Robust Dynamic Implementation

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

March 10, 2015

Abstract

This paper extends the belief-free approach to robust mechanism design to dynamic environments, in which agents obtain information over time. A social choice function (SCF) is *robustly partially implemented* if it is perfect Bayesian incentive compatible for all possible beliefs. It is shown that this is possible if and only if the SCF is ex-post incentive compatible. *Robust full Implementation* imposes the stronger condition that, for all possible beliefs, all the Perfect Bayesian Equilibria induce outcomes consistent with the SCF. Characterizing the set of such equilibria is a key difficulty for studying this problem. This paper shows that, for a weaker notion of equilibrium, the set of all such equilibria can be computed by means of a recursive procedure which combines the logic of rationalizability and backward induction reasoning. These results are then used to show that, in environments with single crossing preferences and well-behaved intertemporal effects, strict ex-post incentive compatibility and a condition which limits the strength of preference interdependencies are sufficient to guarantee robust full implementation.

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

JEL Codes: C72; C73; D82.

1 Introduction

A common criticism to classical theory of mechanism design is that it relies on strong common knowledge assumptions that are unlikely to be satisfied in reality. This viewpoint, often referred to as the 'Wilson doctrine', has recently been revived by a series of papers by Bergemann and Morris (2005, 2009a,b, 2011), which spurred a growing literature on robust mechanism design. Bergemann and Morris' seminal work pursued a 'belief free' approach, investigating conditions under which (full, partial or virtual) implementation can be achieved independent

*[*email*: apenta@ssc.wisc.edu]. Previous versions of this paper circulated under the title "Robust Dynamic Mechanism Design". I am indebted to my advisor, George Mailath, for his unique dedication and for constantly challenging me throughout this project. I thank Andrew Postlewaite, Qingmin Liu, Dirk Bergemann, Stephen Morris, Alessandro Pavan, Marzena Rostek, Bill Sandholm and the anonymous referees for their helpful comments. Financial support from the UW-Madison Graduate School is gratefully acknowledged.

of the agents' beliefs. More recently, alternative and less demanding approaches to robustness have also been proposed.[1] This literature, however, has only considered static settings. In contrast, many situations of economic interest present problems of mechanism design that are inherently dynamic. For instance, consider the problem of a social planner who wants to assign licenses for the provision of a public good to the most productive firm in each period. Firms' productivity is their private information and may change over time. Furthermore, productivity in earlier periods may be informative about later productivity, and later productivity may depend on earlier allocative choices as well (for example, if there is learning-by-doing). In these situations, static mechanisms may not suffice to guarantee a socially desirable outcome, and it is important for the planner to take into account the firms' intertemporal incentives. Moreover, problems of this kind cannot be cast within the framework received by the literature on robust mechanism design, which assumes that agents obtain all the relevant information before the mechanism is set in place.

The present paper extends the belief-free approach to study partial and full implementation in problems of dynamic mechanism design. 'Dynamic mechanism design' can be understood in two ways: first, as the study of dynamic mechanisms (e.g., an ascending auction) in standard (static) environments; second, as the study of mechanism design problems in environments that are inherently dynamic, such as the example above. Our analysis applies to both cases and innovates on the previous literature along both dimensions.[2]

The belief-free approach is often criticized for being excessively demanding. From a theoretical viewpoint, however, this approach represents an important benchmark, particularly to analyze the methodological aspects of robust mechanism design. The insights received from the static literature, for instance, have been adapted to less demanding notions of robustness, which may be more appealing from an applied viewpoint (cf. Section 6). To what extent the methodology developed for static problems can be extended to dynamic settings, however, is not clear, particularly for the study of the full implementation problem. The belief-free approach therefore is a natural starting point to extend the theory of robust mechanism design to dynamic settings. That is both to understand the possible limitations of this important benchmark, and to address the fundamental methodological questions, which may prove useful for future research based on more realistic assumptions.

Until recently, dynamic mechanism design problems were surprisingly neglected by the literature. In recent years, a growing literature has developed to fill this important gap.[3] The present paper departs from this literature in two main respects. First, existing papers on dynamic mechanism design typically assume that the stochastic process that generates payoffs and signals is common knowledge among the agents and known to the designer. As such, the

[1] For environments with partial restrictions on beliefs, see Artemov et al. (2013), Lopomo et al. (2013), Kim and Penta (2013) and Ollár and Penta (2014). For a different, 'second best' approach, see Börgers and Smith (2012a,b) and Yamashita (2012, 2013a,b). This literature is discussed more extensively in Section 6.

[2] Müller (2012a,b) also extends the belief-free approach to study dynamic mechanisms, but he considers virtual implementation and only static environments. This work is further discussed in Section 6.

[3] Among others, see Athey and Segal (2014), Bergemann and Valimaki (2010), Pavan, Segal and Toikka (2013) and the references therein.

approach suffers from the non-robustness problem discussed above. Second, the literature thus far has focused solely on problems of *partial implementation*. That is, the design of 'Perfect Bayesian Incentive Compatible' (PBIC) mechanisms, in which agents truthfully reveal their information in *a* Perfect Bayesian Equilibrium (PBE) of the game. PBIC, however, does not rule out the possibility that other, undesirable PBE exist. The more demanding requirement that no such equilibria exist is referred to as *full implementation*. This is an important question, especially if the dynamic interaction provides agents with more opportunities to collude. At the current state of the literature, however, very little is known on the problem. This is the first paper to address the question of dynamic full implementation, let alone the robustness requirement discussed above.[4] Furthermore, the results are obtained under general assumptions on preferences, which need not be quasilinear nor time separable. This is yet another innovation with respect to the literature on dynamic mechanism design.

The first result of this paper shows that, as far as robust partial implementation is concerned, the main insights from the static literature easily extend to dynamic environments: a Social Choice Function (SCF) is PBIC for all possible beliefs if and only if it is ex-post incentive compatible. This result is best seen as a necessary condition for robust full implementation and enables us to focus on the novel issues that dynamic settings raise for the full implementation problem, which is the main focus of the paper. These novelties involve both the dynamic nature of agents' interaction and the methodology of the analysis.

From a methodological viewpoint, even when the agents' beliefs are known to the designer, characterizing the set of PBE of a given mechanism can be very difficult. This may explain why the full implementation question has not been pursued in dynamic settings. It may thus seem that adding the robustness requirement to the already difficult full implementation problem is doomed to make the problem intractable. This paper introduces and provides foundations to a methodology that avoids the difficulties of computing the set of PBE for all possible beliefs. The key ingredient is the notion of interim perfect equilibrium (IPE). IPE weakens 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.

These results are then applied to study conditions for full implementation in environments

[4] The closest work to dynamic full implementation is Lee and Sabourian (2011), who study repeated full implementation. The main difference between 'dynamic' and 'repeated' implementation is that, in the latter, the distribution of types, SCF and mechanism are the same in every period, hence they do not depend on the previous history.

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

The rest of the paper is organized as follows: Section 2 introduces the formalism for the environments and agents' beliefs; Section 3 introduces mechanisms, the notation for the resulting dynamic games, and the key notions of implementation. The analysis of partial and full implementation is contained in Sections 4 and 5, respectively. Section 6 concludes.

2 Setup

Belief-Free Environments. Consider an environment with n agents and T periods, $T < \infty$. In each period $t = 1, ..., T$, each agent $i = 1, ..., n$ observes a signal $\theta_{i,t} \in \Theta_{i,t} = [\theta_{i,t}^l, \theta_{i,t}^h] \subseteq \mathbb{R}$.[5] For each i, $\Theta_i := \times_{t=1}^T \Theta_{i,t}$ is the set of i's *payoff types*: a payoff-type is a complete sequence of agent i's signals in every period. A *state of nature* is a profile of agents' payoff types, and the set of states of nature is defined as $\Theta := \Theta_1 \times ... \times \Theta_n$. As usual, we let $\Theta_{-i,t} := \times_{j \neq i} \Theta_{j,t}$ and $\Theta_{-i} := \times_{j \neq i} \Theta_j$. A similar notation will be used for other product sets.

In each period t, the social planner chooses an allocation from a non-empty subset of a finitely dimensional Euclidean space, Ξ_t. The set $\Xi = \times_{t=1}^T \Xi_t$ denotes the set of feasible sequences of allocations.[6] Agents are expected utility maximizers, with preferences over sequences of allocations that depend on the realization of Θ: for each $i = 1, ..., n$, preferences are represented by utility functions $u_i : \Xi \times \Theta \to \mathbb{R}$. Thus, the states of nature characterize everybody's preferences over the sets of feasible allocations. A *(belief-free) environment* therefore is defined by a tuple $\mathcal{E} = \langle N, \Xi, \Theta, (u_i)_{i \in N} \rangle$, assumed common knowledge.

Environment $\mathcal{E}$ thus represents agents' *information* and *preferences*, not their beliefs. 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 $y_i^t = (\theta_{i,1}, ..., \theta_{i,t}) \in Y_i^t$, agent i knows that the true state of nature $\theta^* \in \Theta$ belongs to the set $\{y_i^t\} \times (\times_{\tau=t+1}^T \Theta_{i,\tau}) \times \Theta_{-i}$. For any $\theta \in \Theta$ and $t = 1, ..., T$, we let $y_i^t(\theta) = (\theta_{i,1}, ..., \theta_{i,t})$ denote the history of i's private signals realized at state θ, up to period t. We define $y^t(\theta)$ and $y_{-i}^t(\theta)$ similarly. Histories of allocations will be denoted by $x^t = (\xi_1, ..., \xi_t) \in \times_{\tau=1}^t \Xi_\tau$.

[5] The finite horizon restriction is important for the results on full implementation, which are based on a backwards procedure. The restriction is also maintained by Müller (2012a,b), but not by the (non-robust) dynamic mechanism design literature, which focuses on partial implementation alone (e.g., Pavan et al., 2013).

[6] In this paper the social choice function (SCF, introduced below) is taken as a primitive. Hence, an explicit representation of intertemporal constraints is unnecessary. Possible intertemporal constraints, as well as the designer's objective function, are accommodated implicitly in the SCF, which can be thought of as the argmax of a constrained optimization problem. The analysis that follows therefore accommodates the possibility of intertemporal constraints in the designer's optimization problem, of course provided that the resulting SCF satisfies the conditions stated in the results.

Social Choice Functions. The description of the primitives is completed by a *social choice function* (SCF), $f : \Theta \to \Xi$. We assume that the SCF is such that period-t choices are measurable with respect to the information available in that period. That is, we assume that there exist functions $f_t : Y^t \to \Xi_t$, $t = 1, ..., T$, such that $f(\theta) = \left(f_t\left(y^t(\theta)\right)\right)_{t=1}^T$ for each $\theta \in \Theta$.

Models of Beliefs. At any point in time, agents have (subjective) beliefs about the features of the environment they do not know. These beliefs are distinct from information, which is encoded in the payoff types and directly affects the SCF. Such beliefs are thus modeled separately from players' information. A *model of beliefs* for an environment $\mathcal{E}$ is a tuple $\mathcal{B} = (B_i, \beta_i)_{i \in N}$ such that for each i, B_i is the set of *types*, assumed Polish, and $\beta_i : B_i \to \Delta(\Theta \times B_{-i})$ is a continuous function.[7]

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

Standard models of mechanism design (e.g., references in footnote 3) assume common knowledge of a specific model of beliefs, and assume a single common prior. This corresponds to the case in which $B_i = \{b_i\}$ is a singleton for each i, and beliefs $\beta_i(b_i) \equiv p^* \in \Delta(\Theta)$ are the same for all i and have full support. The further special case of independent types (e.g., Pavan et al. (2013)) requires that there exist $\hat{p}^i \in \Delta(\Theta_i)$ for each i such that $p^* = \otimes_{i \in N} \hat{p}^i$. A further assumption, common in the literature, is that such $\hat{p}^i$ are Markov processes (e.g., Bergemann and Valimaki (2010)). In constrast, here we will be interested on implementation results *for all* possible models of beliefs $\mathcal{B} = (B_i, \beta_i)_{i \in N}$.

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 $t = 1, ..., T$, when a type's beliefs after a history of private signals y_i^t are concentrated on the set $\left\{y_i^t\right\} \times \left(\times_{\tau=t+1}^T \Theta_{i,\tau}\right) \times \Theta_{-i} \times B_{-i}$. The term *ex-post stage* refers to the final realization, when the state of nature is revealed.

3 Mechanisms, Incentive Compatibility and Implementation

A *mechanism* is defined by a set of messages $M_{i,t}$ for every $i \in N$ and $t = 1, ..., T$, and by a collection of outcome functions $(g_t)_{t=1,...,T}$ which assign allocations to each history at

[7] A Polish space is a complete separable metric space. For any X, $\Delta(X)$ denotes the set of probability measures on X, endowed with the corresponding Borel sigma-algebra.

each stage. As usual, for each t we define $M_t = \times_{i\in N} M_{i,t}$. It is assumed that the reported messages are publicly observed at the end of each period. Formally, let ϕ denote the empty history, and define $H^0 := \{\phi\}$. For each $t = 1, ..., T$, the set of *public histories of length* t is defined as $H^t := H^{t-1} \times M_t$, and the set of *public histories* is denoted by $\mathcal{H} := \cup_{\tau=0}^{T} H^{\tau}$. The period-$t$ outcome function is a mapping $g_t : H^{t-1} \times M_t \to \Xi_t$. A mechanism therefore is a tuple $\mathcal{M} = \langle \left((M_{i,t})_{i\in N}, g_t \right)_{t=1}^{T} \rangle$, assumed commond knowledge. We focus throughout on mechanisms in which the sets $M_{i,t}$ are compact subsets of finitely dimensional Euclidean spaces.

A *direct mechanism* is such that $M_{i,t} = \Theta_{i,t}$, and $g_t = f_t$ for all $i \in N$ and $t = 1, ..., T$. That is, 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 as specified by the SCF, that is according to the function $f_t : Y^t \to \Xi_t$. To emphasize the dependence of the direct mechanism on the SCF f, we denote it by $\mathcal{M}^f$.

Each mechanism induces a dynamic game. If agents' beliefs $\mathcal{B} = (B_i, \beta_i)_{i\in N}$ are specified, the resulting game is a standard Bayesian game, which can be analyzed using standard solution concepts, such as Bayes-Nash (BNE) or Perfect Bayesian Equilibrium (PBE). For the analysis of robust implementation, however, it is useful to consider environments in which agents' beliefs are not specified. For this reason we also introduce the notion of a 'belief-free game', which obtains imposing a mechanism $\mathcal{M}$ on a belief-free environemnt $\mathcal{E}$. In these games, solution concepts such as BNE or PBE are not defined. Their analysis therefore requires novel solution concepts, which will be introduced in Section 5.1.

Belief-Free Games. An environment $\mathcal{E}$ and a mechanism $\mathcal{M}$ determine a *belief-free dynamic game*, that is a tuple $(\mathcal{E}, \mathcal{M}) = \langle N, (\mathcal{H}_i, \Theta_i, u_i)_{i\in N} \rangle$. Sets N, Θ_i and payoff functions u_i are as defined in $\mathcal{E}$. The sets $\mathcal{H}_i$ denote the set of i's *private histories*, defined as follows: for each i and t, let $Y_i^t = \times_{\tau=1}^{t} \Theta_{i,\tau}$, $H_i^t := H^{t-1} \times Y_i^t$ and finally $\mathcal{H}_i := \cup_{\tau=1}^{T} H_i^{\tau}$. That is, for each i and t, H_i^t denotes the set of *private histories of length* t for player i. Each private history $h_i^t = \left(h^{t-1}, y_i^t \right) \in H_i^t$ is made of two components: a *public component*, h^{t-1}, which consists of the agents' messages in periods 1 through $t-1$; and a private component, y_i^t, which consists 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^{τ} is a predecessor of h^t (similarly for private histories: $\left(h^{\tau-1}, y_i^{\tau} \right) \prec \left(h^{t-1}, y_i^t \right)$ if and only if $h^{\tau} \prec h^t$ and $y_i^{\tau} \prec y_i^t$.)

Agents' strategies in the belief-free game are measurable functions $s_i : \mathcal{H}_i \to \bigcup_{t=1}^{T} M_{i,t}$ such that $s_i \left(h_i^t \right) \in M_{i,t}$ for each $h_i^t \in \mathcal{H}_i$. The set of i's pure strategies is denoted by S_i. We also define the sets $S = \times_{i\in N} S_i$ and $S_{-i} = \times_{j\neq i} S_j$. For any strategy profile $s \in S$, each realization of $\theta \in \Theta$ induces a terminal allocation $\mathbf{g}^s(\theta) \in \Xi$. Strategic-form payoff functions, $U_i : S \times \Theta \to \mathbb{R}$, are such that $U_i(s, \theta) = u_i(\mathbf{g}^s(\theta), \theta)$ for each s and θ. For each public history h^t and player i, let $S_i \left(h^t \right)$ denote the set of player i's strategies that are consistent with history h^t being observed. Since i's private histories are only informative of the opponents' behavior through the public history, for each $i \in N$, $h_i^t = \left(h^{t-1}, y_i^t\right) \in \mathcal{H}_i$ and $j \neq i$, $S_j\left(h_i^t\right) = S_j\left(h^{t-1}\right)$.[8]

In a direct mechanism, the *truthtelling strategies* are those strategies that, conditionally on having reported truthfully in the past, report each period-t signal truthfully. 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 'on the path'. The set of such strategies is denoted by $S_i^* \subseteq S_i$, and let $S^* := \times_{i \in N} S_i^*$. For later reference, it is useful to introduce the following notion of incentive compatibility:

Definition 1 (Ex-post Incentive Compatibility) *SCF f is* ex post incentive compatible (EPIC) *if the truthtelling strategy is an ex-post equilibrium of the direct mechanism. That is, if for all $i \in N$, $\theta \in \Theta$, $s^* \in S^*$ and $s_i' \in S_i$, $U_i\left(s^*, \theta\right) \geq U_i\left(s_i', s_{-i}^*, \theta\right)$. SCF f is* strictly EPIC *if the inequality holds strictly for all $s_i' \in S_i \backslash S_i^*$.*

Bayesian Games. A tuple $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ determines a dynamic Bayesian game. Strategies in a Bayesian game are measurable mappings $\sigma_i : B_i \rightarrow S_i$. The set of strategies in $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ is denoted by Σ_i. Agent i's information sets in the Bayesian game are $B_i \times (\mathcal{H}_i \cup \{\phi\})$, with generic element (b_i, h_i). At period 0, agents only know their own type b_i. Period-0 histories therefore are of the form $(b_i, \phi) \in B_i \times \{\phi\}$, and for each $t \geq 1$, period-t histories are $\left(b_i, h_i^t\right) \in B_i \times H_i^t$. In the following, we let $h_i^0 \equiv \phi$, so that information sets are written as $\left(b_i, h_i^t\right) \in B_i \times (\mathcal{H}_i \cup \{\phi\})$ for $t \geq 0$. From the point of view of each i, for each $\left(b_i, h_i^t\right) \in B_i \times (\mathcal{H}_i \cup \{\phi\})$ and strategy profile σ, the induced terminal history is a random variable that depends on the realization of the state of nature and of the opponents' types. This random variable is denoted by $\mathbf{g}^{\sigma|\left(b_i, h_i^t\right)}\left(\theta, b_{-i}\right)$. We define the Bayesian game strategic-form continuation payoff functions as follows:

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

Since $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ is a dynamic Bayesian game, we consider Perfect Bayesian Equilibrium (PBE) as solution concept. This requires introducing notation for belief systems, which represent players' beliefs about the payoff state and the opponents' types at each information set of the Bayesian game. Formally, a *system of beliefs* for $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ is a profile $p = (p_i)_{i \in N}$ where each p_i is a collection of conditional beliefs $p_i\left(b_i, h_i^t\right) \in \Delta\left(\Theta \times B_{-i}\right)$, one for each $\left(b_i, h_i^t\right) \in B_i \times (\mathcal{H}_i \cup \{\phi\})$, such that $p_i\left(b_i, \phi\right) = \beta_i\left(b_i\right)$ for all $b_i \in B_i$.

Definition 2 *Fix an environment $\mathcal{E}$, a model of beliefs $\mathcal{B}$, and a SCF f.*

(Truthful PBE) *The assessment (σ, p) is a* truthful perfect Bayesian equilibrium *of the Bayesian game $\left(\mathcal{E}, \mathcal{M}^f, \mathcal{B}\right)$ if it satisfies the following conditions: (i) (σ, p) is a PBE of $\left(\mathcal{E}, \mathcal{M}^f, \mathcal{B}\right)$; (ii) $\sigma_i\left(b_i\right) \in S_i^*$ for all i and $b_i \in B_i$; (iii) beliefs p_i assign probability one to the other agents having reported truthfully at all histories.*

(PBIC) *The SCF f is* perfect Bayesian incentive compatible (PBIC) on $\mathcal{B}$, *if there exists a truthful PBE of $\left(\mathcal{E}, \mathcal{M}^f, \mathcal{B}\right)$.*

[8]Sets $\mathcal{H}_i$ and S_i are endowed with the standard metrics derived from $H^T \times \Theta$. See Appendix A.1 for details.

Condition (i) is self-explanatory. Condition (ii) requires the equilibrium profile to be truthful, and corresponds to the 'on-path truthful' condition of Pavan et al. (2013). Condition (iii) says that agent i always believes that the opponents have been following their equilibrium strategies. It does not necessarily follow from Bayesian updating because the general setup accommodates models of beliefs over a continuum of signals. In these cases, every history has a zero probability. This condition is standard in the literature on dynamic mechanism design (e.g., Pavan et al. (2013), Bergemann and Valimaki (2010)).

Robust Implementation in direct mechanisms. In the following we focus on partial and full implementation in direct mechanisms, discussed respectively in Sections 4 and 5. The restriction to direct mechanisms is standard in the literature on dynamic mechanism design, which has only focused on partial implementation. The restriction is known to be with loss of generality for the full implementation problem, in that more complicated mechanisms may make full implementation easier to achieve.[9] The simplicity of direct mechanisms, however, is an important desideratum from the viewpoint of the Wilson doctrine, and it has the further advantage of making the comparison between partial and full implementation easier. Finally, since restricting the class of mechanisms makes full implementation results harder to obtain, the restriction to direct mechanisms strengthens the positive results obtained in Sections 5.

Standard mechanism design assumes that agents' beliefs are known to the designer, and therefore implementation is defined for a given model $\mathcal{B}$. To address the issue of robustness, we require that partial and full implementation are achieved *for all* possible beliefs:

Definition 3 (Robust Partial Implementation) *A SCF f is* robustly partially implementable in the direct mechanism *if it is PBIC on all models of beliefs.*

Definition 4 (Robust Full Implementation) *SCF f is* robustly fully implementable in the direct mechanism *if for every $\mathcal{B} = (B_i, \beta_i)_{i \in N}$, every PBE-strategy profile σ of the Bayesian game $(\mathcal{E}, \mathcal{M}^f, \mathcal{B})$ is such that $\sigma(b) \in S^*$ for all $b \in B$.*

Definitions 3 and 4 extend to dynamic settings the notions introduced, respectively, by Bergemann and Morris (2005, BM05) and Bergemann and Morris (2009a, BM09). As discussed in the introduction, the 'belief-free' approach to robustness is clearly very demanding. From a theoretical viewpoint, however, it is an important benchmark, particularly to analyze the methodological aspects of robust mechanism design (cf. Section 6). The belief-free approach therefore is the natural starting point to extend robust mechanism design to dynamic settings.

[9] For this reason, the classical literature on Bayesian Implementation typically adopts complex mechanisms in which agents report more than their own type (e.g., Maskin (1999) Postlewaite and Schmeidler (1988), Palfrey and Srivastava (1989) and Jackson (1991)). Full implementation results via simple design of transfers are provided by Ollár and Penta (2014).

4 Robust Partial Implementation

BM05 define robust (partial) implementation as interim incentive compatibility (IIC) on all type spaces. Definition 3 essentially adapts the underlying notion of incentive compatibility, replacing IIC with the standard in the literature on dynamic mechanism design, PBIC. It can thus be seen as the dynamic counterpart of BM05, as well as the belief-free counterpart of the dynamic mechanism design literature.

BM05 show that a SCF is IIC on all type spaces if and only if it is *ex-post incentive compatible*. Since, for any model of beliefs, PBIC implies IIC, an immediate implication of their result is that ex-post incentive compatibility (Def. 1) is a necessary condition for robust implementation in dynamic settings (Def. 3). But since PBIC is in general more demanding than IIC, achieving PBIC for a given model of beliefs in general requires stronger conditions than IIC (e.g., Pavan et al. (2013)). As the next result shows, however, PBIC has no extra bite once it is required *for all* models of beliefs. EPIC therefore is both necessary and sufficient for robust partial implementation (the proof is in Appendix D.1).

Proposition 1 (Partial Implementation) *SCF f is* PBIC on all models of beliefs *if and only if it is* ex post incentive compatible.[10]

Thus, from the viewpoint of partial implementation, the belief-free approach entails the same incentive compatibility conditions in static as in dynamic settings. This result, however, does not mean that this is an inherently static approach: beyond incentive compatibility, dynamics has an important role even within the belief-free approach. The next section on full implementation provides one instance of this general point (for another instance of the same point, see also Müller (2012a,b)).

5 Robust Full Implementation

This section focuses on robust full implementation in direct mechanisms (Def. 4). It can thus be seen as the dynamic counterpart of BM09. In static settings, BM09 define robust full implementation by requiring that, for all type spaces, all the BNE induce truthful revelation. Since the set of all such equilibria can be computed applying rationalizability to the belief-free (static) game, BM09 study conditions to ensure that truthful revelation is the only rationalizable strategy in the direct mechanism. Besides (strict) ex-post incentive compatibility of the SCF, these conditions require the preference interdependencies to be 'not too strong'. Intuitively, the reason is that in an EPIC mechanism, strong preference interdependence determines strong strategic externalities, which are a source of multiplicity and may thus undermine the possibility of full implementation. Dynamic settings present two distinct orders of problems. First,

[10] Note that Proposition 1 concerns the properties of a SCF and of the associated direct mechanism. It does *not* state that, in general games, the set of ex-post equilibria and the set of PBE for all models of beliefs coincide. See Borgers and McQuade (2007) for a discussion of the relations between solution concepts based on sequential rationality and ex-post equilibria in dynamic games.

independent of the robustness requirement, characterizing the set of equilibria is difficult in a dynamic setting. Second, the dynamic structure enriches the possibilities of preference interdependencies as well as strategic externalities, which can both exhibit intertemporal effects.

The next section addresses the first problem. First, we introduce a weakening of PBE, which we call *interim perfect equilibrium (IPE)*, and show that the set of IPE-strategies for all models of beliefs is characterized by a 'backwards procedure' which combines the logic of rationalizability and backward induction. This result is convenient because it allows a recursive analysis of the full implementation problem. Endowed with these results, we turn to the second order of problems in Section 5.3, where we provide sufficient conditions for robust full implementation. Since IPE is weaker than PBE, achieving full implementation through the backwards procedure suffices for robust full implementation, as defined in Definition 4.[11]

5.1 IPE and the backwards procedure

Fix a Bayesian game, $(\mathcal{E}, \mathcal{M}, \mathcal{B})$. We say that a strategy profile σ is 'sequentially rational' with respect to belief system p, if for every $i \in N$ and every $(b_i, h_i^t) \in B_i \times \mathcal{H}_i$, $\sigma_i \in \arg\max_{\sigma_i' \in \Sigma_i} \int_{\Theta \times B_{-i}} U_i(\sigma, \theta, b_{-i}; b_i, h_i^t) \cdot dp_i(b_i, h_i^t)$. For each agent i and for each information set $\left(b_i, h_i^{t-1}\right)$, a strategy profile σ and conditional beliefs $p_i(b_i, h_i^{t-1})$ induce a probability measure $P^{\sigma,p_i}(b_i, h_i^{t-1}) \in \Delta(H^{t-1} \times Y_i^t)$ over the private histories of length t.

Definition 5 (IPE) *An assessment* (σ, p) *is an* Interim Perfect Equilibrium (IPE) *if* σ *is sequentially rational with respect to* p *and if* p *satisfies the following conditions: (B-1) for each* $h_i^t = (y_i^t, h^{t-1}) \in \mathcal{H}_i$ *and* $b_i \in B_i$, $p_i(b_i, h_i^t) \in \Delta(\{y_i^t\} \times (\times_{\tau=t+1}^{T} \Theta_{i,\tau}) \times \Theta_{-i} \times B_{-i})$, *and (B-2) for each* h_i^t *such that* $h_i^{t-1} \prec h_i^t$, *for every measurable* $E \subseteq \Theta \times B_{-i}$, $p_i(b_i, h_i^{t-1})[E] = p_i(b_i, h_i^t)[E] \cdot P^{\sigma,p_i}(b_i, h_i^{t-1})\left[h_i^t\right]$.

Condition (B-1) requires conditional beliefs at each information set to be consistent with the player's private information. Condition (B-2) requires that belief system p_i is consistent with Bayesian updating whenever possible, both 'on-' and 'off-the-path'. The latter condition in particular is not required by 'weak PBE' (cf. Mas-Colell et al. (1995, p.285)). Conditions (B-1) and (B-2) impose essentially no restrictions on the beliefs held at histories that receive zero probability at the preceding node.[12] IPE therefore is weaker than PBE. Also note 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 $\sigma_i(b_i, h_i^{t-1})$ alone, then $P^{\sigma,p_i}(b_i, h_i^{t-1})[h_i^t] = 0$, and agent i's beliefs at (b_i, h_i^t) are unrestricted the same way they would be after an unexpected move of the opponents. As it will be discussed shortly, this feature of IPE is key to obtaining a recursive and tractable characterization of the set of equilibria.

Belief-Free Backwards Rationalizability. We introduce next a solution concept for belief-free dynamic games, which will be shown to characterize the set of all IPE-strategies over all

[11] Note that, contrary to partial implementation, full implementation results in general are stronger if obtained for weaker solution concepts.

[12] Hence, conditions B(i), B(iii) and B(iv) in Fudenberg and Tirole (1991, p.332) need not hold in an IPE.

models of beliefs. The formal definition is notationally cumbersome, and left to Appendix B, but the idea is straightforward. Fix a belief-free game, $(\mathcal{E}, \mathcal{M})$, and a public history of length $T-1$, h^{T-1}. For each payoff-type $y_i^T \in \Theta_i$ of each agent, the continuation game is a (belief-free) static game with strategies $s_i|h^{T-1} \in S_i^{h^{T-1}}$. We apply belief-free rationalizability (e.g., BM09) to this game, and let $R_i(h^{T-1})$ 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 do this for all public histories of length $T-1$. We then proceed backwards: for each public history of length $T-2$, h^{T-2}, we apply rationalizability to the continuation game from h^{T-2}, restricting continuation strategies $s_i|h^{T-2} \in S_i^{h^{T-2}}$ to be rationalizable in the continuation games from histories of length h^{T-1}. We let $R_i(h^{T-2})$ denote 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 ϕ is reached, for which the set of 'Belief-Free Backwards Rationalizable' (BFBR) strategies, R_i^{ϕ}, is computed.

Proposition 2 (Characterization of the set of IPE.) *Fix a belief-free game* $(\mathcal{E}, \mathcal{M})$*. For each i:* $\hat{s}_i \in R_i^{\phi}$ *if and only if* $\exists \mathcal{B} = (B_i, \beta_i)_{i \in N}$ *s.t.* $\exists \hat{b}_i \in B_i$ *and* $(\hat{\sigma}, \hat{p})$ *such that* $(\hat{\sigma}, \hat{p})$ *is an IPE of* $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ *and* $\hat{s}_i = \hat{\sigma}_i(\hat{b}_i)$*.*

The proof of this result can be found in Appendix C. To grasp the basic intuition, notice that an implication of this proposition is that, for each public history h, the set of IPE strategies in the continuation game from h coincides with the set of IPE strategies of the continuation game considered in isolation. Hence, the set of IPE strategies for all models of beliefs has a recursive structure analogous to that of the set of subgame perfect equilibria in complete information games. Such set of equilibria can thus be computed 'backwards', analyzing each continuation game in isolation. The property of IPE highlighted above, that own-deviations are treated the same as the opponents', is key to the possibility of considering continuation games in isolation, which is needed for this result.

Since IPE is weaker than PBE, Propositions 1 and 2 imply that an EPIC SCF is fully robustly implemented by a direct mechanism if all the BFBR-strategies are truthful:

Corollary 1 *Let f be EPIC, and consider the belief-free game induced by direct mechanism associated to SCF f,* $(\mathcal{E}, \mathcal{M}^f)$*. If for all i,* $R_i^{\phi} \neq \emptyset$ *and* $R_i^{\phi} \subseteq S_i^*$*, then f is fully robustly implementable (Def. 4).*

By transforming the original problem into an artificial sequence of static problems, the backwards procedure enables us to build on the insights of the static literature to obtain sufficient conditions that guarantee that all the strategies consistent with this solution concept are truthful. Robust full implementation then follows from Corollary 1. The next section contains an illustrative example. Section 5.3 presents the general results.

5.2 Example: A Dynamic Public Good Problem

Consider an environment with two agents ($n = 2$) and two periods ($T = 2$), and let $\Theta_{i,t} = [0,1]$ for each i and t. In each period, the planner chooses some quantity $q_t \in \Xi_t \equiv \mathbb{R}_+$ of public good. Agent i's marginal utility for the public good in period t is a function $\alpha_{i,t}(\cdot)$ of the realized state: for any $\theta = (\theta_{i,1}, \theta_{i,2}, \theta_{j,1}, \theta_{j,2}) \in \Theta$,

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

where $\gamma \geq 0$ and $\varphi : [0,1]^2 \to \mathbb{R}$ is continuously differentiable and strictly increasing in both arguments (hence, first period signals affect second period valuations). Furthermore, if $\gamma = 0$, this is a private-values setting; for any $\gamma > 0$, agents have interdependent values. Finally, we assume time-separability and transferable utility. Agent i's utility function therefore is:

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

where $\pi_{i,t}$ denotes the transfer at period $t = 1, 2$. Notation $\alpha_{i,t}$ is mnemonic for 'aggregator': functions $(\alpha_{i,t})_{t=1,2}$ aggregate the information available to all the agents up to period t into real numbers which uniquely pin down i's preferences. Assuming a cost of production equal to $c(q_t) = \frac{1}{2}q_t^2$, the optimal levels of public good in the two periods are such that, for any $\theta \in \Theta$:

$$q_1^*(\theta) = \alpha_{i,1}(\theta_1) + \alpha_{j,1}(\theta_1) \text{ and} \tag{2}$$
$$q_2^*(\theta) = \alpha_{i,2}(\theta_1, \theta_2) + \alpha_{j,2}(\theta_1, \theta_2). \tag{3}$$

We consider here robust implementation of the efficient rule, $(q_t^*)_{t=1,2}$. By Proposition 1, ex-post incentive compatibility is a necessary condition for robust implementation. We therefore let $(\pi_{i,t}^*, \pi_{j,t}^*)_{t=1,2}$ denote the ex-post incentive compatible transfers, and consider the SCF $f = (q_t^*, \pi_{i,t}^*, \pi_{j,t}^*)_{t=1,2}$.[13] To see that f is indeed EPIC, for any (θ, m), let $\Delta_i(\theta, m) := \alpha_{i,2}(m) - \alpha_{i,2}(\theta)$. Given a profile of first period reports $\hat{m}_1 = (\hat{m}_{i,1}, \hat{m}_{j,1})$ and private signals $(\hat{\theta}_{i,1}, \hat{\theta}_{i,2})$, i's best response $m_{i,2}^*$ to point beliefs $(\theta_{j,1}, \theta_{j,2}, m_{j,2})$ at the second period satisfies:[14]

$$\Delta_i\left(\hat{\theta}_{i,1}, \hat{\theta}_{i,2}, \theta_{j,1}, \theta_{j,2}, \hat{m}_1, m_{i,2}^*, m_{j,2}\right) = 0. \tag{4}$$

[13] The EPIC transfers π^* are defined as follows: for any $\theta \in \Theta$,

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

[14] For the sake of illustration, we ignore here the possibility of corner solutions, which do not affect the fundamental insight. Corner solutions will be discussed in Section 5.3.

Similarly, given private signal $\hat{\theta}_{i,1}$ and point beliefs about the future own signal and message and about the opponents' signals and reports in both periods, $(\theta_{i,2}, m_{i,2}, \theta_{-i}, m_{-i})$, the first period best-response $m^*_{i,1}$ for agent i satisfies:

$$\begin{aligned} m^*_{i,1} - \hat{\theta}_{i,1} = {} & \gamma\left(\theta_{j,1} - m_{j,1}\right) \\ & + \frac{\partial \varphi\left(m^*_{i,1}, m_{i,2}\right)}{\partial m_{i,1}} \cdot \Delta_i\left(\hat{\theta}_{i,1}, \theta_{i,2}, m^*_{i,1}, m_{i,2}, \theta_{-i}, m_{-1}\right) \end{aligned} \tag{5}$$

It is now easy to verify that $f = (q^*_t, \pi^*_{i,t}, \pi^*_{j,t})_{t=1,2}$ is EPIC: for any θ, if agent i has reported truthfully in the past ($m_{i,1} = \theta_{i,1}$) and he expects the opponents to report truthfully ($\theta_{-i} = m_{-i}$), then (4) is satisfied if and only if i reports truthfully in the second period ($m_{i,2} = \theta_{i,2}$). Furthermore, if $\Delta_i(\theta, m) = 0$, the right-hand side of (5) is zero if the opponents report truthfully in the first period, and so it is optimal to report $m^*_{i,1} = \hat{\theta}_{i,1}$. Notice that this is the case for any $\gamma \geq 0$, which means that f is robustly partially implementable for all $\gamma \geq 0$.

Ex-post incentive compatibility, however, does not suffice for full implementation, as non-truthful equilibria may also exist. To address this problem, we apply the backwards procedure R^{ϕ} introduced above. First, notice that (4) implies that, conditional on having reported truthfully in the first period ($m_{i,1} = \theta_{i,1}$), truthful revelation in the second period is a best-response to truthful revelation of the opponent, irrespective of the realization of θ. If i misreported in the first period ($m_{i,1} \neq \theta_{i,1}$), then (maintaining $m_{j,t} = \theta_{j,t}$ for $t = 1, 2$) 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^c_i: a strategy that reports truthfully at the beginning of the game and at every truthful history, but in which earlier misreports are followed by further misreports, to correct the impact of the previous misreports on the value of the aggregator $\alpha_{i,2}$.[15]

We show next that, if $\gamma < 1$, the self-correcting strategy profile is the only profile surviving the backwards procedure introduced above. Hence, given the result in Proposition 2, the self-correcting strategy is the only IPE-strategy, hence the only PBE-strategy, for any model of beliefs. Since s^c induces truthful revelation, this implies that f is fully robustly implementable if $\gamma < 1$. To this end, fix the profile of first period reports, $\hat{m}^1$. Given $\theta_i = (\theta_{i,1}, \theta_{i,2})$ and $m_{i,2} \in M_{i,2}$, let $w_i(\hat{m}_{i,1}, m_{i,2}, \theta_i) = [\varphi(\hat{m}_{i,1}, m_{i,2}) - \varphi(\theta_{i,1}, \theta_{i,2})]$ denote type θ_i's implied over-report of the value of φ. Then, equation (4) can be interpreted as saying that the optimal over-report of $\varphi(\theta_i)$ is equal to $-\gamma$ times the (expected) under-report of $\varphi(\theta_j)$. Let $\underline{\mathrm{w}}^0_j$ and $\bar{\mathrm{w}}^0_j$ denote, respecively, the minimum and maximum possible values of $w_i(\hat{m}_{i,1}, m_{i,2}, \theta_i)$ over $(m_{i,2}, \theta_i)$. Then, if i is rational, the optimal over-report for type θ_i at history $\hat{m}_1$, $w^*_i(\theta_i, \hat{m}_1)$, is bounded above and below, respectively, by $\bar{\mathrm{w}}^1_i \equiv \gamma \cdot \underline{\mathrm{w}}^0_j$ and $\underline{\mathrm{w}}^1_i \equiv -\gamma \cdot \bar{\mathrm{w}}^0_j$. Recursively, define $\bar{\mathrm{w}}^k_i = -\gamma \cdot \underline{\mathrm{w}}^{k-1}_j$ and $\underline{\mathrm{w}}^k_i = -\gamma \bar{\mathrm{w}}^{k-1}_j$. Also, for each k and i, let $z^k_i \equiv [\bar{\mathrm{w}}^k_i - \underline{\mathrm{w}}^k_i]$ denote the distance between the maximum and lowest possible over-report at step k. Substituting, we obtain a

[15] In this example, the self-correcting strategy can be related to the 'strongly truthful' strategy of Pavan et al. (2013), redefining i's second period signals as $\varphi(\theta_i)$. This transformation, however, is not always possible in the general environments of Section 5.3. Footnote 17 discusses this point in some detail.

system of difference equations, $\mathbf{z}^k = \mathbf{\Gamma} \cdot \mathbf{z}^{k-1}$, where:

$$\mathbf{z}^k = \begin{pmatrix} z_i^k \\ z_j^k \end{pmatrix} \text{ and } \Gamma = \begin{bmatrix} 0 & \gamma \\ \gamma & 0 \end{bmatrix}. \tag{6}$$

Notice that the continuation game from $\hat{m}_1$ is dominance solvable if and only if $\mathbf{z}^k \to 0$ as $k \to \infty$. In that case, for each θ_i, $w_i^* (\theta_i, \hat{m}_1) \to 0$, and so the continuation of the self-correcting strategy is uniquely rationalizable in the continuation game. In this example, this is the case 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 (5) 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 $\gamma < 1$. Then, if $\gamma < 1$, the *self-correcting strategy* is the only strategy surviving the backwards procedure R^{ϕ}, hence the only strategy played as part of PBE, for any model of beliefs. It follows that, if preference interdependencies are not too strong, f is robustly fully implementable.

Key properties and their generalizations. The next section generalizes the key insights of this example to environments with monotone aggregators of information (EMA, Def. 6). As in the example above, these environments have the property that for each agent and in each period, all the available information (across time and agents) can be summarized by T one-dimensional statistics (one for each period), which jointly pin down the agent's preferences. Preferences, however, need not be additively separable over periods nor quasilinear. To accommodate the more general class of preferences, the notion of self-correcting strategy will be generalized. A 'contraction property' will be introduced to formalize the idea of bounding preference interdependencies, which generalizes the condition $\gamma < 1$ in the example. The main results show that, if the SCF is strictly EPIC and preferences satisfy such a contraction property and properly defined single-crossing conditions, then s^c is the only strategy that survives the backwards procedure. The SCF is therefore robustly fully implementable.

As shown by the example, despite EPIC is satisfied here, the analysis still presents non trivial dynamics due to the intertemporal interaction and the full implementation requirement: in the presence of strategic uncertainty, ruled out in the partial implementation approach, depending on the agent's beliefs about future signals and others' strategies, misreporting in one period may be a best response even if the SCF is EPIC. It is only thanks to the methodology based on BFBR that we can ensure that, if $\gamma < 1$, $\Delta(\theta, m) = 0$ independent of the agent's current choice, and hence we can analyze the problem *as if* it is a static one. Further 'intertemporal effects', which do not arise in the example above, are possible in the general framework (e.g., an example with 'path dependent' preferences is discussed in Sections 5.3.1 and 5.3.2).

5.3 Robust Full Implementation

Definition 6 (EMA) *An Environment admits* monotone aggregators *(EMA) if, for each i, and for each $t = 1, ..., T$, there exists an* aggregator *function $\alpha_{i,t} : Y^t \rightarrow \mathbb{R}$ and a valuation function $v_i : \Xi \times \mathbb{R}^T \rightarrow \mathbb{R}$ that satisfy the following conditions:*

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

2. *$\alpha_{i,t}$ and v_i are continuous and $\alpha_{i,t}$ is strictly increasing in $\theta_{i,t}$.*

3. *For any $y_i^t, \hat{y}_i^t \in Y_i^t$, if $\alpha_{i,t}\left(y_i^t, y_{-i}^t\right) > \alpha_{i,t}\left(\hat{y}_i^t, y_{-i}^t\right)$ for some $y_{-i}^t \in Y_{-i}^t$, then $\alpha_{i,t}\left(y_i^t, \hat{y}_{-i}^t\right) > \alpha_{i,t}\left(\hat{y}_i^t, \hat{y}_{-i}^t\right)$ for all $\hat{y}_{-i}^t \in Y_{-i}^t$.*

Assuming the existence of the aggregators and the valuation functions (condition 1), per se, entails little loss of generality. The bite of the representation derives mainly from the continuity and monotonicity conditions (2), and from condition (3), which guarantees that i's private histories of signals Y_i^t can be ordered in terms of the induced values of the period-t aggregator $\alpha_{i,t}$.[16]

The next definition generalizes the idea of self-correcting strategy introduced in the example.

Definition 7 (Self-correcting strategy) *The* self-correcting strategy, *$s_i^c \in S_i$, is such that for each $t = 1, ..., T$ and public history $h^{t-1} = \tilde{y}^{t-1}$, and for each $h_i^t = \left(h^{t-1}, y_i^t,\right)$,*

$$s_i^c\left(h_i^t\right) = \arg\min_{m_{i,t}\in\Theta_{i,t}} \left\{ \max_{y_{-i}^t \in Y_{-i}^t} \left|\alpha_{i,t}\left(y_i^t, y_{-i}^t\right) - \alpha_{i,t}\left(\tilde{y}_i^{t-1}, m_{i,t}, y_{-i}^t\right)\right| \right\}. \tag{7}$$

In words, conditional on past truthful revelation, strategy 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 misresports. Strategy s^c therefore induces truthful reporting on its path, hence $s^c \in S^*$.

Given private history $h_i^t = \left(h^{t-1}, y_i^t\right)$ (and the induced report $s_i^c\left(h_i^t\right)$), let $\hat{y}_{-i}^t\left(h_i^t\right)$ be s.t.:

$$\hat{y}_{-i}^t\left(h_i^t\right) \in \arg\max_{y_{-i}^t \in Y_{-i}^t} \left|\alpha_{i,t}\left(y_i^t, y_{-i}^t\right) - \alpha_{i,t}\left(\tilde{y}_i^{t-1}, s_i^c\left(h_i^t\right), y_{-i}^t\right)\right|. \tag{8}$$

Then, the definition of s_i^c and the properties of $\alpha_{i,t}$ (Def. 6) imply that, for any $h_i^t = \left(h^{t-1}, y_i^t,\right)$,

[16] To understand the restriction entailed by condition (3), suppose that aggregator function $\alpha_{i,2}$ in the example of Section 5.2 is replaced by the following: $\alpha_{i,2}(\theta) = (\theta_{i,2} + \theta_{j,2}) + (\theta_{i,1} + \theta_{j,1}) \cdot \mathbf{1}\{\theta_{j,1} \geq \theta_{i,1}\}$ (where $\mathbf{1}\{\cdot\}$ denotes the indicator function). To see how this violates condition (3), consider $y_i^2 = (3/4, 3/4)$ and $\hat{y}_i^2 = (1/2, 1/2)$. Then, $\alpha_{i,2}(y_i^2, y_j^2) > \alpha_{i,2}(\hat{y}_i^2, y_j^2)$ if $y_j^2 = (1/2, 1/2)$, but $\alpha_{i,2}(y_i^2, y_j^2) < \alpha_{i,2}(\hat{y}_i^2, y_j^2)$ if $y_j^2 = (3/4, 3/4)$. Hence, y_i^2 does not imply an unambiguously higher aggregated value than $\hat{y}_i^2$ does: whether one history of signals induces a higher aggregated value than the other (hence a higher marginal utility for q_2 in the example) depends on the signals of the other player. This is ruled out by condition (3), which requires one history of signals to be 'unambiguously higher' than the other, in terms of the induced aggregated value.

three cases are possible:

$$\alpha_{i,t}\left(y_i^t, y_{-i}^t\right) = \alpha_{i,t}\left(\tilde{y}_i^{t-1}, s_i^c\left(h_i^t\right), y_{-i}^t\right) \text{ for all } y_{-i}^t \in Y_{-i}^t, \tag{9}$$

$$\alpha_{i,t}\left(y_i^t, \tilde{y}_{-i}^t\right) > \alpha_{i,t}\left(\tilde{y}_i^{t-1}, s_i^c\left(h_i^t\right), \hat{y}_{-i}^t\left(h_i^t\right)\right) \text{ and } s_i^c\left(h_i^t\right) = \theta_{i,t}^h, \tag{10}$$

$$\alpha_{i,t}\left(y_i^t, \tilde{y}_{-i}^t\right) < \alpha_{i,t}\left(\tilde{y}_i^{t-1}, s_i^c\left(h_i^t\right), \hat{y}_{-i}^t\left(h_i^t\right)\right) \text{ and } s_i^c\left(h_i^t\right) = \theta_{i,t}^l. \tag{11}$$

Equation (9) 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 5.2, suppose that the first period under- (respectively, over-) report is so low (high), that even reporting the highest (lowest) possible message in the second period is not enough to correct the implied value of φ. These histories correspond, respectively, to cases (10) and (11), associated with the highest and lowet period-t reports $\theta_{i,t}^h$ and $\theta_{i,t}^l$, and in the example they induce corner solutions at the second period.[17]

We introduce next a 'contraction property' which bounds the strength of preference interdependence, and provides a multi-period extension of the analogous condition for static environments in BM09. To introduce the condition formally, some extra notation is needed: for each set of strategy profiles $D \subseteq S$, and for each private history $h_i^t = (h^{t-1}, y_i^t)$, let $D_i(h_i^t) := \{m_{i,t} : \exists (s_i, s_{-i}) \in D \text{ s.t. } s_i(h_i^t) = m_{i,t}\}$ and $D_i\left(h^{t-1}\right) := \bigcup_{y_i^t \in Y_i^t} D_i\left(h^{t-1}, y_i^t\right)$. Let $\mathbf{s}_i\left[D_i\left(h^{t-1}\right)\right]$ denote the set of pairs $\left(m_{i,t}, y_i^t\right) \in M_{i,t} \times Y_i^t$ such that $m_{i,t} \in D_i(h^{t-1}, y_i^t)$, and $\mathbf{s}_i^c\left[h^{t-1}\right]$ the set of $\left(m_{i,t}, y_i^t\right)$ such that $m_{i,t} = s_i^c\left(h^{t-1}, y_i^t\right)$.

Definition 8 (Contraction Property) *An environment with monotone aggregators of information satisfies the* Contraction Property *if, for each non-empty $D \subseteq S$ such that $D \neq \{s^c\}$ and for each public history $h^{t-1} = \tilde{y}^{t-1}$ such that $\mathbf{s}\left[D\left(h^{t-1}\right)\right] \neq \mathbf{s}^c\left[h^{t-1}\right]$, there exists y_i^t and $m_{i,t}' \in D_i\left(h^{t-1}, y_i^t\right)$, $m_{i,t}' \neq s_i^c\left(h^{t-1}, y_i^t\right)$, such that:*

$$\textit{sign}\left[s_i^c\left(h^{t-1}, y_i^t\right) - m_{i,t}'\right] = \textit{sign}\left[\alpha_{i,t}\left(y_i^t, y_{-i}^t\right) - \alpha_{i,t}\left(\tilde{y}^{t-1}, m_{i,t}', m_{-i,t}'\right)\right], \tag{12}$$

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

To interpret the condition, and make it more easily comparable to BM09's, it is useful to

[17] Suppose that the aggregator functions satisfy the following 'reduction property': $\forall i, \forall t$, $\exists \varphi_{i,t} : Y_i^t \to \mathbb{R}$ and $\hat{\alpha}_{i,t} : \mathbb{R} \times Y_{-i}^t$ s.t. $\alpha_{i,t}\left(y_i^t, y_{-i}^t\right) = \hat{\alpha}_{i,t}\left(\varphi_{i,t}\left(y_i^t\right), y_{-i}^t\right)$ for all $\left(y_i^t, y_{-i}^t\right)$ (the example in Section 5.2 has this property). Then, agents' signals can be relabeled as follows: for each i and t, let the 'transformed signal' be $\theta_{i,t}' = \varphi_{i,t}\left(y_i^t(\theta)\right)$. In the relabeled model, $\alpha_{i,t}$ only depends on $\theta_{i,t}'$ (not on $\theta_{i,\tau}'$ for $\tau < t$). Strategy s_i^c therefore induces truthful revelation of the transformed signal at every period. The self-correcting strategy is thus related to the 'strongly truthful' strategies considered, for instance, by Pavan et al. (2013) to study PBIC in Markovian environments. This connection, however, is not perfect: first, the environments of Def. 6 need not satisfy the 'reduction property', and the relabeling that transforms the self-correcting into a strongly truthful strategy may not be possible; second, s_i^c is defined in belief-free environments, which do not include a stochastic process, hence no assumptions are made on the evolution of signals over time. Thus, even if the reduction property is satisfied, so that the period-t aggregator $\alpha_{i,t}$ is completely pinned down by the (relabeled) period-t signal $\theta_{i,t}' = g_{i,t}\left(y_i^t\right)$, the environment need not be Markovian: i's beliefs about future signals, for instance, may depend on earlier signals, even if $\alpha_{i,t}$ does not.

rewrite the argument of the right-hand side of (12) as follows:

$$\alpha_{i,t}\left(y_i^t, y_{-i}^t\right) - \alpha_{i,t}\left(\tilde{y}^{t-1}, m'_{i,t}, m'_{-i,t}\right)$$
$$= \left[\alpha_{i,t}\left(\tilde{y}^{t-1}, s_i^c\left(h^{t-1}, y_i^t\right), y_{-i}^t\right) - \alpha_{i,t}\left(\tilde{y}^{t-1}, m'_{i,t}, m'_{-i,t}\right)\right] \tag{13}$$
$$+ \kappa_i\left(h^{t-1}, y_i^t, y_{-i}^t\right),$$
$$\text{where } \kappa_i\left(h^{t-1}, y_i^t, y_{-i}^t\right) = \alpha_{i,t}\left(y_i^t, y_{-i}^t\right) - \alpha_{i,t}\left(\tilde{y}^{t-1}, s_i^c\left(h^{t-1}, y_i^t\right), y_{-i}^t\right). \tag{14}$$

First note that the term κ_i represents the extent by which the self-correcting strategy is incapable of offsetting the previous misreports. To understand how the contraction property formalizes the idea that preference interdependences are not too strong (restriction $\gamma < 1$ in the example), consider first a public history $\tilde{y}^{t-1}$ along which i has reported truthfully ($\tilde{y}_i^{t-1} = y_i^{t-1}$). At such a history, the self-correcting strategy requires that i reports $\theta_{i,t}$ truthfully, so that $\kappa_i(h^{t-1}, y_i^t, y_{-i}^t) = 0$. Then, the condition boils down to requiring that, given $\tilde{y}^{t-1}$ and i's period-t signal $\theta_{i,t}$, for any $m'_{i,t} \neq \theta_{i,t}$ and for all y_{-i}^t and $m'_{-i,t}$, $sign\left[\theta_{i,t} - m_{i,t}\right] = sign\left[\alpha_{i,t}\left((y_i^{t-1}, \theta_{i,t}), y_{-i}^t\right) - \alpha_{i,t}\left(\tilde{y}^{t-1}, m'_{i,t}, m'_{-i,t}\right)\right]$. That is, the direct impact of i's private signal $\theta_{i,t}$ on the aggregator $\alpha_{i,t}$ is always sufficiently strong that the difference in the aggregated value between the true signals and the reported signals always has the same sign as the difference between the true and reported signal of agent i by itself, regardless of others' reports in this period ($m'_{-i,t}$), or whether their earlier reports were truthful or not ($\tilde{y}_{-i}^{t-1}$ is given, but the condition is requested for all y_{-i}^t). The same logic applies to other histories with the property that $\kappa_i = 0$ (which, by equations (9)-(11), occur whenever $s_i^c(h_i^t) \in (\theta_{i,t}^l, \theta_{i,t}^h)$), with the only difference that the 'truthful report' $\theta_{i,t}$ is now replaced by the 'self-correcting report', $s_i^c\left(h_i^t\right)$. To account for the possibility that, at some histories, the self-correcting strategy is not sufficient to offset the previous misreports (the corner solutions in the example of Section 5.2), the contraction property further requires that the sign of the impact on the aggregator $\alpha_{i,t}$ is not offset by the previous misreports, measured by $\kappa_i > 0$.[18]

Thus, similar to BM09's analogous condition, the contraction property limits the strength of the preference interdependence. The key difference here is that payoff types are revealed over time, and the strength of the preference interdependence may vary from period to period. The condition above ensures that such preference interdependence remains 'small' at any point in time, for all possible reports that may have already been revealed.

The last assumption to obtain the full implementation result is a single-crossing condition. As usual, single-crossing conditions allow to sort types with respect to the implemented allocation. The key difference in dynamic environments is that this sorting must also take into account the intertemporal incentives. The single-crossing condition therefore will involve restrictions that are both 'within' and 'between' periods. To cleanly separate the two, we first consider 'aggregator based' SCFs (Section 5.3.1), and show that a simple 'within period'

[18] Appendix D.4 illustrates how this complexity may be avoided by adopting simple mechanisms with extended message spaces, so that any possible past misreport can be corrected, inducing $\kappa_i\left(h^{t-1}, y_i^t, y_{-i}^t\right) = 0$ at all histories. For a similar trick, see Pavan (2008).

single-crossing condition suffices to obtain full implementation in this case. We then discuss the restrictiveness of the aggregator-based assumption, and the extra complications due to relaxing it. The general results are provided in Section 5.3.2.

5.3.1 Aggregator-Based SCF

Consider the SCF in the example of Section 5.2: the SCF has the property that 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 9 (AB-SCF) *The SCF $f = (f_t)_{t=1}^T$ is* aggregator-based *if for each t, $\alpha_{i,t}\left(y^t\right) = \alpha_{i,t}\left(\tilde{y}^t\right)$ for all i implies $f_t\left(y^t\right) = f_t\left(\tilde{y}^t\right)$.*

We next introduce a standard single-crossing condition (SCC), applied to every period.

Definition 10 (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 $\xi, \xi' \in \Xi$ s.t. $\xi_\tau = \xi'_\tau$ for all $\tau \neq t$, for each $a^*_{i,-t} \in \mathbb{R}^{T-1}$ and for each $\alpha_{i,t} < \alpha'_{i,t} < \alpha''_{i,t}$, $v_i(\xi, \alpha_{i,t}, a^*_{i,-t}) > v_i(\xi', \alpha_{i,t}, a^*_{i,-t})$ and $v_i(\xi, \alpha'_{i,t}, a^*_{i,-t}) = v_i(\xi', \alpha'_{i,t}, a^*_{i,-t})$ implies $v_i(\xi, \alpha''_{i,t}, a^*_{i,-t}) < v^t_i(\xi', \alpha''_{i,t}, a^*_{i,-t})$.*

Equivalently: for any two allocations ξ and ξ' that only differ in their period-t component, for any $a^*_{i,-t} \in \mathbb{R}^{T-1}$, the difference $\delta_{i,t}\left(\xi, \xi', \alpha_{i,t}\right) = v_i\left(\xi, \alpha_{i,t}, a^*_{i,-t}\right) - v_i\left(\xi', \alpha_{i,t}, a^*_{i,-t}\right)$, as a function of $\alpha_{i,t}$, crosses zero at most once (Figure 1.a, p. 20). If $T = 1$, SCC-1 coincides with BM09's condition.

Proposition 3 (Full Implementation: AB-SCF) *In an environment with monotone aggregators (Def. 6) satisfying SCC-1 (Def. 10) and the* contraction property *(Def. 8), if an* aggregator-based *social choice function satisfies* Strict EPIC *(Definition 1), then $R^\phi = \{s^c\}$.*

The argument of the proof, which can be found in Appendix D.2, is analogous to that of the example in Section 5.2. 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 $t+1, ..., T$ agents follow s^c, a misreport at period t only affects the period-t aggregator. Since the SCF is aggregator-based, the allocations at periods $\tau \neq t$ are fixed, hence the problem at period-t is essentially static. The contraction property and the 'within period' SCC-1 therefore imply that the self-correcting strategy is followed at stage t. And so on. Clearly, the entire argument relies on the methodological results of Section 5.1, which enable us to study continuation games in isolation, and to apply the backwards procedure.

An appraisal of the aggregator-based assumption. Consider the important special case of *time-separable preferences*: for each i and $t = 1, ..., T$, there exist an aggregator function

$\alpha_{i,t} : Y^t \to \mathbb{R}$ and a valuation function $v_i^t : \Xi_t \times \mathbb{R} \to \mathbb{R}$ such that for each $(\xi^*, \theta^*) \in \Xi \times \Theta$,

$$u_i (\xi^*, \theta^*) = \sum_{t=1}^{T} v_i^t \left(\xi_t^*, \alpha_{i,t} \left(y^t (\theta^*) \right) \right).$$ [19]

In this case, the condition that the SCF is aggregator-based (Def. 9) can be interpreted as saying that the SCF only responds to changes in preferences. These preferences, however, cannot accommodate phenomena of path-dependence such as habit formation or learning-by-doing. If preferences are path-dependent, the aggregator-based assumption is too restrictive. For instance, suppose that agents in the example of Section 5.2 have the following preferences:

$$u_i (q_1, q_2, \pi_{i,1}, \pi_{i,2}, \theta) = \alpha_{i,1} (\theta_1) \cdot q_1 + \pi_{i,1} + \left[\alpha_{i,2} (\theta_1, \theta_2) \cdot F(q_1) \cdot q_2 + \pi_{j,2} \right]. \tag{15}$$

Then, the marginal utility of q_2 also depends on the amount of public good provided in the first period. Then, the efficient rule for the second period is $q_2^* (\theta) = [\alpha_{i,2} (\theta) + \alpha_{j,2} (\theta)] \cdot F(q_1)$, which is not aggregator-based. To allow for path-dependent preferences, therefore, it is important to relax the aggregator-based assumption.

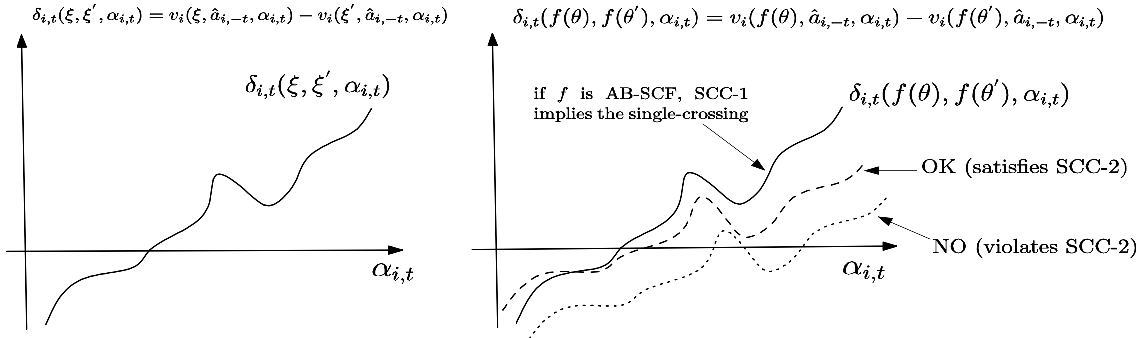
5.3.2 Relaxing the AB-assumption

The problem with relaxing the aggregator-based assumption is that a one-shot deviation from s^c at period-t may induce different allocations in period-t and also in subsequent periods. Hence, the 'within period' single-crossing condition (SCC-1) may not suffice to conclude the inductive step in the proof of Proposition 3, and guarantee that strategy s^c is played at period-t for all $t \leq T$. To avoid this problem, some bound is needed on the impact that a one-shot deviation has on the outcome of the SCF. The next single-crossing condition guarantees that such intertemporal effects of one-shot deviations are not too strong.

Definition 11 (SCC-2) *An environment with monotone aggregators of information satisfies SCC-2 if, for each i: for each $\theta, \theta' \in \Theta$ such that $\exists t \in \{1, ..., T\}$ s.t. $y^\tau (\theta) = y^\tau (\theta')$ for all $\tau < t$ and $\alpha_j^\tau (\theta) = \alpha_j^\tau (\theta')$ for all $\tau > t$ and for all j, for each $a_{i,-t}^* \in \mathbb{R}^{T-1}$ and for each $\alpha_{i,t} < \alpha'_{i,t} < \alpha''_{i,t}$, $v_i(f(\theta), \alpha_{i,t}, a_{i,-t}^*) > v_i(f(\theta'), \alpha_{i,t}, a_{i,-t}^*)$ and $v_i(f(\theta), \alpha'_{i,t}, a_{i,-t}^*) = v_i(f(\theta'), \alpha'_{i,t}, a_{i,-t}^*)$ implies $v_i(f(\theta'), \alpha''_{i,t}, a_{i,-t}^*) < v_i^t(f(\theta'), \alpha''_{i,t}, a_{i,-t}^*)$.*

SCC-2 compares the allocations chosen for any two 'similar' states of nature, θ and θ'. These states 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 (Def. 6), the preferences induced by states θ and θ' only differ along the dimension of the period-t aggregator. The condition requires a

[19] These preferences are time-separable in the sense that each v_i^t does not depend on the allocations chosen in periods other than t. Function v_i^t, however, may depend on previous signals.



(a) SCC-1: for any ξ, ξ' that only differ at period t, *strict-single crossing* is satisfied with respect to the period-t aggregator.

(b) SCC-2: for any θ, θ' that are identical up to period-t and such that they only differ for the implied period-t aggregator, *strict single-crossing* is satisfied with respect to the period-t aggregator.

single-crossing condition for the corresponding outcomes to hold along this direction. From a graphical viewpoint, this condition can be interpreted as follows: suppose that θ and θ' are as in Definition 11. Then, if the SCF is aggregator-based and the environment satisfies SCC-1 (Def. 10), 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.b). If f is not aggregator based, allocations at periods $\tau > t$ may differ under $f(\theta)$ and $f(\theta')$, shifting the curve $\delta_{i,t}(f(\theta), f(\theta'), \alpha_{i,t})$. SCC-2 guarantees that the single-crossing property is maintained under this shift (Figure 1.b).[20]

The economic intuition of SCC-2 is also straightforward. In static environments, single-crossing conditions allow agents' types to monotonically sort themselves with respect to the chosen allocation. By requiring a single-crossing condition to hold within each period, condition SCC-1 (Def. 10) is the natural extension of the basic idea to a multi-period setting. With aggregator based SCF, such period-by-period single-crossing condition suffices to achieve full implementation in the dynamic mechanism (Proposition 3). Without the aggregator based assumption, letting types at time t sort themselves relative to period-t allocations does not guarantee that the required monotonicity also holds when forward-looking intertemporal considerations are taken into account. SCC-2 requires precisely such intertemporal effects to be sufficiently well-behaved that they do not upset the single-crossing property at time t. In this sense, SCC-2 can be seen as consisting of two components: the first is a standard single-crossing condition 'within' each period (as in SCC-1); the second component requires the intertemporal effects, which may alter the 'within period' incentives, not to be too strong.[21]

[20] Within the special case of time-separable preferences, SCC-2 is indeed quite permissive. For instance, the preferences in equation (15) satisfy SCC-2 for any choice of $F : \mathbb{R} \to \mathbb{R}$.

[21] In a Markovian setting, Pavan et al. (2013) introduce a single-crossing condition with a similar flavour, also to ensure that types sort themselves monotonically at every period, taking into account the intertemporal incentives. The latter are defined in terms of the agents' beliefs about the future, as entailed by the underlying stochastic process, and therefore have the Markovian property that at every period they are pinned down by the current signal $\theta_{i,t}$. This property of agents' beliefs is key to the tractability of Markovian settings. As explained in footnote 17, however, this property need not hold in the present, belief-free, setting: even if α_i^t were pinned down by the period-t signals, beliefs about the future would still be unrestricted and may depend, for instance, on previous signals. The two conditions therefore are very different at a formal level (also, the condition in Pavan et al. (2013) requires that preferences are differentiable, which is not assumed here.) Nonetheless, both

The next Proposition and Corollary state the main result of the paper.

Proposition 4 (Full Implementation) *In an environment with monotone aggregators (Def. 6) satisfying the* contraction property *(Def. 8), if a SCF f is* Strictly EPIC *(Definition 1) and satisfies SCC-2 (Def. 11), then $R^{\phi} = \{s^c\}$.*

Corollary 2 *Since $s^c \in S^*$, if the assumptions of Propositions 3 or 4 are satisfied, f is fully robustly implementable by the direct mechanism.*

5.3.3 Transferable Utility

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

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

where for each $t = 1, ..., T$, $v_i^t : \left(\times_{\tau=1}^{t} Q_\tau\right) \times \mathbb{R} \to \mathbb{R}$ is the period-t valuation of the common component. Notice that functions $v_i^t : \left(\times_{\tau=1}^{t} Q_\tau\right) \times \mathbb{R} \to \mathbb{R}$ are defined over the entire history $(q_1, ..., q_t)$: this allows period-t valuation of the current allocation (q_t) to depend on the previous allocative decisions $(q_1, ..., q_{t-1})$. This allows us to accommodate the path-dependencies in preferences discussed above.[22]

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

It is easy to verify that, in environments with transferable utility, if agents' preferences over the common component $Q^* = \times_{t=1}^{T} Q_t$ satisfy SCC-1, and $\chi : \Theta \to Q$ is aggregator-based, then for any transfer scheme $\left(\pi_{i,t}\left(y^t\right)\right)_{i=1}^{n}$, the SCF $f^{\chi,\pi}$ satisfies SCC-2. More generally, if χ and agents' preferences over Q^* satisfy SCC-2, then $f^{\chi,\pi}$ satisfies SCC-2 for any transfer scheme $\left(\pi_{i,t}\left(y^t\right)\right)_{i=1}^{n}$. Given this, the following corollary of Proposition 4 is immediate:

conditions ensure that types sort themselves monotonically with respect to the future outcomes. This is useful for applying recursive techniques, which is done here in distribution-free environments, and by Pavan et al. (2013) in Markovian settings.

[22] The special case of *path-independent preferences* corresponding to the example in section 5.2 is such that period-t valuation are functions $v_i^t : Q_t \times \mathbb{R} \to \mathbb{R}$.

Corollary 3 *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 π that make χ strictly ex-post incentive compatible. Then: $f^{\chi,\pi}$ is fully robustly implemented.*

6 Concluding Remarks

Alternative Approaches to Robustness. As already discussed in Section 3, the belief-free approach has been criticized for imposing an unnecessarily demanding notion of robustness. Nonetheless, it has provided an important theoretical benchmark, particularly to understand the methodological aspects of robust mechanism design. The analysis of belief-free full implementation has highlighted an important connection between robust mechanism design and strategic uncertainty, and identified rationalizability as a central concept in robust mechanism design.[23] The insights of the belief-free approach have also been extended by a more recent literature that studies environments in which the designer has some, albeit limited, information on agents' beliefs. Artemov, Kunimoto and Serrano (2013) pursue this approach within the context of virtual implementation, Kim and Penta (2013) and Lopomo, Rigotti and Shannon (2013) from the viewpoint of partial implementation, and Ollár and Penta (2014) in terms of full implementation. The latter paper also shows that the insights of the belief-free literature can be usefully adapted to address both more realistic notions of robustness, as well as to obtain full implementation via simple mechanisms, which consist of standard transfer schemes. Another strand of the literature has pursued a different, 'second best' approach (e.g., Börgers and Smith (2012a,b), and Yamashita (2012, 2013b, 2013c)). Börgers and Smith (2012a,b), for instance, show that letting the SCF depend on agents' beliefs may be useful in situations in which the first-best is belief-free but unattainable. Exploring these insights in dynamic settings seems a promising and important direction for future research, but understanding the limitations of the belief-free approach and the more fundamental methodological problems seems a natural first step to extend the robust approach to dynamic settings.

A 'Backward Induction Reasoning' Approach. The methodological results in Section 5.1 are based on solution concepts developed in Penta (2012b), which argues that the backwards procedure R^{ϕ} characterizes the predictions of backward induction reasoning in games with incomplete information. Two alternative epistemic characterizations are provided, in terms of 'common belief in future rationality' (CBFR), and in terms of 'common belief in rationality and in belief persistence' (CBRBP).[24] Intuitively, CBFR represents the idea, typical of backward induction reasoning, that deviations are unintentional, or 'mistakes' (a connection

[23]The connection between structural and strategic uncertainty, relevant to robust mechanism design, has emerged in several contexts. See, for instance, Bergemann and Morris (2009a), Oury and Tercieux (2012) and Yamashita (2013a). From a general game theoretic viewpoint, see Battigalli and Siniscalchi (2003), Weinstein and Yildiz (2007, 2013) and Penta (2012a,b, 2013).

[24]Perea (2012) independently introduced a related solution concept for complete information games, and provided a similar epistemic characterization to CBFR.

to ‘trembles’ is also explored). Under CBFR, players start out with common belief in rationality. If, at any point in the game, an unexpected move is observed, players may accept that the deviation was a mistake and maintain common belief in rationality for the continuation game.[25] CBRBP instead represents the idea, also inherent to backward induction reasoning, that players never change their beliefs about the continuation strategies of the opponents, even in the face of evidence that contradicts beliefs on earlier components of those strategies. In this sense, this paper has pursued a ‘backward induction’ approach. The related papers by Müller (2012a,b) instead adopt a solution concept based on the assumption of Common Strong Belief in Rationality (CSBR), which is a forward induction concept (Battigalli and Siniscalchi, 2002). Forward induction, however, is highly sensitive to the common knowledge assumptions, which may be problematic from the viewpoint of robustness. For instance, applying CSBR under common knowledge of no belief-restrictions (as formalized in belief-free models), need not be the same as applying it to all possible type spaces (cf. Müller (2012a,b)).

Dynamic Mechanisms in Static Environments. Consider an environment in which agents obtain all the relevant information before the planner has to make a decision. Although static mechanisms may be a viable option, the designer may still have reasons to adopt a dynamic mechanism (e.g. an ascending auction). In an environment with complete information, Bergemann and Morris (2007) recently argued that dynamic mechanisms may improve on static ones by reducing agents’ strategic uncertainty: They show that applying backward induction to an ascending auction guarantees full robust implementation for a larger set of parameters than applying rationalizability to its sealed-bid counterpart. Using the approach of this paper, it can be shown that this result is not robust to the introduction of incomplete information. The reason is that, under a backward induction logic, unexpected past reports may be interpreted as possibly unintentional mistakes, and as such they may be attritubed to any type of the opponents. If, as in belief-free settings, all restrictions on beliefs are relaxed, backward induction has essentially no bite in the ascending auction, because histories to do not ‘robustly’ convey information. The case for dynamic mechanisms in static environments therefore must rely on more complicated mechanisms or on stronger epistemic assumptions, which enable agents to draw stronger inferences from past histories. Müller (2012a) obtains some results in this sense, under CSBR. CSBR, however, is a strong epistemic assumption, which implies sophisticated forward induction reasoning. Also, Müller (2012a) adopts non-direct mechanisms with properties similar to the unbounded mechanisms used in the classical literature on implementation (see references in footnote 9). These mechanisms have been criticized, among others, by Jackson (1992).

[25] The view of deviations as ‘mistakes’ contrasts with the logic of *forward induction*, which requires instead that unexpected moves be rationalized (if possible) as purposeful deviations.

Appendix

A Topological structures and Conditional Probability Systems.

A.1 Topological structures.

Sets $\Theta_{i,t} \subseteq \mathbb{R}^{n_{i,t}}$ and $M_{i,t} \subseteq \mathbb{R}^{\nu_{i,t}}$ are non-empty and compact, for each i and t (Sections 2 and 3). Let $n_t = \sum_{i \in N} n_{i,t}$ and $\nu_t = \sum_{i \in N} \nu_{i,t}$. For each h_i^t, $\tau < t$, let $\rho_\tau\left(h_i^t\right)$ denote the tuple $(\theta_{i,\tau}, m_\tau)$ at period τ along history h_i^t. For each $k \in N$, let $d_{(k)}$ denote the Euclidean metric on $\mathbb{R}^k$. We endow the sets H_i with the following metrics, $d^i (i \in N)$, defined as: For each $h_i^t, h_i^\tau \in H_i$ (w.l.o.g.: let $\tau \geq t$) s.t. $h_i^t = ((\theta_{i,k}, m_k)_{k=1}^{t-1}, \theta_i^t)$ and $\hat{h}_i^\tau = ((\hat{\theta}_{i,k}, m_k)_{k=1}^{\tau-1}, \hat{\theta}_i^t)$,

$$d^i\left(h_i^t, \hat{h}_i^\tau\right) = \sum_{k=1}^{t-1} d_{\left(n_{i,k}+\nu_k\right)}\left(\rho_k\left(h_i^t\right), \rho_k\left(\hat{h}_i^\tau\right)\right) + d_{n_{i,t}}\left(\theta_{i,t}, \hat{\theta}_{i,t}\right) + \sum_{k=t+1}^{\tau} 1.$$

It can be checked that $\left(\mathcal{H}_i, d^i\right)$ are complete, separable metric spaces. Sets of strategies are endowed with the supmetrics d_{S_i} defined as:

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

Under these topological structures, lemma 2.1 in Battigalli (2003) implies that $S_i(h)$ is closed for every h, hence the CPSs introduced in the next section are well-defined.

A.2 Conditional Probability Systems

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

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

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

Axiom 3 *For all $A \in A$, $B, C \in C$, 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 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 Ω in C are the conditioning events, each inducing beliefs over Ω; $\Delta(\Omega)$ 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 $h_i^t = \left(h^{t-1}, y_i^t\right) \in \mathcal{H}_i$, the corresponding event $\left[h_i^t\right]$ is defined as:

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

Under the maintained assumptions and topological structures, sets $\left[h_i^t\right]$ 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 $\mu^i\left(h_i^t\right)$ instead of $\mu^i\left(\left[h_i^t\right]\right)$.

B The Backwards Procedure

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

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

For any $h_i^t = \left(h^{t-1}, y_i^t\right)$ and $\pi \in \Delta\left(\left\{y_i^t\right\} \times \left(\times_{\tau=t+1}^{T} \Theta_{i.\tau}\right) \times \Theta_{-i} \times S_{-i}^{h^{t-1}}\right)$, let $BR_i\left(\pi; h_i^t\right) \subseteq S_i^{h_i^t}$ denote the set of continuation strategies from h_i^t that are best response to conjectures π over the payoff states and the opponents' continuation strategies. That is:

$$BR_i\left(\pi; h_i^t\right) = \arg\max_{s_i \in S_i^{h_i^t}} \int_{(\theta, s_{-i}) \in \Theta \times S_{-i}^{h^{t-1}}} U_i\left(s_i, s_{-i}, \theta; h^{t-1}\right) \cdot d\pi$$

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

- **[t = T]** For each $h_i^T = \left(h^{T-1}, y_i^T\right)$, let $R_i^0\left(h_i^T\right) = S_i^{h_i^T}$, and for each $k = 1, 2, ...$, let

$$R_j^{k-1}\left(h^{T-1}\right) = \left\{(\theta_j, s_j) : s_j \in R_j^{k-1}\left(h^{T-1}, y_j^T(\theta_j)\right)\right\},$$
$$R_{-i}^{k-1}\left(h^{T-1}\right) = \times_{j \neq i} R_j^{k-1}\left(h^{T-1}\right),$$
$$R_i^k\left(h^{T-1}, y_i^T\right) = \left\{s_i \in R_i^{k-1}\left(h^{T-1}, y_i^T\right) : \begin{array}{ll} \text{(R.1)} & \exists \pi \in \Delta\left(\left\{y_i^T\right\} \times \Theta_{-i} \times R_{-i}^{k-1}\left(h^{T-1}\right)\right), \\ \text{(R.2)} & s_i \in BR_i\left(\pi; h^{T-1}, y_i^T\right) \end{array}\right\}$$
$$R_i\left(h^{T-1}, y_i^T\right) = \bigcap_{k=1}^{\infty} R_i^k\left(h^{T-1}, y_i^T\right) \text{ and } R_i\left(h^{T-1}\right) = \left\{\left(y_i^T, s_i\right) : s_i \in R_i^{k-1}\left(h^{T-1}, y_i^T\right)\right\}.$$

Notice that $R_i\left(h^{T-1}\right)$ consists of pairs of types $\theta_i = y_i^T$ and continuation strategies

$s_i \in S_i^{\left(h^{T-1}, y_i^T\right)}$, hence $R_i\left(h^{T-1}\right) \subseteq S_i^{h^{T-1}}$: an element of $R_i\left(h^{T-1}\right)$ therefore is a function from types to actions in the last stage following public history h^{T-1}.

- $[\mathbf{t} = \mathbf{T} - \mathbf{1}, \ldots \mathbf{0}]$ For each $h_i^t = \left(h^{t-1}, y_i^t\right)$, let

$$R_i^0\left(h^{t-1}, y_i^t\right) = \Big\{s_i \in S_i^{h_i^t} : \forall h^t \text{ s.t. } h^{t-1} \prec h^t, \\ \forall y_i^{t+1} = \left(y_i^t, \theta_{i,t+1}\right),\ s_i|\left(h^t, y_i^{t+1}\right) \in R_i\left(h^t, y_i^{t+1}\right)\Big\},$$

and for each $k = 1, 2, \ldots$, let

$$\begin{aligned}
R_j^{k-1}\left(h^{t-1}\right) &= \left\{(\theta_j, s_j) : s_j \in R_j^{k-1}\left(h^{t-1}, y_j^t(\theta_j)\right)\right\}, \\
R_{-i}^{k-1}\left(h^{t-1}\right) &= \times_{j \neq i} R_j^{k-1}\left(h^{t-1}\right), \\
R_i^k\left(h^{t-1}, y_i^t\right) &= \left\{s_i \in R_i^{k-1}\left(h^{t-1}, y_i^t\right) : \begin{array}{ll} \text{(R.1)} & \exists \pi \in \Delta\left(\left\{y_i^t\right\} \times \left(\times_{\tau=t+1}^T \Theta_{i.\tau}\right) \times \Theta_{-i} \times R_{-i}^{k-1}\left(h^{t-1}\right)\right), \\ \text{(R.2)} & s_i \in BR_i\left(\pi; h^{t-1}, y_i^t\right) \end{array}\right\} \\
R_i\left(h^{t-1}, y_i^t\right) &= \bigcap_{k=1}^{\infty} R_i^k\left(h^{t-1}, y_i^t\right) \text{ and } R_i\left(h^{t-1}\right) = \left\{\left(y_i^t, s_i\right) : s_i \in R_i^{k-1}\left(h^{t-1}, y_i^t\right)\right\}.
\end{aligned}$$

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

C Proof Proposition 2

The proof of Proposition 2 is based on a solution concept for belief-free games in extensive form, Backwards Extensive Form Rationalizability ($\mathcal{BR}$). Proposition 2 obtains by first showing that $\mathcal{BR}$ characterizes the set of IPE strategies (Proposition 5) and then that $\mathcal{BR}$ can be computed via the backwards procedure R^{ϕ} (Proposition 6).

Backwards Extensive Form Rationalizability. Similar to rationalizability, $\mathcal{BR}$ is a non-equilibrium solution concept. Agents form conjectures about everyone's behavior, which may or may not be consistent with each other. To avoid confusion, we refer to this kind of beliefs as 'conjectures', retaining the term 'beliefs' only for those introduced in Section 2. Agent i's conjectures are represented by CPS $\mu^i = \left(\mu^i\left(h_i^t\right)\right)_{h_i^t \in \mathcal{H}_i} \in \Delta^{\mathcal{H}_i}(\Theta \times S)$. Given a CPS $\mu^i \in \Delta^{\mathcal{H}_i}(\Theta \times S)$ and a history $h_i^t = \left(h^{t-1}, y_i^t\right)$, strategy s_i expected payoff at h_i^t, given μ^i, is defined as:

$$U_i\left(s_i, \mu^i; h_i^t\right) = \int_{\Theta \times S_{-i}} U_i\left(s_i, s_{-i}, \theta; h^{t-1}\right) \cdot dmarg_{\Theta \times S_{-i}} \mu^i\left(h_i^t\right). \tag{16}$$

Strategy s_i is sequentially rational with respect to $\mu^i \in \Delta^{\mathcal{H}_i}(\Theta \times S)$, written $s_i \in r_i\left(\mu^i\right)$, if and only if for each $h_i^t \in H_i$ and each $s_i' \in S_i$, $U_i\left(s_i, \mu^i; h_i^t\right) \geq U_i\left(s_i', \mu^i; h_i^t\right)$. If $s_i \in r_i\left(\mu^i\right)$, we say that conjectures μ^i 'justify' strategy s_i.

Definition 12 *For each $i \in N$, let $\mathcal{BR}_i^0 = S_i$. Define recursively, for $k = 1, 2, ..., \mathcal{BR}_i^{k-1} = \times_{i=1,...,n} \mathcal{BR}_i^{k-1}$, $\mathcal{BR}_{-i}^{k-1} = \times_{j \neq i} \mathcal{BR}^{k-1}$ and*

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

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

Proposition 5 *Fix a game $(\mathcal{E}, \mathcal{M})$. For each i: $\hat{s}_i \in \mathcal{BR}_i$ if and only if $\exists \mathcal{B} = (B_i, \beta_i)_{i \in N}$, $\hat{b}_i \in B_i$ and $(\hat{\sigma}, \hat{p})$ such that: (i) $(\hat{\sigma}, \hat{p})$ is an IPE of $(\mathcal{E}, \mathcal{M}, \mathcal{B})$ and (ii) $\hat{s}_i = \hat{\sigma}_i \left(\hat{b}_i\right)$.*

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

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

and for each h,

$$\bar{S}_j^h = \left\{s_j^h \in S_j^h : \exists s'_j \in \bar{S}_j \text{ s.t. } s'_j | h = s_j^h\right\}$$

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

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

where m_j^τ 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 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 $L_{h_i^t} : \Theta \times B_{-i} \times S_{-i} \to \Theta \times B \times S$ such that

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

(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 $E \subseteq \Theta \times B \times S$,

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

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

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

(Conditional beliefs $\lambda_i\left(h_i^t\right)$ at histories h_i^t s.t. $\lambda_i\left(h_i^{t-1}\right)\left[h_i^t\right]>0$ are determined via Bayesian updating, from the definition of CPS. See appendix A.2). Define the CPS $\mu^i \in \Delta^{\mathcal{H}_i}(\Theta\times S)$ s.t. $\forall h_i^t \in H_i$, $\mu^i\left(h_i^t\right)=\text{marg}_{\Theta\times S}\lambda_i\left(h_i^t\right)$. By construction, $\hat{s}_i \in r_i\left(\mu^i\right)$. We only need to show that conditions (2) and (3) in the definition of $\mathcal{BR}$ are satisfied by μ^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 μ^i constructed above satisfies $\text{supp}\left(\mu^i(\phi)\right)\subseteq \Theta\times\{\hat{s}_i\}\times\mathcal{BR}_{-i}^k$ and

$$supp\left(marg_{S^{h^{t-1}}}\mu^i(\phi)\right)=supp\left(marg_{S^{h^{t-1}}}\mu^i\left(h_i^t\right)\right)\subseteq \bar{S}^{h^{t-1}}$$

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

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

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

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

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

$$M_{s_i,h_i^t}(\theta,s_i,s_{-i})=\left(\theta,m_{h_i^t}(s_{-i})\right).$$

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

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

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

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

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

Next we show that $\mathcal{BR}$ can be computed via the backwards procedure R^{ϕ} introduced above.

Proposition 6 *$\mathcal{BR}_i = R_i^{\phi}$ for each i.*

Proof: **Step 1 ($R_i^{\phi} \subseteq \mathcal{BR}_i$):** let $\hat{s}_i \in R_i^{\phi}$. Then, for each $h_i^t = \left(h^{t-1}, y_i^t\right)$, $s_i|h_i^t \in R_i\left(h^{t-1}, y_i^t\right)$ (equivalently: $s_i^{h^{t-1}} \in R_i\left(h^{t-1}\right)$). Notice that for each h^{t-1} and $s_i^{h^{t-1}} \in R_i\left(h^{t-1}\right)$, there exists $s_i \in R_i^{\phi}$ 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 $\rho_j^{h^{t-1}} : R_j\left(h^{t-1}\right) \to R_j^{\phi}$ such that: $\forall\ s_j^{h^{t-1}} \in R_j\left(h^{t-1}\right)$

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

(Functions $\rho_j^{h^{t-1}}$ assign to strategies in $R_j\left(h^{t-1}\right)$, strategies in R_j^{ϕ} 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 $\varphi_j^{h^{t-1}} : S_j \to S_j\left(h^{t-1}\right)$ be a measurable function such that

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

where m_j^{τ} 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 $\times_{j\neq i}\varphi_j^{h^{t-1}}$.)

For each h^{t-1}, define the measurable mapping $\varrho_{-i}^{h^{t-1}} : R_{-i}\left(h^{t-1}\right) \to S_{-i}\left(h^{t-1}\right)$ such that $\forall s_{-i}^{h^{t-1}} \in R_{-i}\left(h^{t-1}\right)$,

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

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

The initial step is trivially satisfied ($\mathcal{BR}_i^0 = S_i = R_i^0(\phi)$). For the inductive step, suppose that the statement is true for $n = 0, ..., k-1$: Since $\hat{s}_i \in R_i^k(\phi)$, for each $h_i^t = \left(h^{t-1}, y_i^t\right)$ there exists $\pi^{h_i^t} \in \Delta\left(\Theta \times S_{-i}^{h^{t-1}}\right)$ that satisfies

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

and such that $\pi^{\phi}\left(\Theta \times R_{-i}^{k-1}(\phi)\right) = 1$ and for all $h_i^t \neq \phi$,

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

Now, consider the CPS $\mu^i \in \Delta^{\mathcal{H}_i}(\Theta \times S)$ such that, for all measurable $E \subseteq \Theta \times S_{-i}$,

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

By definition of CPS, $\mu^i(\phi)$ defines $\mu\left(h_i^t\right)$ for all h_i^t s.t. $\mu^i(\phi)\left[h_i^t\right] > 0$. Let h_i^t be such that $\mu^i(\phi)\left[h_i^{t-1}\right] > 0$ and $\mu^i(\phi)\left[h_i^t\right] = 0$. Define the measurable mapping $M_{h_i^t} : \Theta \times R_{-i}^{h^{t-1}} \to$

$\Theta \times S\left(h^{t-1}\right)$ such that for all $\left(\theta, s_{-i}^{h^{t-1}}\right) \in \Theta \times S\left(h^{t-1}\right)$,

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

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

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

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

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

Step 2 ($\mathcal{BR}_i \subseteq R_i^\phi$): let $\hat{s}_i \in R_i^\phi$ and $\mu^i \in \Delta^{\mathcal{H}_i}(\Theta \times S)$ be such that $\hat{s}_i \in r_i\left(\mu^i\right)$. For each $h_i^t = \left(h^{t-1}, y_i^t\right)$, define the mapping $\psi_{h_i^t} : S_{-i} \to S_{-i}^{h^{t-1}}$ s.t. $\psi_{h_i^t}\left(s_{-i}\right)|h^{t-1} = s_{-i}|h^{t-1}$ 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 $\Psi_{h_i^t} : \Theta \times S \to \Theta \times S_{-i}^{h^{t-1}}$ such that

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

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

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

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

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

The argument proceeds by induction on the length of histories.

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

Inductive Step: suppose that for each $\tau = t+1, ..., T$, $\hat{s}_i \in \mathcal{BR}_i$, implies $\hat{s}_i|h_i^\tau \in R_i\left(h^{\tau-1}, y^\tau\right)$ for each $h_i^\tau = \left(h^{\tau-1}, y_i^\tau\right)$. We will show that for each k, $h_i^t = \left(h^{t-1}, y_i^t\right)$, $\hat{s}_i|h_i^t \in R_i^k\left(h^{t-1}, y^t\right)$.

We proceed by induction on k: under the inductive hypothesis on τ, $\hat{s}_i|h_i^t \in R_i^0\left(h^{t-1}, y^t\right)$. For the inductive step on k, suppose that $\hat{s}_i \in \mathcal{BR}_i$, implies $\hat{s}_i|h_i^t \in R_i^n\left(h^{t-1}, y^t\right)$ for $n = 0, ..., k-1$, and suppose (as contrapositive) that $\hat{s}_i|h_i^t \notin R_i^k\left(h^{t-1}, y^t\right)$. Then, for $\pi^{h_i^t}$ defined as above, it must be that $\text{supp}\left(\pi^{h_i^t}\right) \nsubseteq \Theta \times R_{-i}^{k-1}\left(h^{t-1}\right)$, which, under the inductive hypothesis on n, implies that $\exists s_{-i} \in \text{supp}\left(marg_{S_{-i}}\mu^i\left(h_i^t\right)\right)$ s.t. $\nexists s'_{-i} \in \mathcal{BR}_{-i} : s'_{-i}|h^{t-1} = s_{-i}|h^{t-1}$, which contradicts that μ^i justifies $\hat{s}_i$ in $\mathcal{BR}_i$.■

Proposition 2 follows from propositions 5 and 6. QED.

D Implementation Results

D.1 Proof of Proposition 1

As explained in the main text, the 'only if' part follows from Bergemann and Morris (2005). For the 'if' part, let f be EPIC, and fix a model of beliefs $\mathcal{B} = (B_i, \beta_i)_{i \in N}$. It will be shown that there exists a truthful PBE of $\left(\mathcal{E}, \mathcal{M}^f, \mathcal{B}\right)$. Let σ^* be a 'strongly truthful' profile, in the sense that for every i, and for every h_i^t, period-t signal $\theta_{i,t}$ is reported truthfully. Let belief system $\left(p^i\right)_{i \in N}$ be consistent with σ^* and such that: $\forall i \in N$, $\forall b_i \in B_i$, $p^i\left(b_i, \phi\right) = \beta_i\left(b_i\right)$ and for each $h_i^t = \left(h^{t-1}, y_i^t\right) \in H_i$, where $h^{t-1} = \left(\tilde{y}_i, \tilde{y}_{-i}^{t-1}\right)$,

$$supp\left(marg_{\Theta_{-i}} p^i\left(b_i, h_i^t\right)\right) \subseteq \left\{\tilde{y}_{-i}^t\right\} \times \left(\times_{\tau=t+1}^T \Theta_{-i,\tau}\right). \tag{17}$$

That is, at unexpected histories, each i believes that the opponents have reported truthfully: If unexpected reports were observed, player i revises his beliefs about the opponents' types, not their behavior. Notice that if $U_i\left(s^*, \theta\right) \geq U_i\left(s'_i, s^*_{-i}, \theta\right)$ for all θ (cf. Def. 1), then for any $\pi^i \in \Delta\left(\Theta \times B_{-i}\right)$,

$$\int_{\Theta \times B_{-i}} U_i\left(\sigma^*, \theta, b_{-i}; b_i, \phi\right) \cdot d\pi^i \geq \int_{\Theta \times B_{-i}} U_i\left(s'_i, \sigma^*_{-i}, \theta, b_{-i}; b_i, \phi\right) \cdot d\pi^i. \tag{18}$$

Hence, the incentive compatibility constraints are satisfied at the beginning of the game, i.e. for $\pi^i = p^i\left(b_i, \phi\right)$, and so at all histories 'on the path'. At other histories, the belief system satisfies (17), which implies that for every h_i^t,

$$\int_{\Theta \times B_{-i}} U_i\left(\sigma^*, \theta, b_{-i}; b_i, h_i^t\right) \cdot dp^i\left(b_i, h_i^t\right) = \int_{\Theta \times B_{-i}} U_i\left(\sigma^*, \theta, b_{-i}; b_i, \phi\right) \cdot dp^i\left(b_i, h_i^t\right)$$

and

$$\int_{\Theta \times B_{-i}} U_i\left(s'_i, \sigma^*_{-i}, \theta, b_{-i}; h_i^t\right) \cdot dp^i\left(b_i, h_i^t\right) = \int_{\Theta \times B_{-i}} U_i\left(s'_i, \sigma^*_{-i}, \theta, b_{-i}; b_i, \phi\right) \cdot dp^i\left(b_i, h_i^t\right).$$

But then, for each h_i^t and letting $\pi^i = p^i\left(b_i, h_i^t\right)$, equation (18) implies that:

$$\int_{\Theta \times B_{-i}} U_i\left(\sigma^*, \theta, b_{-i}; b_i, h_i^t\right) \cdot dp^i\left(b_i, h_i^t\right) \geq \int_{\Theta \times B_{-i}} U_i\left(s'_i, \sigma^*_{-i}, \theta, b_{-i}; b_i, h_i^t\right) \cdot dp^i\left(b_i, h_i^t\right).$$

That is, σ^* is sequentially rational with respect to p. By construction, (p, σ^*) is a truthful PBE of the Bayesian game.■

D.2 Proof of Proposition 3.

By contradiction, suppose $R^{\phi} = D \neq \{s^c\}$. By continuity of u_i and compactness of Θ, $D\left(h^t\right)$ is compact for each h^t. It will be shown that for each t and for each public history h^{t-1}, $s\left[D\left(h^{t-1}\right)\right] = s^c\left[h^{t-1}\right]$, 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: [It will be proven that $s\left[D\left(h^{T-1}\right)\right] = s^c\left[h^{T-1}\right]$ for each h^{T-1}]. Suppose, by contradiction, that $\exists h^{T-1} = \left(\tilde{y}^{T-1}, x^{T-1}\right) : s\left[D\left(h^{T-1}\right)\right] \neq s^c\left(h^{T-1}\right)$. Then, by the contraction property,

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

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

$$\delta\left(h^{T-1}, y_i^T\right) := \min_{\left(y_{-i}^T, \theta'_{-i,T}\right) \in Y_{-i}^T \times D_{-i}\left(h^{T-1}, y_{-i}^T\right)} \left[\alpha_{i,T}\left(y_i^T, y_{-i}^T\right) - \alpha_{i,T}\left(\tilde{y}^{T-1}, \theta'_{i,t}, \theta'_{-i,t}\right)\right] \quad (19)$$

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

For any $\varepsilon > 0$, let

$$\psi\left(h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon\right) = \max_{\theta_{-i,T} \in \Theta_{-i,T}} \left\{\alpha_{i,T}\left(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}\right) - \alpha_{i,T}\left(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}\right)\right\} \quad (20)$$

(again, compactness of $\Theta_{-i,T}$ guarantees that $\psi(\cdot)$ is well-defined). Since $\alpha_{i,T}$ is strictly increasing in $\theta_{i,T}$, $\psi\left(h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon\right)$ is increasing in ε and $\psi\left(h^{T-1}, y_i^T, \theta'_{i,T}, \varepsilon\right) \to 0$ as $\varepsilon \to 0$. Let $\left(f_t\left(\tilde{y}^t\right)\right)_{t=1}^{T-1} = x^{T-1}$. From strict EPIC, we have that for each ε,

$$\begin{aligned}
&v_i\left(x^{T-1}, f_T\left(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}\right), \alpha_{i,T}\left(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}\right), \alpha_{i,-T}\left(\tilde{y}^{T-1}\right)\right) \\
&> v_i\left(x^{T-1}, f_T\left(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}\right), \alpha_{i,T}\left(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}\right), \alpha_{i,-T}\left(\tilde{y}^{T-1}\right)\right) \\
&\text{and} \\
&v_i\left(x^{T-1}, f_T\left(\tilde{y}^{T-1}, \theta'_{i,T} + \varepsilon, \theta_{-i,T}\right), \alpha_{i,T}\left(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}\right), \alpha_{i,-T}\left(\tilde{y}^{T-1}\right)\right) \\
&< v_i\left(x^{T-1}, f_T\left(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}\right), \alpha_{i,T}\left(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}\right), \alpha_{i,-T}\left(\tilde{y}^{T-1}\right)\right)
\end{aligned}$$

Thus, by continuity, there exists $a_{i,T}(\varepsilon)$ such that

$$\alpha_{i,T}\left(\tilde{y}^{T-1},\theta'_{i,T},\theta_{-i,T}\right) < a_{i,T}(\varepsilon) < \alpha_{i,T}\left(\tilde{y}^{T-1},\theta'_{i,T}+\varepsilon,\theta_{-i,T}\right) \tag{21}$$

$$\text{such that}$$

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

From single-crossing condition SCC-1 (Def. 10),

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

Thus, to reach the contradiction, it suffices to show that for any $y^T_{-i} \in Y^T_{-i}$, $\alpha_{i,T}\left(y^T_i, y^T_{-i}\right) > a_{i,T}(\varepsilon)$: If this is the case, reporting $\theta'_{i,T}$ is (conditionally) strictly dominated by reporting $\theta'_{i,T}+\varepsilon$ at $h^T_i = \left(h^{T-1}, y^T_i\right)$, hence it cannot be that $D_i = R^{\phi}_i$ and $\theta'_{i,T} \in D_i\left(h^{T-1}, y^T_i\right)$. To this end, it suffices to choose ε sufficiently small that

$$\psi\left(h^{T-1}, y^T_i, \theta'_{i,T}, \varepsilon\right) < \delta\left(h^{T-1}, y^T_i\right) \tag{22}$$

and operate the substitutions as follows:

$$\begin{aligned} \alpha_{i,T}\left(y^T_i, y^T_{-i}\right) &\geq \alpha_{i,T}\left(\tilde{y}^{T-1},\theta'_{i,T},\theta'_{-i,T}\right) + \delta\left(h^{T-1}, y^T_i\right) \\ &\geq \alpha_{i,T}\left(\tilde{y}^{T-1},\theta'_{i,T}+\varepsilon,\theta'_{-i,T}\right) + \delta\left(h^{T-1}, y^T_i\right) - \psi\left(h^{T-1}, y^T_i, \theta'_{i,T}, \varepsilon\right) \\ &> \alpha_{i,T}\left(\tilde{y}^{T-1},\theta'_{i,T}+\varepsilon,\theta'_{-i,T}\right) \\ &> a_{i,T}(\varepsilon) \end{aligned}$$

(the first inequality follows from (19), the second from (20), the third from (22) and the fourth from (21)). Thus: $\alpha_{i,T}\left(y^T_i, y^T_{-i}\right) > a_{i,T}(\varepsilon)$ for any y^T_{-i}. This concludes the initial step.

Inductive Step: [for $t = 1, ..., T$, it will be shown that if $s\left[D\left(h^{\tau}\right)\right] = s^c\left[h^{\tau}\right]$ for all h^{τ} s.t. $\tau \geq t$, then $s\left[D\left(h^{t-1}\right)\right] = s^c\left[h^{t-1}\right]$ for all h^{t-1}]. By contradiction, suppose that there exists $h^{t-1} = \left(\tilde{y}^{t-1}, x^{t-1}\right) : s\left[D\left(h^{t-1}\right)\right] \neq s^c\left(h^{t-1}\right)$. Then, by the contraction property,

$$\begin{aligned} &\exists y^t_i \text{ and } \theta'_{i,t} \in D_i\left(h^{t-1}, y^t_i\right) : \theta'_{i,t} \neq s^c_i\left(h^{t-1}, y^t_i\right) \text{ such that:} \\ &sign\left[s^c_i\left(h^{t-1}, y^t_i\right) - \theta'_{i,t}\right] = sign\left[\alpha_{i,t}\left(y^t_i, y^t_{-i}\right) - \alpha_{i,t}\left(\tilde{y}^{t-1},\theta'_{i,t},\theta'_{-i,t}\right)\right] \\ &\text{for all } y^t_{-i} = \left(y^{t-1}_{-i}, \theta_{-i,t}\right) \text{ and } \theta'_{-i,t} \in D_{-i}\left(h^{t-1}, y^t_{-i}\right). \end{aligned}$$

Fix such y^t_i and $\theta'_{i,t} \neq s^c_i\left(h^{t-1}, y^t_i\right)$, and suppose that $s^c_i\left(h^{t-1}, y^t_i\right) > \theta'_{i,t}$. Similar to the initial step, it will be shown that there exists $\theta^{\varepsilon}_{i,t} = \theta'_{i,t} + \varepsilon$ for some $\varepsilon > 0$ such that for any conjecture consistent with D_{-i}, playing $\theta^{\varepsilon}_{i,t}$ is strictly better than playing $\theta'_{i,t}$ at history

$h_i^t = \left(h^{t-1}, y_i^t\right)$, contradicting the hypothesis that $R^{\phi} = D$. For any $\varepsilon > 0$, set $\theta_{i,t}^{\varepsilon} = \theta_{i,t}' + \varepsilon$; for each realization of signals $\tilde{\theta}_i = \left(\tilde{\theta}_{i,k}\right)_{k=1}^{T}$ and opponents' reports $\tilde{m}_{-i} = \left(\tilde{m}_{-i,k}\right)_{k=t}^{T}$, for each $\tau > t$, denote by $s_{i,\tau}^{c}\left(\theta_{i,t}^{\varepsilon}, \tilde{m}_{-i}, \tilde{\theta}_i\right)$ the action taken at period τ 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 properties 1 and 2 of the aggregators functions (Def. 6) and definition of s^c (Def. 7), $s_{i,\tau}^{c}\left(\theta_{i,t}^{\varepsilon}, \tilde{m}_{-i}, \tilde{\theta}_i\right)$ is continuous in ε, and converges to $s_{i,\tau}^{c}\left(\theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_i\right)$ as $\varepsilon \to 0$.)

For each realization $\tilde{\theta}_i = \left(\tilde{\theta}_{i,k}\right)_{k=1}^{T}$ and reports $\tilde{m}_{-i} = \left(\tilde{m}_{-i,k}\right)_{k=t}^{T}$ and for each $\tau > t$, $s_{i,\tau}^{c}\left(\theta_{i,t}', \tilde{m}_{-i}, \tilde{\theta}_i\right)$ may be one of five cases (cf. equations 9-11):

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

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

   and we can choose ε sufficiently small that $s_{i,\tau}^{c}\left(\theta_{i,t}^{\varepsilon}, \tilde{m}_{-i}, \tilde{\theta}_i\right) \in \left(\theta_{i,T}^{l}, \theta_{i,T}^{h}\right)$, and hence

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

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

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

   at $\hat{y}_{-i}^{\tau}\left(h_i^{\tau}\right)$ defined as in equation (8), then we can choose ε sufficiently small that $s_{i,\tau}^{c}\left(\theta_{i,t}^{\varepsilon}, \tilde{m}_{-i}, \tilde{\theta}_i\right) = \theta_{i,T}^{h}$ as well.

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

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

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

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

In either case,

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

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

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

at $\hat{y}_{-i}^\tau \left(h_i^t \right)$ defined as in equation (8), then we can choose ε sufficiently small that $s_{i,\tau}^c \left(\theta_{i,t}^\varepsilon, \tilde{m}_{-i}, \tilde{\theta}_i \right) = \theta_{i,T}^l$ as well.

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

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

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

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

In either case,

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

That is, for each $\tau > t$, and for each $\left(\tilde{\theta}_i, \tilde{m}_{-i} \right)$, in all five cases there exists $\bar{\varepsilon} \left(\tilde{\theta}_i, \tilde{m}_{-i}, \tau \right) > 0$ such that:

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

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

$\tau > t$ (this is because 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 by following s_i^c at periods $\tau > t$ and playing $\theta'_{i,t}$ or $\theta^{\varepsilon}_{i,t}$ at history h_i^t, respectively ξ' and ξ^{ε}, are such that $\xi'_{\tau} = \xi^{\varepsilon}_{\tau}$ for all $\tau \neq t$.

Consider types of player i, $\theta'_i, \theta^{\varepsilon}_i \in \Theta_i$ such that for each $\tau < t$, $\theta'_{i,\tau} = \theta^{\varepsilon}_{i,\tau} = \hat{\theta}_{i,\tau}$ (where $\hat{\theta}_{i,\tau}$are the types actually reported on the path); for all $\tau > t$, $\theta_{i,\tau} = s^c_{i,\tau}$ as above; while at t respectively equal to $\theta^{\varepsilon}_{i,t}$ and $\theta'_{i,t}$. Thus, the induced allocations are ξ^{ε} and ξ' discussed above, and for each $\tau \neq t$, $\alpha_{i,\tau}(\theta^{\varepsilon}) = \alpha_{i,\tau}(\theta') \equiv \hat{a}_{i,\tau}$.

From strict EPIC, we have that for any θ_{-i}

$$v_i \left(\xi^{\varepsilon}, \alpha_{i,t} (\theta^{\varepsilon}_i, \theta_{-i}), \{\hat{a}_{i,\tau}\}_{\tau \neq t} \right) > v_i \left(\xi', \alpha_{i,t} (\theta^{\varepsilon}_i, \theta_{-i}), \{\hat{a}_{i,\tau}\}_{\tau \neq t} \right)$$

and

$$v_i \left(\xi^{\varepsilon}, \alpha_{i,t} (\theta'_i, \theta_{-i}), \{\hat{a}_{i,\tau}\}_{\tau \neq t} \right) < v_i \left(\xi', \alpha_{i,t} (\theta'_i, \theta_{-i}), \{\hat{a}_{i,\tau}\}_{\tau \neq t} \right)$$

Thus, by continuity, there exists $a_{i,t}(\varepsilon) \in \left(\alpha_{i,t} \left(\tilde{y}^{t-1}, \theta'_{i,t}, \theta_{-i,t} \right), \alpha_{i,t} \left(\tilde{y}^{t-1}, \theta^{\varepsilon}_{i,t}, \theta_{-i,t} \right) \right)$ such that $v_i \left(\xi^{\varepsilon}, a_{i,t}(\varepsilon), \{\hat{a}_{i,\tau}\}_{\tau \neq t} \right) = v_i \left(\xi', a_{i,t}(\varepsilon), \{\hat{a}_{i,\tau}\}_{\tau \neq t} \right)$. From SCC-1,

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

whenever $a^* > a_{i,t}(\varepsilon)$.

Thus, since the continuations in periods $\tau > t$ are the same under both $\theta'_{i,t}$ and $\theta^{\varepsilon}_{i,t}$, to reach the desired contradiction it suffices to show that for any $y^t_{-i} \in Y^t_{-i}$, $\alpha_{i,t}\left(y^t_i, y^t_{-i}\right) > a^t(\varepsilon)$. (This, for any realization of $\tilde{\theta}_{-i}$).

As in the initial step, define:

$$\delta := \min_{\left(y^t_{-i}, \theta'_{-i,t}\right) \in Y^t_{-i} \times B_{-i}\left(h^{t-1}, y^t_{-i}\right)} \left[\alpha_{i,t}\left(y^t_i, y^t_{-i}\right) - \alpha_{i,t}\left(\tilde{y}^{t-1}, \theta'_{i,t}, \theta'_{-i,t}\right) \right] \tag{23}$$

For any $\varepsilon > 0$, let

$$\psi(\varepsilon) = \max_{\theta_{-i,t} \in \Theta_{-i,t}} \left\{ \alpha_{i,t}\left(\tilde{y}^{t-1}, \theta^{\varepsilon}_{i,t}, \theta_{-i,t}\right) - \alpha_{i,t}\left(\tilde{y}^{t-1}, \theta'_{i,t}, \theta_{-i,t}\right) \right\} \tag{24}$$

Since $\alpha_{i,t}$ is strictly increasing in $\theta_{i,t}$, $\psi(\varepsilon)$ is increasing in ε and $\psi(\varepsilon) \to 0$ as $\varepsilon \to 0$. To obtain the desired contradiction, it suffices to choose ε sufficiently small that $\psi(\varepsilon) < \delta$, and operate the substitutions as follows:

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

D.3 Proof of Proposition 4.

The proof is very similar to that of Proposition 3.

Initial Step: $[s\left[D\left(h^{T-1}\right)\right] = s^c\left[h^{T-1}\right]$ for each $h^{T-1}]$. The initial step is the same, to conclude (in analogy with equation 21), that there exists $a^T(\varepsilon)$ such that

$$\alpha_{i,T}\left(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}\right) < a_{i,T}(\varepsilon) < \alpha_{i,T}\left(\tilde{y}^{T-1}, \theta'_{i,T}+\varepsilon, \theta_{-i,T}\right) \tag{25}$$

such that

$$\begin{aligned}&v_i\left(x^{T-1}, f_T\left(\tilde{y}^{T-1}, \theta'_{i,T}+\varepsilon, \theta_{-i,T}\right), a_{i,T}(\varepsilon), \alpha_{i,-T}\left(\tilde{y}^{T-1}\right)\right) \\ &= v_i\left(x^{T-1}, f_T\left(\tilde{y}^{T-1}, \theta'_{i,T}, \theta_{-i,T}\right), a_{i,T}(\varepsilon), \alpha_{i,-T}\left(\tilde{y}^{T-1}\right)\right)\end{aligned} \tag{26}$$

Then, SCC-2 (Def. 11) implies that

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

whenever $a^* > a^T(\varepsilon)$. From this point, the argument proceeds unchanged, concluding the initial step.

Inductive Step: [for $t = 1, ..., T-1$: if $s[D(h^\tau)] = s^c[h^\tau]$ for all h^τ and all $\tau > t$ then $s\left[D\left(h^t\right)\right] = s^c\left[h^t\right]$ for all h^t]. The argument proceeds as in Proposition 3, to show that for each $\tau > t$, and for each $\left(\tilde{\theta}, \tilde{m}_{-i}\right)$, if continuation strategies are self-correcting, there exists $\bar{\varepsilon}\left(\tilde{\theta}, \tilde{m}_{-i}, \tau\right) > 0$ such that:

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

Consider types of player i, $\theta'_i, \theta^{\varepsilon}_i \in \Theta_i$ such that for each $\tau < t$, $\theta'_{i,\tau} = \theta^{\varepsilon}_{i,\tau} = \hat{\theta}_{i,\tau}$ (the one actually reported on the path), and for all $\tau > t$, $\theta_{i,\tau} = s^c_{i,\tau}$ as above, while at t respectively equal to $\theta^{\varepsilon}_{i,t}$ 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 θ_{-i}

$$v_i\left(\xi^{\varepsilon},\alpha_{i,t}\left(\theta_i^{\varepsilon},\theta_{-i}\right),\{\hat{a}_{i,\tau}\}_{\tau\neq t}\right)>v_i\left(\xi',\alpha_{i,t}\left(\theta_i^{\varepsilon},\theta_{-i}\right),\{\hat{a}_{i,\tau}\}_{\tau\neq t}\right)$$

and

$$v_i\left(\xi^{\varepsilon},\alpha_{i,t}\left(\theta_i',\theta_{-i}\right),\{\hat{a}_{i,\tau}\}_{\tau\neq t}\right)<v_i\left(\xi',\alpha_{i,t}\left(\theta_i',\theta_{-i}\right),\{\hat{a}_{i,\tau}\}_{\tau\neq t}\right)$$

Thus, by continuity, there exists $a_{i,t}(\varepsilon)$

$$\alpha_{i,t}\left(\tilde{y}^{t-1},\theta'_{i,t},\theta_{-i,t}\right)<a_{i,t}(\varepsilon)<\alpha_{i,t}\left(\tilde{y}^{t-1},\theta^{\varepsilon}_{i,t},\theta_{-i,t}\right)\tag{27}$$

such that

$$v_i\left(\xi^{\varepsilon},a_{i,t}(\varepsilon),\{\hat{a}_{i,\tau}\}_{\tau\neq t}\right)=v_i\left(\xi',a_{i,t}(\varepsilon),\{\hat{a}_{i,\tau}\}_{\tau\neq t}\right)$$

From SCC-2 (Def. 11),

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

whenever $a^*>a_{i,t}(\varepsilon)$

To reach the desired contradiction it suffices to show that for any $y^t_{-i}\in Y^t_{-i}$, $\alpha_{i,t}\left(y_i^t,y^t_{-i}\right)>a_{i,t}(\varepsilon)$. The remaining part of the proof is identical to Proposition 3.□

D.4 'Quasi-direct' Mechanisms.

This section shows how simple enlarged mechanisms may avoid incurring into the problem of corner solutions, which allows us to modify the contraction property (Definition 8) by guaranteeing that the sign condition holds with $\kappa_i\left(h^{t-1},y^t\right)=0$ at every history (equation 14). This way, the dynamic contraction property is more directly comparable with Bergemann and Morris' (2009a) static counterpart .

Let $\hat{\alpha}_{i,t}:\mathbb{R}^{nt}\to\mathbb{R}$ be a continuous extension of $\alpha_{i,t}:Y^t\to\mathbb{R}$ from Y^t to $\mathbb{R}^{nt}$, strictly increasing in the component that extends $\theta_{i,t}$ and constant in all the others on $\mathbb{R}^{nt}\backslash Y^t$. Set $m^-_{i,1}=\theta^l_{i,1}$ and $m^+_{i,1}=\theta^h_{i,1}$, and for each $t=1,...,T$, let $\hat{\Theta}_{i,t}=\left[m^-_{i,t},m^+_{i,t}\right]$, and $\hat{Y}^t_i=\times_{\tau=1}^t\hat{\Theta}_{i,\tau}$ where for all $t=2,...,T$, $m^-_{i,t}$ and $m^+_{i,t}$ are recursively defined so as to satisfy:

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

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

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 $h_i^t=\left(h^{t-1},y_i^t\right)$, the self-correcting report $s_i^c\left(h_i^t\right)$ satisfies equation (9), that is s^c is capable of fully offsetting previous misreports: messages in $\hat{\Theta}_{i,t}\backslash\Theta_{i,t}$

are used whenever equations (10) or (11) 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 $\hat{\Theta}_{i,t} \backslash \Theta_{i,t}$ has been reported at some period $\tau \leq t$, let $g_t\left(\hat{y}^t\right) = f_t(\theta)$ for some θ such that $\alpha_{i,\tau}(\theta) = \alpha_{i,\tau}(\hat{y}^\tau)$ for all i and $\tau \leq t$, $f_t(\theta) = f_t(\theta')$.

References

1. Artemov, G., T. Kunimoto and R. Serrano (2013) “Robust Virtual Implementation with Incomplete Information: Towards a Reinterpretation of the Wilson Doctrine”, *Journal of Economic Theory,* vol. 148, (2013), 424-447.

2. Athey, S. and I. Segal (2014), “An Efficient Dynamic Mechanism”, *Econometrica*, forthcoming

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 (2002), “Strong Belief and Forward Induction Reasoning”, *Journal of Economic Theory*, **106**, 356–391.

5. Battigalli, P. and M. Siniscalchi (2003), “Rationalization and Incomplete Information ,” *The B.E. Journal of Theoretical Economics* 3(1).

6. Bergemann, D. and S. Morris (2005) “Robust Mechanism Design”, *Econometrica* 73(6), 1771-1813.

7. Bergemann, D. and S. Morris (2007) “An Ascending Auction for Interdependent Values: Uniqueness and Robustness to Strategic Uncertainty”, *AEA papers and proceedings*, May 2007.

8. Bergemann, D. and S. Morris (2009a) “Robust Implementation in Direct Mechanisms”, *Review of Economic Studies*, **76**, 1175-1204.

9. Bergemann, D. and S. Morris (2009b) “Robust virtual implementation,” *Theoretical Economics*, 4(1).

10. Bergemann, D. and S. Morris (2011) “Robust Implementation in General Mechanisms”, *Games and Economic Behavior,* 71(2), pp. 261-281.

11. Bergemann, D. and J. Valimaki (2010) “The Dynamic Pivot Mechanism,” *Econometrica*, 78(2), pages 771-789.

12. Börgers, T. and T. McQuade (2007) “Information Invariant Equilibria of Extensive Games,” *The B.E. Journal of Theoretical Economics (Contributions)*, Vol. 7 (2007), Article 42

13. Börgers and Smith (2012a) "Robustly Ranking Mechanisms", *American Economic Review Papers and Proceedings*, (2012) 325-329

14. Börgers and Smith (2012b) "Robust Mechanism Design and Dominant Strategy Voting Rules," *Theoretical Economics*, forthcoming

15. Fudenberg, D. and J. Tirole (1991), "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53, 236-290.

16. Fudenberg, D. and J. Tirole (1995), *Game Theory*, MIT Press, Cambridge (MA).

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. Jackson, M. (1991), "Bayesian Implementation", *Econometrica* 59, 461-477.

19. Jackson, M. O. (1992), "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," *Review of Economic Studies,* 59(4), 757–75.

20. Kim, K. and A. Penta (2013) "Efficient Auctions and Robust Mechanism Design: A New Approach," *mimeo*, UW-Madison

21. Lee, J. and H., Sabourian (2011) "Efficient Repeated Implementation," *Econometrica*, 79(6), 1967-1994.

22. Lopomo, G., L. Rigotti and C. Shannon (2013) "Uncertainy in Mechanism Design," *mimeo.*

23. Maskin, E. (1999), "Nash Equilibrium and Welfare Optimality", *Review of Economic Studies*, Vol. 66 (1), pp. 23-38.

24. Müller, C. (2012a) "Robust Virtual Implementation under Common Strong Belief in Rationality" *mimeo*, Carnegie-Mellon.

25. Müller, C. (2012b) "Robust Implementation in Weakly Rationalizable Strategies" *mimeo*, Carnegie-Mellon.

26. Ollár, M. and A. Penta (2014) "Full Implementation and Belief Restrictions", *mimeo*, Univ. of Wisconsin-Madison.

27. Oury, M. and O. Tercieux (2012), "Continuous Implementation," *Econometrica*, 80, 1605-1637.

28. Palfrey, T. and Srivastrava (1989) "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem", *Journal of Political Economy*, 97, 668-691.

29. Pavan, A., Y Segal and J. Toikka (2011) "Dynamic Mechanism Design" *mimeo*, Stanford Univ.

30. Penta, A. (2012a), "Higher Order Uncertainty and Information: Static and Dynamic Games", *Econometrica*, Vol. 80, No. 2 (March, 2012), 631-660.

31. Penta, A. (2012b), "Backward Induction Reasoning in Games with Incomplete Information'", *mimeo*, Univ. of Wisconsin-Madison.

32. Penta, A. (2013), "On the Structure of Rationalizability on Arbitrary Spaces of Uncertainty", *Theoretical Economics*, Vol.8, 405-430.

33. Perea, A. (2014), "Belief in the Opponents'Future Rationality", *Games and Economic Behavior*, Vol. 83, 231–254.

34. Postlewaite, A. and D. Schmeidler (1988) "Implementation in Differential Information Economies", *Journal of Economic Theory*, 39, 14-33.

35. 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.

36. Weinstein, J. and M. Yildiz (2007): "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements," *Econometrica,* 75, 365-400.

37. Weinstein, J. and M. Yildiz (2012): "A Structure Theorem for Rationalizability in Infinite Horizon-Games," *Review of Economic Studies*, forthcoming

38. Yamashita, T. (2012) "A Necessary Condition for Implementation in Undominated Strategies, with Applications to Robustly Optimal Trading Mechanisms," *mimeo*, Toulouse School of Economics

39. Yamashita, T. (2013a) "Strategic and Structural Uncertainties in Mechanism Design", *mimeo*, Toulouse School of Economics.

40. Yamashita, T. (2013b) "Robust Auctions with Interdependent Values", *mimeo*, Toulouse School of Economics.

41. Yamashita, T. (2013c) "Robust trading mechanisms to strategic uncertainty", *mimeo*, Toulouse School of Economics