as above, while at  $t$  respectively equal to  $\theta_{i,t}^\varepsilon$  and  $\theta'_{i,t}$ . Thus, the induced allocations are  $\xi^\varepsilon$  and  $\xi'$  discussed above, and for each  $\tau \neq t$ ,  $\alpha_i^\tau (\theta^\varepsilon) = \alpha_i^\tau (\theta') \equiv \hat{\alpha}_i^\tau$ .

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

$$\begin{aligned} v_i \left( \xi^\varepsilon, \alpha^t(\theta^\varepsilon), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) &> v_i \left( \xi', \alpha^t(\theta^\varepsilon), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) \\ &\text{and} \\ v_i \left( \xi^\varepsilon, \alpha^t(\theta'), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) &< v_i \left( \xi', \alpha^t(\theta'), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) \end{aligned}$$

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

$$\begin{aligned} \alpha_i^t(\tilde{y}^{t-1}, \theta'_{i,t}, \theta_{-i,t}) &< a^t(\varepsilon) < \alpha_i^t(\tilde{y}^{t-1}, \theta_{i,t}^\varepsilon, \theta_{-i,t}) \\ &\text{such that} \\ v_i \left( \xi^\varepsilon, a^t(\varepsilon), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) &= v_i \left( \xi', a^t(\varepsilon), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) \end{aligned} \tag{27}$$

From the Single Crossing Condition,

$$\begin{aligned} v_i \left( \xi^\varepsilon, a^*, \{\hat{a}_i^\tau\}_{\tau \neq t} \right) &> v_i \left( \xi', a^*, \{\hat{a}_i^\tau\}_{\tau \neq t} \right) \\ &\text{whenever } a^* > a^t(\varepsilon). \end{aligned}$$

Thus, since the continuations in periods  $\tau > t$  are the same under both  $\theta'_{i,t}$  and  $\theta_{i,t}^\varepsilon$ , 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  $\tilde{\theta}_{-i}$ ).

As in the initial step, define:

$$\delta := \min_{\substack{y_{-i}^t \in Y_{-i}^t \text{ and} \\ \theta'_{-i,t} \in B_{-i}(h^{t-1}, y_{-i}^t)}} [\alpha_i^t(y_i^t, y_{-i}^t) - \alpha_i^t(\tilde{y}^{t-1}, \theta'_{i,t}, \theta'_{-i,t})] \tag{28}$$

For any  $\varepsilon > 0$ , let

$$\psi(\varepsilon) = \max_{\theta_{-i,t} \in \Theta_{-i,t}} \left\{ \alpha_i^t(\tilde{y}^{t-1}, \theta_{i,t}^\varepsilon, \theta_{-i,t}) - \alpha_i^t(\tilde{y}^{t-1}, \theta'_{i,t}, \theta_{-i,t}) \right\} \tag{29}$$

Since  $\alpha_i^t$  is strictly increasing in  $\theta_{i,t}$ ,  $\psi(\varepsilon)$  is increasing in  $\varepsilon$  and  $\psi(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ .

To obtain the desired contradiction, it suffices to choose  $\varepsilon$  sufficiently small that

$$\psi(\varepsilon) < \delta \tag{30}$$

and operate the substitutions as follows

$$\begin{aligned} \alpha_i^t(y_i^t, y_{-i}^t) &\geq \alpha_i^t(\tilde{y}^{t-1}, \theta'_{i,t}, \theta'_{-i,t}) + \delta \\ &\geq \alpha_i^t(\tilde{y}^{t-1}, \theta_{i,t}^\varepsilon, \theta'_{-i,t}) + \delta - \psi(\varepsilon) \\ &> \alpha_i^t(\tilde{y}^{t-1}, \theta_{i,t}^\varepsilon, \theta'_{-i,t}) \\ &> a^t(\varepsilon). \end{aligned}$$

### C.3 Proof of Proposition 6.

The proof is very similar to those of proposition 5.

**Initial Step:**  $[s [B (h^{T-1})]] = s^c [h^{T-1}]$  for each  $h^{T-1}$ .

The initial step is the same, to conclude (in analogy with equation 25), that there exists  $a^T (\varepsilon)$  such that

$$\alpha_i^T (\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}) < a^T (\varepsilon) < \alpha_i^T (\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}) \quad (31)$$

such that

$$\begin{aligned} & v_i (x^{T-1}, f_T (\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), a^T (\varepsilon), \alpha^{-T} (\tilde{y}^{T-1})) \\ &= v_i (x^{T-1}, f_T (\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), a^T (\varepsilon), \alpha^{-T} (\tilde{y}^{T-1})) \end{aligned} \quad (32)$$

From the ‘‘Strengthened SCC’’ (def. 17),

$$\begin{aligned} & v_i (x^{T-1}, f_T (\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}), a^*, \alpha^{-T} (\tilde{y}^{T-1})) \\ & > v_i (x^{T-1}, f_T (\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}), a^*, \alpha^{-T} (\tilde{y}^{T-1})) \end{aligned}$$

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

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

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

The argument proceeds as in proposition 5, to show that for each  $\tau > t$ , and for each  $(\tilde{\theta}, \tilde{m}_{-i})$ , if continuation strategies are self-correcting, there exists  $\bar{\varepsilon} (\tilde{\theta}, \tilde{m}_{-i}, \tau) > 0$  such that:

$$\begin{aligned} & \text{for all } \varepsilon \in (0, \bar{\varepsilon} (\tilde{\theta}, \tilde{m}_{-i}, \tau)), \\ & \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right) \\ &= \alpha_i^\tau \left( \tilde{y}_i^{t-1}, \theta'_{i,t}, \left( s_{i,k}^c \left( \theta'_{i,t}, \tilde{m}_{-i}, \tilde{\theta}_i \right) \right)_{k=t+1}^\tau, y_{-i}^\tau \right) \text{ for all } y_{-i}^\tau. \end{aligned}$$

Consider types of player  $i$ ,  $\theta'_i, \theta_i^\varepsilon \in \Theta_i^*$  such that for each  $\tau < t$ ,  $\theta'_{i,\tau} = \theta_{i,\tau}^\varepsilon = \hat{\theta}_{i,\tau}$  (the one actually reported on the path), for all  $\tau > t$  and  $\theta_{i,\tau} = s_{i,\tau}^c$  as above, while at  $t$  respectively equal to  $\theta_{i,t}^\varepsilon$  and  $\theta'_{i,t}$ . By construction, such types are such that for any  $\tau \neq t$ ,  $\alpha_i^\tau (\theta^\varepsilon) = \alpha_i^\tau (\theta')$ .

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

$$\begin{aligned} & v_i \left( f (\theta^\varepsilon), \alpha^t (\theta^\varepsilon), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) > v_i \left( f (\theta'), \alpha^t (\theta^\varepsilon), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) \\ & \text{and} \\ & v_i \left( f (\theta^\varepsilon), \alpha^t (\theta'), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) < v_i \left( f (\theta'), \alpha^t (\theta'), \{\hat{a}_i^\tau\}_{\tau \neq t} \right) \end{aligned}$$

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

$$\alpha_i^t(\tilde{y}^{t-1}, \theta'_{i,t}, \theta_{-i,t}) < a^t(\varepsilon) < \alpha_i^t(\tilde{y}^{t-1}, \theta^\varepsilon_{i,t}, \theta_{-i,t}) \quad (33)$$

such that

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

From the single crossing condition,

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

whenever  $a^* > a^t(\varepsilon)$

To reach the desired contradiction it suffices to show that for any  $y_{-i}^t \in Y_{-i}^t$ ,  $\alpha_i^t(y_i^t, y_{-i}^t) > a^t(\varepsilon)$ . The remaining part of the proof is identical to proposition 5.  $\square$

## References

1. AUMANN, R. (1974), "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67-96.
2. ATHEY, S. AND I. SEGAL (2007), "An Efficient Dynamic Mechanism" *mimeo*, Stanford Univ.
3. BATTIGALLI, P. (2003), "Rationalizability in infinite, dynamic games with incomplete information", *Research in Economics* 57 (2003) 1-38
4. BATTIGALLI, P. AND M. SINISCALCHI (2003), "Rationalization and Incomplete Information", *Advances in Theoretical Economics*, **3**, Article 3.
5. BATTIGALLI, P. AND M. SINISCALCHI (2007), "Interactive Epistemology in Games with Payoff Uncertainty" *Research in Economics*, 61, 165-184.
6. BEN-PORATH, E. (1997), "Rationality, Nash Equilibrium and Backwards Induction in Perfect Information Games," *Review of Economic Studies*, **64**, 23-46.
7. BERGEMANN, D. AND S. MORRIS (2005) "Robust Mechanism Design", *Econometrica* 73(6), 1771-1813.
8. BERGEMANN, D. AND S. MORRIS (2007) "An Ascending Auction for Interdependent Values: Uniqueness and Robustness to Strategic Uncertainty", *AEA papers and proceedings*, May 2007.
9. BERGEMANN, D. AND S. MORRIS (2009) "Robust Implementation in Direct Mechanisms", *Review of Economic Studies*,

10. BERGEMANN, D. AND J. VALIMAKI (2008) "The Dynamic Pivot Mechanism"
11. BIKHCHANDANI (2006), "Ex Post Implementation in Environments with Private Goods", *Theoretical Economics*, 1, 369-393.
12. BRANDENBURGER, A. AND E. DEKEL (1987), "Rationalizability and Correlated Equilibria", *Econometrica*, 55, 1391-1402.
13. ESO, P. AND E. MASKIN (2002) "Multi-Good Efficient Auctions with Multidimensional Information", Discussion paper, Northwestern Univ. and Institute for Advanced Studies.
14. FUDENBERG, D. AND J. TIROLE (1991), "Perfect Bayesian Equilibrium and Sequential Equilibrium", *Journal of Economic Theory*, 53, 236-290.
15. GERSHOV AND B. MOLDOVANU (2009a) "Learning About the Future and Dynamic Efficiency" *American Economic Review*, forthcoming
16. GERSHOV AND B. MOLDOVANU (2009b) "Optimal Search, Learning and Implementation", *mimeo*, Univ. of Bonn.
17. HARSANYI, J. (1967-68), "Games of Incomplete Information Played by Bayesian Players. Parts I, II, III," *Management Science*, 14, 159-182, 320-334, 486-502.
18. HARSANYI, J. AND R. SELTEN (1988), *A general Theory of Equilibrium Selection in Games*, MIT Press, Cambridge, MA.
19. JACKSON, M. (1991), "Bayesian Implementation", *Econometrica* 59, 461-477.
20. JEHIEL, P., M. MEYER-TER-VEHN, B. MOLDOVANU AND W. ZAME (2006) "The Limits of Ex-Post Implementation", *Econometrica*, 74, 585-610.
21. KUNIMOTO, T. AND O. TERCIEUX (2009) "Implementation with Near-Complete Information: The Case of Subgame Perfection", *mimeo*, PSE.
22. NEEMAN, Z. (2004) "The Relevance of Private Information in Mechanism Design," *Journal of Economic Theory* 117 (2004), 55-77.
23. MUELLER, C. (2009) "Robust Virtual Implementation under Common Strong Belief in Rationality" *mimeo*, University of Minnesota.
24. PALFREY, T. AND SRIVASTAVA (1989) "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem", *Journal of Political Economy*, 97, 668-691.

25. PAVAN, A. (2007) “Long-term Contracting in a Changing World”, *mimeo*, Northwestern Univ.
26. PAVAN, A., Y SEGAL AND J. TOIKKA (2009) “Dynamic Mechanism Design” *mimeo*, Stanford Univ.
27. PEARCE, D. (1984), “Rationalizable Strategic Behavior and the Problem of Perfection,” *Econometrica*, **52**, 1029-1050.
28. PENTA, A. (2009), “Strong Interim Perfect Equilibrium: A technical appendix to ‘Robust Dynamic Mechanism Design’”, *mimeo*, UPenn.
29. PEREA, A. (2009), “Belief in the Opponents’Future Rationality”, *mimeo*, Maastricht University.
30. PICKETTY, T. (1999) “The Information-Aggregation Approach to Political Institutions”, *European Economic Review*, 43, 791-800.
31. POSTLEWAITE, A. AND D. SCHMEIDLER (1988) “Implementation in Differential Information Economies”, *Journal of Economic Theory*, 39, 14-33.
32. WILSON, R. (1987), “Game-Theoretic Analyses of Trading Processes” in *Advances in Economic Theory: Fifth World Congress*, ed. by T. Bewley. Cambridge, U.K.: Cambridge Univ. Press, Chap.2, 33-70.