Full Implementation and Belief Restrictions*

Mariann Ollár  
University of Pennsylvania, Econ. Dept. and Warren Center

Antonio Penta  
University of Wisconsin-Madison, Department of Economics

October 26, 2015

Abstract

We introduce a framework to study the problem of full implementation via the design of simple transfer schemes, under general restrictions on agents’ beliefs. Our construction suggests a simple design principle, in which belief restrictions are used to weaken the strategic externalities of the baseline belief-free transfers, so as to induce mechanisms that ensure uniqueness. Importantly, our results require minimal restrictions on agents’ beliefs, specifically on moments of the distribution of types. These moment conditions arise naturally in applications and in practical problems of mechanism design.

Keywords: Full Implementation, Robust Mechanism Design, $\Delta$-Rationalizability, Interdependent Values, Moment Conditions, Nice Games, Uniqueness, Strategic Externalities

1 Introduction

The problem of multiplicity is a key concern for the design of real-world mechanisms and institutions. Unless all the solutions of a mechanism are consistent with the outcome the designer wishes to implement, the designer may not confidently assume that the proposed mechanism will perform well. This is a well-known criticism of the partial implementation approach to mechanism design, which requires only that there exists one strategy profile consistent with the chosen solution concept that guarantees desirable outcomes. The full implementation approach (cf., Maskin, 1999) overcomes the problem of multiplicity, but in pursuit of greater generality, the existing literature has typically adopted rather complicated mechanisms.\(^1\) Thus, while it addresses an important practical concern, the full implementation literature overall has provided limited insight into how real-world institutions could be designed to avoid the problem of multiplicity.

Another well-known limitation of the classical (Bayesian) approach is its excessive reliance on common knowledge assumptions. This criticism, often referred to as the ‘Wilson doctrine’, has recently received considerable attention in the literature on robust implementation. It is fair to say, however, that the aims of the Wilson doctrine, ‘[...] to conduct useful analyses of practical problems [...]’ (Wilson, 1987), are still far from being fulfilled. This is due to two main limitations.

---

*We are particularly grateful to Laurent Mathevet, Stephen Morris, Ken Hendricks, Larbi Alaoui, George Mailath, Meg Meyer, Andy Postlewaite, Rakesh Vohra and Bill Sandholm for the helpful comments. We also thank seminar audiences at Stanford, NYU, UPenn, Oxford, UW-Madison, Pompeu Fabra, the SAET Conferences in Paris and Tokyo, and the 2015 ESWC in Montreal.

\(^1\)See Jackson (1992) for an influential criticism of the tail-chasing mechanisms typically used in this literature.
On the one hand, most of this literature has focused on environments in which the designer has no information about the agents’ beliefs. This extreme assumption represents a useful benchmark to address foundational questions, but significantly limits the possible applications of the theory to practical problems of mechanism design. On the other hand, as far as full implementation is concerned, the literature has focused on characterization results which offer very little insights on the properties that more realistic mechanisms should satisfy, in order to ensure full implementation. In this paper we address these points pursuing a more pragmatic approach to full implementation, based on mechanisms with a clear economic interpretation (namely, transfer schemes that only depend on agents’ payoff-relevant information) and which rely on more realistic assumptions of common knowledge, intermediate between the classical and the ‘belief-free’ approaches.

For the sake of illustration, consider the problem of efficient implementation. In environments with single-crossing preferences, the generalized VCG transfers of Cremer and McLean (1985) guarantee partial implementation of the efficient allocation in an ex-post equilibrium, with essentially no restrictions on the strength of the preference interdependence. Hence, independent of the agents’ beliefs, truthful revelation (hence efficiency) is always achievable as part of a Bayes-Nash equilibrium (cf. Bergemann and Morris, 2005). The problem with this mechanism is that it typically admits also inefficient equilibria, which can be ruled out if and only if the interdependence in agents’ valuations is not too strong (cf. Bergemann and Morris, 2009). This characterization, however, is often regarded as a negative result, as in many cases preference interdependence is strong, and there is nothing the designer can do about it.

In this paper we shift the focus of the analysis from preference interdependence to the strategic externalities in the mechanism, which - unlike preferences - can be affected by the designer. The problem with the VCG transfers, for instance, is that when agents’ preferences exhibit strong interdependence, the strategic externalities in the mechanism are strong, in that players’ best responses are strongly affected by others’ strategies. This in turn generates multiplicity of equilibria, and hence failure of full implementation. But if the designer has some information about the agents’ beliefs, then preferences and strategic externalities need not be aligned: the strategic externalities can be weakened, so as to ensure uniqueness, even if preference interdependence is strong. Clearly, to ensure that the unique solution implements the designer’s objective, the strategic externalities should be weakened in a way that preserves incentive compatibility – if not in the ex-post sense, then at least for the beliefs consistent with the designer’s information (cf. Mathevet, 2010).  

While efficient implementation is one of our leading examples, our analysis covers general implementation problems with interdependent values, under varying assumptions on agents’ beliefs. For this reason, we adopt the solution concept of Δ-Rationalizability (Battigalli and Siniscalchi, 2003), which extends rationalizability to environments with incomplete information and general assumptions on agents’ beliefs. The resulting notion of implementation provides a unified framework to study full implementation under general belief restrictions, thereby allowing for varying degrees of robustness.  

---

2 On the ‘belief free’ approach to mechanism design, see Bergemann and Morris (2005, 2009a, 2009b, 2011) for static mechanisms and Mueller (2013, 2015) and Penta (2015) for dynamic ones. A thorough account of this literature is provided by Bergemann and Morris (2012). We discuss the related literature in Section 6.

3 The idea of modifying ex-post incentive compatible transfers using information about beliefs is based on a clever insight of Mathevet (2010). Apart from the broad idea, however, there are important differences, which we discuss in Section 6. (See also Mathevet and Taneva (2013) and Healy and Mathevet (2012)).

4 For instance, as the belief restrictions are varied, our notion of implementation includes as special cases both belief-free implementation and Oury and Tercieux’s (2012) ICR-implementation.
we show that Δ-Rationalizability is also convenient from a practical viewpoint, thereby providing a strong argument for its adoption in applied problems. In particular, this solution concept enables a key conceptual innovation of our approach. Namely, the robustness of a mechanism is determined contextually with its design, and as such it can be regarded as being chosen by the designer, in the same way that transfers are. This change in perspective allows us to move beyond the existing characterization results, and to provide constructive insights about what can be done when the conditions for belief-free implementation are not met.

Our general analysis parallels the example above. First we derive the ‘canonical transfers’, a generalization of well-known necessary conditions for ex-post incentive compatible payment schemes. Depending on the environment, and particularly on the strength of the preference interdependence, the canonical transfers may induce overly strong strategic externalities, which are problematic from the viewpoint of full implementation. The second part of our design then exploits the belief restrictions to reduce the strategic externalities, so as to induce a contractive mechanism which guarantees uniqueness. The conditions that guarantee full implementation relate the strength of the preference interdependence to the information embedded in the designer’s assumptions on agents’ beliefs. This information, in particular, takes the form of simple moment conditions, which represent very weak restrictions on agents’ beliefs. Moment conditions arise naturally in applications, and can be easily estimated from previous data on the performance of the mechanism, an important desideratum from a practical viewpoint (see, e.g., Deb and Pai (2013)).

Overall, our results suggest a simple design strategy: start out with the canonical transfers, and then compensate each agent for the strategic externality he faces given everybody’s reports. To deter agents from misreporting their types in order to inflate their compensation, each agent i is also asked to pay a fee equal to the expected value of the compensation given his report:

\[
t_i(\theta) = t_i^*(\theta) + CSE_i(\theta_i, \theta_{-i}) - E(CSE_i|\theta_i)
\]

The first term we add to the canonical transfers reduces the strategic externalities and ensures that the mechanism is contractive; the last term, derived from the designer’s information about agents’ beliefs, restores incentive compatibility. Full implementation follows.

Note that this argument also suggests an interesting tension between the robustness of the partial implementation result (achieved by the VCG mechanism, for instance, in an ex-post equilibrium), and the possibility of achieving full implementation (which, if preference interdependence is strong, necessarily requires information about beliefs). Importantly, however, our implementation results in general rely on weak assumptions about agents’ beliefs, in that they require that only some moments of the type distribution are common knowledge.

We also discuss how our general results can be applied to important special cases. In particular, we consider environments that satisfy standard single-crossing properties and a ‘public concavity’ condition, which generalizes important classes of models in the applied literature. Under these restrictions on preferences, our results can be summarized as follows: (i) in the Bayesian settings commonly considered in the classical and applied literature, full implementation is always possible if types are independent or affiliated, regardless of preference interdependence. Moreover, this can be ensured in dominant strategies; (ii) within these settings, ex-post incentive compatibility is possible
if and only if (interim) dominant-strategy implementation is – a result reminiscent of Manelli and Vincent’s (2010) and Gershkov et al. (2013) BIC-DIC equivalence, but with interdependent values; (iii) in non-Bayesian environments, in which only the conditional averages of types are common knowledge, implementation can always be achieved, provided that the conditional averages of the opponents’ types are constant or increasing in an agent’s own type.

Finally, we show that contractive mechanisms have further desirable properties, such as low sensitivity to misspecifications of agents’ beliefs. This result holds under the most general assumptions of our model, and suggests further notions of robustness as well as a novel concept of approximate implementation.

The rest of the paper is organized as follows: Section 2 introduces the general model, the notions of belief-restrictions and moment conditions, as well as some leading examples. Section 3 introduces the general notion of implementation. Section 4 provides the general results on full implementation via transfers and moment conditions. Section 5 provides the applications and the sensitivity analysis. The related literature is discussed in Section 6. Section 7 concludes.

2 Model

Environments and Mechanisms. We consider standard environments with transferable utility. We denote by \( I = \{1, \ldots, n\} \) the set of agents, by \( X \) the set of social outcomes and by \( t_i \in \mathbb{R} \) the private transfer to agent \( i \in I \). Agents’ preferences depend on the realization of the state of the world \( \theta \in \Theta = \times_{i \in I} \Theta_i \). When \( \theta \) is realized, agent \( i \) privately observes the \( i \)-th component, \( \theta_i \in \Theta_i \). We refer to \( \theta_i \in \Theta_i \) as agent \( i \)’s payoff type (or just as ‘type’), and let \( \theta_{-i} \in \Theta_{-i} = \times_{j \neq i} \Theta_j \) denote the type profile of \( i \)’s opponents. For each \( i \in I \), we let \( v_i : X \times \Theta \to \mathbb{R} \) denote agent \( i \)’s valuation function. We maintain throughout that, for each \( i \), \( \Theta_i = [0, 1] \) and that \( X \) is a convex and compact subset of a Euclidean space. This model accommodates general externalities in consumption, including both pure cases of private and public goods.

The tuple \( \mathcal{E} = (I, (\Theta_i, v_i)) \) defines the (payoffs) environment, and we assume it is common knowledge among the agents. Payoff types thus represent agents’ information about preferences. If \( v_i \) is constant in \( \theta_{-i} \) for every \( i \), then the environment is one of private values. If not, the environment has interdependent values.

A decision rule (or allocation rule) is a mapping \( d : \Theta \to X \), which assigns to each payoff state the social outcome that the designer wishes to implement. We say that an allocation rule is responsive if for any \( i, \theta_i \) and \( \theta_i' \) such that \( \theta_i \neq \theta_i' \), there exists \( \theta_{-i} \in \Theta_{-i} \) such that \( d(\theta_i, \theta_{-i}) \neq d(\theta_i', \theta_{-i}) \). We impose the following assumptions on \( (\mathcal{E}, d) \):

**Assumption 1 (Environment):** (i) for each \( i, v_i \) is three times continuously differentiable, and (ii) \( d : \Theta \to X \) is responsive and twice continuously differentiable.

For simplicity, we will use the standard notation \( \partial f(x)/\partial x \) for all derivatives, with the understanding that when \( X \) is multidimensional, \( \partial v_i(x, \theta)/\partial x \) and \( \partial d(\theta)/\partial \theta \) should be regarded as the vectors of partial derivatives, and \( \partial v_i(x, \theta)/\partial x \cdot \partial d(\theta)/\partial \theta \) denote their inner product.

In general, a mechanism is a tuple \( \mathcal{M} = ((M_i)_{i \in I}, g) \), where \( g: \times_{i \in I} M_i \to X \times \mathbb{R}^n \). For the reasons discussed in the introduction, we restrict ourselves to using direct mechanisms, in which the sets of messages are \( M_i = \Theta_i \), and the common component \( x \in X \) is chosen according to \( d \).
A direct mechanism is thus uniquely determined by a transfer scheme \((t_i)_{i \in I}, t_i : \Theta \rightarrow \mathbb{R}\), which specifies the (possibly negative) transfer to agent \(i\), for each possible profile of reports \(m \in \Theta\). (To distinguish the report from the state, we maintain the notation \(m\) even though \(M = \Theta\).)

Besides ensuring a clear economic interpretation, studying direct mechanisms allows an easier comparison with the literature on partial implementation, by making transparent what features of an incentive compatible mechanism may or may not be problematic for full implementation.

Any direct mechanism \(M\) induces a (belief-free) game \(G^M = \langle I, (\Theta_i, M_i, U_i)_{i \in I} \rangle\), where \(I\) is the set of players, \(\Theta_i\) the set of \(i\)'s payoff types, \(M_i\) is the set of \(i\)'s actions and ex-post payoff functions \(U_i : M \times \Theta \rightarrow \mathbb{R}\) are such that \(U_i(m; \theta) = v_i(d(m), \theta) + t_i(m)\). For every \(\theta_i \in \Theta_i\), \(\mu_i \in \Delta(\Theta_{-i} \times M_{-i})\) and \(m_i \in M_i\), we let \(EU_{\theta_i}^{\mu_i}(m_i)\) denote player \(i\)'s expected payoff from message \(m_i\), if \(i\)'s type is \(\theta_i\) and his conjectures are \(\mu_i\):

\[
EU_{\theta_i}^{\mu_i}(m_i) := \int_{\Theta_{-i} \times M_{-i}} U_i(m_i, m_{-i}; \theta_i, \theta_{-i}) \, d\mu_i.
\]

We also define \(BR_{\theta_i}(\mu_i) := \arg \max_{m_i \in M_i} EU_{\theta_i}^{\mu_i}(m_i)\).

**Belief Restrictions.** We model agents’ beliefs separately from the environment \(E\). This is because, whereas information about beliefs may be useful in designing a mechanism, agents’ beliefs are not directly relevant to the designer’s objectives, \(d\). We model ‘belief restrictions’ as sets of possible beliefs for each type of every player. Formally, \(B = \langle (B_{\theta_i})_{\theta_i \in \Theta_i} \rangle_{i \in I}\) where \(B_{\theta_i} \subseteq \Delta(\Theta_{-i})\) for each \(\theta_i \in \Theta_i\) and \(i \in I\), assumed common knowledge. If \(B\) and \(B'\) are such that \(B_{\theta_i} \subseteq B'_{\theta_i}\) for all \(\theta_i\) and \(i\), we write \(B \subseteq B'\). We maintain the following assumptions throughout:

**Assumption 2 (Beliefs):** For each \(i \in I\) and \(\theta_i \in \Theta_i\), \(B_{\theta_i}\) is non-empty, closed and convex.

This formulation is fairly general. For instance, if \(B_{\theta_i}\) is a singleton for every \(\theta_i\) and \(i\), then the pair \((E, B)\) is a standard Bayesian environment, in which agents’ hierarchies of beliefs are uniquely pinned down by their payoff types. The further special case of a common prior model requires that \(B_{\theta_i} = \{b_{\theta_i}\}\) are such that there exists \(p \in \Delta(\Theta)\) s.t. \(b_{\theta_i} = p(\cdot|\theta_i) \in \Delta(\Theta_{-i})\) for each \(i\) and \(\theta_i\). If, furthermore, \(B_{\theta_i} = B_{\theta'_i}\) for all \(i\) and all \(\theta_i, \theta'_i \in \Theta_i\), then we obtain the case of independent types (cf. Example 1). At the opposite extreme, if \(B_{\theta_i} = \Delta(\Theta_{-i})\) for every \(\theta_i\) and every \(i\), then there are no commonly known restrictions on beliefs, and the pair \((E, B)\) coincides with the belief-free environments that are common in the literature on robust mechanism design (see footnote 2). Such ‘vacuous restrictions’ are thus denoted by \(B^{BF}\). Our model also accommodates settings, intermediate between the Bayesian and belief-free cases, in which some common knowledge restrictions are maintained but not to the point that belief hierarchies are uniquely determined by the payoff types. In those cases, the tuple \(B\) represents the designer’s partial information about agents’ beliefs. Our results apply to all of these cases.

For reasons that will be illustrated in the following, we will distinguish between the belief restrictions in \(B\) and the beliefs with respect to which full implementation may be obtained. From this viewpoint, it is useful to think of \(B\) as the most that the designer is willing to assume about agents’ beliefs. Clearly, if \(B \subseteq B'\), then \(B'\) entails weaker restrictions than \(B\).
2.1 Leading Examples: key insights and their generalizations

In this section we illustrate the main ideas by means of two simple examples.

Example 1 (Full Implementation in a Common Prior Model) Consider an environment with two agents, \(i = 1, 2\). A benevolent social planner chooses some quantity \(x \in X \subseteq \mathbb{R}_+\) of a public good, with cost of production \(c(x) = \frac{1}{2}x^2\). Players’ valuation functions are \(v_i(x, \theta) = (\theta_i + \gamma \theta_j) x\), where \(\gamma \geq 0\) is a parameter of preference interdependence: if \(\gamma = 0\), this is a private-value setting; if \(\gamma > 0\), values are interdependent. Suppose that the planner knows that, for all \(i\)’s, types are i.i.d. draws from a uniform distribution over \(\Theta_i = [0, 1]\), denoted by \(\nu_{\Theta_i}\), and that this is common knowledge among the agents. This is a standard common prior environment, with independently distributed types and interdependent values. In this model, the planner’s information about agents’ beliefs is represented by belief restrictions \(\mathcal{B} = ((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}\) such that \(B_{\theta_i} = \{\tilde{\nu}_{\Theta_j}\}\) for every \(i, j \neq i\) and \(\theta_i \in \Theta_i\).

The objective of the social planner is to implement the efficient level of public good, that is \(d^*(\theta) = (1 + \gamma)(\theta_1 + \theta_2)\). This decision rule can be partially implemented through the VCG mechanism, with transfers

\[
t_i^{VCG}(m) = -(1 + \gamma)\left(\frac{1}{2}m_i^2 + \gamma m_i m_j\right).
\]

Given the VCG mechanism, for any pair \((\theta_j, m_j)\) of player \(j\)’s type and report, the ex-post best-reply function for type \(\theta_i\) of player \(i\) is

\[
BR_i^{VCG}(\theta_j, m_j) = \text{proj}_{[0,1]}(\theta_i + \gamma (\theta_j - m_j)).
\]

Observe that for any \(\gamma \geq 0\) and for any realization of \(\theta_i\), truthful revelation \((m_i(\theta_i) = \theta_i)\) is a best response to the opponent’s truthful strategy \((m_j(\theta_j) = \theta_j)\). This is the well-known ex-post incentive compatibility of the VCG mechanism. Partial implementation of the efficient allocation is thus guaranteed independent of agents’ beliefs. Furthermore, if \(\gamma < 1\), equation (2) is a contraction, and its iteration delivers truthful revelation as the only rationalizable strategy. In this case, the VCG mechanism also guarantees full robust implementation (Bergemann and Morris (2009a)). If \(\gamma \geq 1\), on the other hand, this mechanism fails to robustly implement the efficient allocation rule. (In the general symmetric case with \(n\) agents, it can be shown that no mechanism achieves belief-free full implementation if \(\gamma \geq 1/(n-1)\).)

Hence, with weak interdependence in valuations, the designer need not rely on the common prior: the VCG mechanism ensures full implementation in the belief-free model \(\mathcal{B}^{BF} \supset \mathcal{B}\). If the interdependence is strong, however, full implementation fails, even under the \(\mathcal{B}\)-restrictions. For instance, if \(\gamma = 2\) and types are independently and uniformly distributed, the strategy profile \((\hat{m}_1(\theta_1) = 1, \hat{m}_2(\theta_2) = 0)\) is also a Bayes Nash equilibrium. Furthermore, this equilibrium is inefficient, as it implements \(x = 3\) regardless the state.

Being designed to achieve ex-post incentive compatibility, the VCG mechanism ignores any information about agents’ beliefs. We propose next a different set of transfers, which do exploit

\[5\]For any \(y \in \mathbb{R}\), we let \(\text{proj}_{[0,1]}(y) := \arg\min_{\theta_i \in [0,1]} |\theta_i - y|\) denote the ‘projection’ of \(y\) on the interval \([0,1]\).
some information contained in the common prior (namely, that \( \mathbb{E}(\theta_j|\theta_i) = 0.5 \) for all \( \theta_i \) and \( i \)):

\[
    t^*_i (m) := -(1 + \gamma) \left( \frac{1}{2} m_i^2 + \gamma m_i \mathbb{E}(\theta_j|\theta_i) \right) = -(1 + \gamma) \left( \frac{1}{2} m_i^2 + \gamma m_i \cdot 0.5 \right). 
\]

These transfers induce the following best response function:

\[
    BR^*_i (\hat{m}_j (\cdot)) = \text{proj}_{[0,1]} (\theta_i + \gamma [\mathbb{E}(\theta_j|\theta_i) - 0.5]). 
\]

Since, under the common prior, \( \mathbb{E}(\theta_j|\theta_i) = 0.5 \) for all \( \theta_i \), the term in square brackets cancels out for all types. Truthful revelation therefore is strictly dominant, independent of the strength of preference interdependence, \( \gamma \). In fact, this would be the case for any beliefs that satisfy the moment condition “\( \mathbb{E}(\theta_j|\theta_i) = 0.5 \) for all \( \theta_i \)”.

The precise definition of \( \mathcal{B}^* \) clearly depends on the particular moment condition we used to design the transfers (that is, “\( \mathbb{E}(\theta_j|\theta_i) = 0.5 \) for all \( \theta_i \) and \( \hat{i} \)”). This is only one of infinitely many conditions that are consistent with the designer’s information \( \mathcal{B} \). Had we used a different condition to design the transfers, say “\( \mathbb{E}(L(\theta_j)|\theta_i) = f(\theta_i) \) for all \( \theta_i \) and \( \hat{i} \)”, full implementation may have obtained for different beliefs \( \mathcal{B}' \supset \mathcal{B} \): namely, \( \mathcal{B}' = ((B_{\theta_i}')_{\theta_i \in \Theta_i})_{i \in I} \) such that \( B_{\theta_i}' = \{b_i \in \Delta(\Theta_j) : \int L(\theta_j) \cdot db_i = f(\theta_i)\} \). Thus, it is not only true that the set \( \mathcal{B} \), which represents the designer’s information, need not coincide with the set of beliefs with respect to which implementation is obtained (such as \( \mathcal{B}^{BF} \) or \( \mathcal{B}^* \) in the example), but the latter set is itself determined by the planner’s choice of the mechanism.

In Section 3 we introduce the general notion of implementation, and formalize the sense in which the strength of the strategic externalities, not of the preference interdependence, is key for full implementation. The two go hand in hand in belief-free environments, but need not coincide if the designer has some information about agents’ beliefs. In Section 4 we use these results to develop a general design principle which consists of using properly chosen belief restrictions to weaken the strategic externalities of a baseline ‘canonical’ mechanism. We show that moment conditions, formally introduced in Section 2.2, are particularly suited to this task. In the example above, for instance, the moment condition “\( \mathbb{E}(\theta_j|\theta_i) = 0.5 \) for all \( \theta_i \) and \( \hat{i} \)” enabled us to completely offset the strategic externalities of the VCG mechanism, thereby ensuring full implementation in dominant strategies. In the general case in which strategic externalities cannot be completely eliminated, our design principle pursues contractiveness of the best replies, to ensure that truthful revelation is the unique rationalizable outcome. The next example illustrates the point in the context of a non-Bayesian model.
Example 2 (Full Implementation without a Common Prior) Consider an environment with three agents, \( i = 1, 2, 3 \), and assume that agents commonly know that types \( \theta_i \in [0, 1] \) are i.i.d. draws from some distribution \( \Phi \). The distribution itself, however, is not necessarily known by the agents, and most importantly is unknown to the designer. This environment therefore provides an example both of non-Bayesian belief restrictions and of a situation in which the designer possibly knows less than what is commonly known by the agents.

Preferences are similar to the previous example, except that the factor of preference interdependence may be heterogenous. That is, \( v_i(x, \theta) = (\theta_i + \gamma \theta_j + \delta \theta_k) x \) where \( \gamma, \delta \in \mathbb{R}, x \in \mathbb{R}_+ \) denotes the quantity of public good, and where we let \( j := i + 1 \mod 3 \) and \( k := i + 2 \mod 3 \). If the cost of production is the same as in the previous example, the efficient allocation rule is \( d^*(\theta) = \kappa (\theta_1 + \theta_2 + \theta_3) \) where \( \kappa \equiv (1 + \gamma + \delta) \). The VCG transfers are \( t_i^{VCG}(m) = -\kappa (0.5 m_i^2 + m_i (\gamma m_j + \delta m_k)) \), which induce the following best reply:

\[
BR_{\theta_i}^{VCG} = \text{proj}_{[0,1]}(\theta_i + \mathbb{E}(\gamma (\theta_j - m_j) + \delta (\theta_k - m_k) | \theta_i)).
\]

Now, suppose that \( \gamma = 4/3 \) and \( \delta = -2/3 \). With these parameter values, any report profile is rationalizable, and belief-free implementation fails. The following transfers instead achieve full implementation: \( t_i^*(m) = t_i^{VCG}(m) + \kappa m_i \gamma (m_j - m_k) \). With these transfers, the best reply is:

\[
BR_{\theta_i}^* = \text{proj}_{[0,1]}(\theta_i + \gamma \mathbb{E}(\theta_j - \theta_k | \theta_i) + (\gamma + \delta) \mathbb{E}(m_k - \theta_k | \theta_i))
= \text{proj}_{[0,1]}(\theta_i + (\gamma + \delta) \mathbb{E}(m_k - \theta_k)).
\]

The simplification in the second line is due to the fact that, under the maintained assumptions, \( \mathbb{E}(\theta_j - \theta_k | \theta_i) = 0 \). Unlike in the previous example, strategic externalities are not completely eliminated in this case. However, for the values of parameters specified above, the term \( (\gamma + \delta) = 2/3 < 1 \). Hence, the best-replies induce a contraction, which delivers truthful revelation as the only rationalizable profile. Similar to the previous example, full implementation only relies on common knowledge of the moment condition used in the design of transfers (in this case, \( \gamma \mathbb{E}(\theta_j - \theta_k | \theta_i) = 0 \) for all \( \theta_i \)). Formally, the belief restrictions \( B \) in this model are such that \( B_{\theta_i} = \{ b_i \in \Delta(\Theta_{-i}) : \exists \phi \in \Delta([0,1]) \text{ s.t. } b_i = \sum_{j \neq i} \phi \} \), whereas transfers \( t^* \) achieve full implementation for the weaker restrictions \( B' \subseteq B \), such that \( B'_{\theta_i} = \{ b_i \in \Delta(\Theta_{-i}) : \int (\theta_k - \theta_j) db_i = 0 \} \).

2.2 Moment Conditions

As illustrated by the previous examples, we will exploit an important class of belief restrictions: moment conditions. In this section we introduce the concept formally.

**Definition 1** A \( B \)-consistent moment condition is defined by a collection \( \rho = (L_i, f_i)_{i \in I} \) such that, for every \( i \in I \), \( L_i : \Theta_{-i} \to \mathbb{R} \) and \( f_i : \Theta_i \to \mathbb{R} \) are twice continuously differentiable and such that:

\[
\int_{\Theta_{-i}} L_i(\theta_{-i}) \cdot db_i = f_i(\theta_i) \text{ for all } i, \theta_i \text{ and } b_i \in B_{\theta_i}. \quad (6)
\]

We denote the set of \( B \)-consistent moment conditions by \( \varrho(B) \), and for each \( \rho \in \varrho(B) \) we let
\[ \mathcal{B}^o = \{(B_{\theta_i}^o)_{\theta_i \in \Theta_i})_{i \in I} \text{ be such that, for each } i \text{ and } \theta_i, \]

\[ B_{\theta_i}^o = \left\{ b_i \in \Delta(\Theta_{-i}) : \int_{\Theta_{-i}} L_i(\theta_{-i}) \cdot db_i = f_i(\theta_i) \right\}. \]

In words, a moment condition \( \rho = (L_i, f_i)_{i \in I} \) is consistent with \( \mathcal{B} \) if the latter imply that agents commonly believe that, for every \( i \), his expectation of moment \( L_i(\theta_{-i}) \) of the opponents’ types varies with his type \( \theta_i \) according to \( f_i \). The tuple \( \mathcal{B}^o = ((B_{\theta_i}^o)_{\theta_i \in \Theta_i})_{i \in I} \) in turn denotes the belief restrictions in which only common knowledge of \( \rho \) is maintained. It is easy to see that, for any \( \mathcal{B} \) and \( \rho \in \varrho(\mathcal{B}) \), \( \mathcal{B}^o \supseteq \mathcal{B} \) (that is, \( \mathcal{B}^o \) entails weaker restrictions than \( \mathcal{B} \)).

The next three examples show how moment conditions are implicit in standard models.

**Example 3 (Unobserved Heterogeneity)** Suppose that types \( \theta_i \) are i.i.d. draws from a distribution \( F_\eta \), where \( \eta \) is a parameter drawn from some distribution \( H \). The realization of \( \eta \) is observed by the agents but not by the designer (e.g., Aradillas-Lopez et al., 2013. See also Example 2). This model entails many moment conditions. For instance, irrespective of further details about the involved distributions, it is common knowledge in this model that \( \mathbb{E}(\theta_l - \theta_k | \theta_i) = 0 \) for any \( \theta_i \) and \( i \neq l, k \). This is represented by setting \( L_i(\theta_{-i}) = \theta_l - \theta_k \) for some \( l, k \neq i \) and \( f_i(\theta_i) = 0 \) for any \( \theta_i \). (This is precisely the moment condition used in Example 2.) \( \square \)

**Example 4 (Fundamental Value Models)** Consider a model in which types can be decomposed into a fundamental component and a noise component, i.e. \( \theta_i = \theta_0 + \varepsilon_i \) where \( \theta_0 \) is drawn from a normal distribution and \( \varepsilon_i \)’s are i.i.d. across agents and independent of \( \theta_0 \) (e.g., Rostek and Weretka (2012)). Unlike the previous example, \( \theta_0 \) is not observed by the agents. Yet, the moment condition with \( L_i(\theta_{-i}) = \theta_k - \theta_l \) and \( f_i(\theta_i) = 0 \) holds in this environment. Examples of such information structures include financial models with intrinsic values (e.g., Grossman and Stiglitz (1980) and Hellwig (1980)) and common value auctions. \( \square \)

**Example 5 (Spatial Values)** Consider an environment with two distinct groups of agents (e.g., by geographic location, technology, etc.). Agents inherit the type of their group, drawn independently from a distribution \( F \), with mean \( \mathbb{E} F \), and are assigned to group 1 independently with probability \( p \). An agent’s group and type are his private information (e.g., Ausubel and Baranov (2010)). This information structure admits the moment equation \( \mathbb{E}(\theta_j | \theta_i) = p(i) \theta_i + (1 - p(i)) \mathbb{E} F(v_j) \), where \( p(i) = p \) if \( i \) belongs to group 1, and \( (1 - p) \) otherwise. The corresponding moment condition thus obtains setting \( L_i(\theta_{-i}) = \theta_j \) for some \( j \neq i \) and \( f_i(\theta_i) = p(i) \theta_i + (1 - p(i)) \mathbb{E} F \). \( \square \)

Moment conditions arise naturally in many real-world settings, in which knowledge of some moments of the distribution is a much more basic and realistic kind of information than the one typically assumed by the standard common prior models. Consider the following examples:

**Example 6 (Uncorrelated types without a prior)** Suppose that the designer has data showing no significant correlations across agents. His information, however, does not include the entire distribution of players’ types, but only some moments \( \rho \) of that distribution. In this case, the designer’s information itself consists of moment conditions (that is, \( \mathcal{B} = \mathcal{B}^o \)). For example, if types are uncorrelated, for each \( i, j \) and \( \theta_i \), we have \( \mathbb{E}(\theta_j | \theta_i) = y_j \) for some \( y_j \in \mathbb{R} \). In this case, a moment condition obtains letting \( L_i(\theta_{-i}) = \theta_j \) and \( f_i(\theta_i) = y_j \). \( \square \)
Example 7 (Estimation-based Conditions) Consider a situation in which past data facilitate conditional predictions of agents’ types in the form of linear regressions. Linear regressions are nothing but moment conditions, with \( L_i (\theta_{-i}) = \theta_j \) for \( j \neq i \) and \( \tilde{f}_i (\theta_i) = \hat{c}_i + \hat{a}_i \theta_i \) (where \( \hat{a}_i \) and \( \hat{c}_i \) are the estimated coefficients). Alternatively, past data may only report aggregate statistics of the distributions, so that only conditional expectations of the average of types may be allowed. In this case, a moment condition is obtained letting \( L_i (\theta_{-i}) = \frac{1}{n-1} \sum_{j \neq i} \theta_j \), and so on. \( \square \)

In fact, econometric methods often provide a description of the environment in terms of conditional moments of the distributions, rather than a single ‘common prior’. In these cases, the very belief-restrictions \( \mathcal{B} \) can be specified as the set of all beliefs consistent with such moment conditions, taken as a primitive. Examples 6 and 7 are instances of this kind of situation.

Observe that, in general, any belief restriction entails common knowledge of some moment conditions (that is, \( \varrho (\mathcal{B}) \neq \emptyset \) for any \( \mathcal{B} \)). At a minimum, condition (6) is trivially satisfied for any constant functions \( L_i (\cdot) = \tilde{f}_i (\cdot) = \tilde{y} \). In a belief-free environment, only such trivial moment conditions are commonly known. (Conversely, \( \mathcal{B}' \equiv \mathcal{B}'^{BF} \) whenever \( \rho = (\tilde{L}_i, \tilde{f}_i)_{i \in I} \) consists of such trivial moment conditions). In general, the stronger the belief-restrictions (i.e., the smaller the sets \( \mathcal{B} \)), the richer the set of moment conditions: \( \varrho (\mathcal{B}') \subseteq \varrho (\mathcal{B}) \) if \( \mathcal{B} \subseteq \mathcal{B}' \). Hence, common prior models are ‘maximal’ in the set of moment conditions: if \( \mathcal{B} \) is a common prior model, any collection of functions \( L_i : \Theta_{-i} \rightarrow \mathbb{R} \), satisfies \( \{ L_i, f_i^L \}_{i \in I} \in \varrho (\mathcal{B}) \) for \( f_i^L (\theta_i) := \mathbb{E} L_i (\theta_{-i}) (\theta_i) \), and therefore the designer has maximum freedom to choose a suitable moment condition (cf. Section 5).

3 Implementation

Given the environment \( \mathcal{E} \) and a direct mechanism \( \mathcal{M} \), the ensuing \( G^\mathcal{M} \) is a ‘belief-free’ game, in that it does not contain any information about agents’ beliefs. We introduce belief restrictions via the solution concept, \( \Delta \)-Rationalizability (Battigalli and Siniscalchi, 2003). \( \Delta \)-Rationalizability characterizes the behavioral implications of common certainty of players’ rationality and of a set of exogenous restrictions on players’ beliefs. The latter are referred to as ‘\( \Delta \)-restrictions’, formally defined as a collection \( \Delta = ((\Delta_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I} \) such that \( \Delta_{\theta_i} \subseteq \Delta(\Theta_{-i} \times \mathcal{M}_{-i}) \) for every \( i \) and \( \theta_i \).

\( \Delta \)-Rationalizability consists of an iterated deletion procedure in which, for each type \( \theta_i \), a given report \( m_i \) survives the \( k \)-th round of deletion if and only if it can be justified by conjectures in \( \Delta_{\theta_i} \) that are consistent with the previous rounds of deletion:

**Definition 2 (\( \Delta \)-Rationalizability)** Fix a set of \( \Delta \)-restrictions. For every \( i \in I \), let \( R_i^{\Delta,0} = \Theta_i \times M_i \) and for each \( k = 1, 2, \ldots \), let \( R_i^{\Delta,k-1} = \times_{j \neq i} R_i^{\Delta,k-1} \),

\[
R_i^{\Delta,k} = \{(\theta_i, m_i) : m_i \in B R_{\theta_i} (\mu_i) \text{ for some } \mu_i \in \Delta_{\theta_i} \cap \Delta (R_i^{\Delta,k-1})\},
\]

and \( R_i^{\Delta} = \bigcap_{k \geq 0} R_i^{\Delta,k-1} \).

The set of \( \Delta \)-rationalizable messages for type \( \theta_i \) is defined as \( R_i^{\Delta} (\theta_i) := \{m_i : (\theta_i, m_i) \in R_i^{\Delta}\} \).

Here we consider \( \Delta \)-restrictions which merely capture the idea that, in the game ensuing from the mechanism, the belief restrictions \( \mathcal{B} \) are common knowledge. Formally, for any \( \mathcal{B} = \)}
\[
((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}, \text{ we let } \Delta^B = \left( (\Delta^B_{\theta_i})_{\theta_i \in \Theta_i} \right)_{i \in I} \text{ be such that, for every } i \in I \text{ and every } \theta_i \in \Theta_i, \\
\Delta^B_{\theta_i} := \{ \mu_i \in \Delta(\Theta_{-i} \times M_{-i}) : \text{marg}_{\theta_{-i}}\mu_i \in B_{\theta_i} \}. \tag{7}
\]

**Definition 3 (Full Implementation)** Fix an allocation rule \(d\), a direct mechanism \(\mathcal{M} = (d, t)\) and belief restrictions \(\mathcal{B}\). We say that:

1. \(\mathcal{M}\) implements \(d\) with respect to \(\mathcal{B}\), if \(R^I d \Delta^B (\theta_i) = \{\theta_i\}\) for all \(\theta_i\) and all \(i \in I\).
2. \(\mathcal{M}\) implements \(d\) in (strictly) \(\mathcal{B}\)-dominant strategies, if \(R^I d \Delta^B (\theta_i) = \{\theta_i\}\) for all \(\theta_i\) and all \(i\).

We say that \(d\) is \(\mathcal{B}\)-Implementable (resp., \(\mathcal{B}\)-DS Implementable) if there exists a direct mechanism that \(\mathcal{B}\)-implements \(d\) (resp., implements \(d\) in \(\mathcal{B}\)-dominant strategies).

The basic notion of implementation (part 1) requires that truthful revelation is the only \(\Delta^B\)-Rationalizable strategy.\(^6\) This notion presents several advantages. First, as the belief restrictions are varied, \(\Delta^B\)-Rationalizability coincides with various versions of rationalizability, some of which play an important role in the literature on robustness and on implementation.\(^7\) Second, \(\Delta^B\)-Rationalizability in general is a very weak solution concept. This is important because, opposite to partial implementation, full implementation results are stronger if obtained with respect to a weaker solution concept. Sufficient conditions for full Implementation therefore guarantee full implementation with respect to any refinement of \(\Delta^B\)-Rationalizability. Finally, it can be shown that \(\Delta^B\)-Rationalizability characterizes the set of all Bayes-Nash equilibrium strategies, taking the union over all type spaces that are consistent with the model of beliefs \(\mathcal{B}\). Full \(\mathcal{B}\)-Implementation therefore can be seen as a shortcut to analyze standard questions of Bayesian Implementation for general restrictions on beliefs.

Part 2 of the definition considers a stronger notion of implementation. As shown in Example 1, if truthful implementation is achieved with only one round of \(\Delta^B\)-rationalizability, then truthful revelation is strictly dominant for all the beliefs consistent with \(\mathcal{B}\). In this case, full implementation actually obtains independent of higher order beliefs, so the belief restrictions need not even be common knowledge among the agents. The concept of \(\mathcal{B}\)-dominant implementation therefore entails a strong notion of robustness, and it is stronger than the condition in point 1. For instance, if \(\mathcal{B}\) is a standard common prior model, then \(\mathcal{B}\)-DS Implementation is equivalent to truthful revelation being strictly dominant in the interim normal form of the Bayesian game.

**Remark 1** If \(d\) is \(\mathcal{B}'\)-implementable and \(\mathcal{B}' \supseteq \mathcal{B}\), then \(d\) is also \(\mathcal{B}\)-implementable. It follows that if \(d\) is \(\mathcal{B}'\)-implementable for some \(\rho \in \varrho(\mathcal{B})\), then \(d\) is \(\mathcal{B}\)-implementable.

This remark formalizes the idea, discussed in Section 2.1, that achieving implementation with respect to \(\mathcal{B}\) (the beliefs consistent with the designer's information) is the minimum objective for the designer. The general notion of implementation, however, implicitly accounts for the possibility

---

\(^6\) A weaker notion of implementation would allow \(\Delta^B\)-rationalizable misreports, provided that they all induce the same outcome as the true type profile (i.e., \(m \in R^B(\theta)\) implies \(d(m) = d(\theta)\)). However, it can be shown that the two notions coincide under the maintained assumptions 1 and 2 (see Ollár and Penta, 2014).

\(^7\) In particular, if \(\mathcal{B}\) is a Bayesian model, then \(\Delta^B\)-rationalizability coincides with interim correlated rationalizability (ICR; Dekel, Fudenberg and Morris, 2007). ICR-Implementation has been studied by Oury and Terrieux (2012). If instead the belief restrictions are vacuous (\(\mathcal{B} = B^B\)), then \(\Delta^B\)-rationalizability coincides with ‘belief free’ rationalizability (e.g., Bergemann and Morris, 2009). See Section 6 for further connections.
of achieving full implementation for weaker belief restrictions \( \mathcal{B}' \supseteq \mathcal{B} \), which would ensure a more robust result. In Example 1, for instance, depending on the parameter \( \gamma \), full implementation could be obtained with respect to \( \mathcal{B}^{BF} \) or \( \mathcal{B}^* \), both of which are weaker than the designer’s information in that example.\(^8\) Hence, if \( d \) is \( \mathcal{B}^\rho \)-implementable for some \( \rho \in \varrho (\mathcal{B}) \), \( \mathcal{B} \)-implementation is achieved in a ‘more robust’ sense (that is, relying on weaker common knowledge assumptions).

A close inspection of Definition 3 suffices to see that, in order to achieve \( \mathcal{B} \)-implementation, the truthful profile must be a mutual best response for all types, and for all conjectures allowed by the \( \Delta^\mathcal{B} \)-restrictions. Hence, some notion of incentive compatibility is clearly necessary for full \( \mathcal{B} \)-implementation. For any direct mechanism, and for every \( i \in I \), let \( C_i^T \subseteq \Delta (\Theta_i \times M_{-i}) \) denote the set of truthful conjectures of player \( i \) that is, player \( i \)’s conjectures that assign probability one to his opponents reporting truthfully. Formally:

\[
C_i^T = \{ \mu \in \Delta (\Theta_i \times M_{-i}) : \mu ( \{ (\theta_i, m_{-i}) : m_{-i} = \theta_{-i} \} ) = 1 \}.
\]

**Definition 4** Given belief restrictions \( \mathcal{B} \), a direct mechanism \( \mathcal{M} \) is (strictly) \( \mathcal{B} \)-incentive compatible (B-IC) if for all \( \theta_i \in \Theta_i \) and for all \( \mu \in \Delta^\mathcal{B}_i \cap C_i^T \) and \( \theta'_i \neq \theta_i \), \( EU^\mathcal{B}_i (\theta_i) > EU^\mathcal{B}_i (\theta'_i) \).

It is easy to verify that, if \( \mathcal{B} = \mathcal{B}^{BF} \), then B-IC coincides with (strict) ex-post incentive compatibility (EPIC); if \( \mathcal{B} \) represents a standard type space, then B-IC coincides with standard interim (or Bayesian) incentive compatibility (IIC). Clearly, the weaker the belief restrictions, the stronger the B-IC condition. The following result is straightforward, from Definitions 3 and 4:

**Remark 2** B-IC is a necessary condition for B-Implementation.

As usual, incentive compatibility does not suffice for full implementation. In Ollár and Penta (2014) we provide a full characterization of B-Implementation in terms of B-IC and another condition - ‘B-direct monotonicity’ (B-DM) - which is related to several notions of monotonicity in the implementation literature.\(^9\) While conceptually important, these notions are not particularly suited to providing insights on the design of transfers for full implementation. Thus, rather than discussing the full characterization, we focus here on sufficient conditions which provide a clearer economic intuition and that will inform the design of transfers for full implementation (Section 4). To this end, it is useful to introduce the following definition, which extends the idea of ‘nice games’ – defined by Moulin (1984) for complete information (see Weinstein and Yildiz (2011) for an application) – to the incomplete information games induced by a direct mechanism:

**Definition 5** Fix a set of belief restrictions \( \mathcal{B} = (\mathcal{B}_{\theta_i})_{\theta_i \in \Theta_i, i \in I} \). A mechanism \( \mathcal{M} \) is ‘nice’ with respect to \( \mathcal{B} \) (or \( \mathcal{B} \)-nice) if \( G^\mathcal{M} = \{ I, (\Theta_i, M_i, U_i)_{i \in I} \} \) satisfies the following properties: (i) For each \( i \), \( U_i : M \times \Theta \rightarrow \mathbb{R} \) is twice continuously differentiable; (ii) for every \( i \in I \), \( \theta_i \in \Theta_i \) and \( \mu \in \Delta^\mathcal{B}_i \), the expected payoff function \( EU^\mathcal{B}_{\theta_i} : M_i \rightarrow \mathbb{R} \) is strictly concave. (For convenience, we will use the term ‘nice’ instead of ‘\( \mathcal{B}^{BF} \)-nice’.)

\(^8\)The distinction between the maintained assumptions on beliefs over the environment and the beliefs with respect to which implementation is achieved is not completely new to the literature, though it typically remains implicit. For instance, within the partial implementation literature, ex-post incentive compatibility is often sought even in common prior environments. See, for instance, in Myerson (1981) and Cremer and McLean (1985, 1988).

\(^9\)For instance, if \( \mathcal{B} = \mathcal{B}^{BF} \), then B-DM coincides with Bergemann and Morris’ (2009a) contraction property; if \( \mathcal{B} \) is a Bayesian model, then B-DM is closely related to Oury and Terrieux’s (2012) ICRR-monotonicity, which in turn is related to robust monotonicity (Bergemann and Morris, 2011) and to Bayesian monotonicity (Jackson (1991) and Postlewaite and Schmeidler (1986)). See Ollár and Penta (2014) for details.
Theorem 1 Let mechanism $\mathcal{M}$ be $\mathcal{B}$-incentive compatible and nice with respect to $\mathcal{B}$. Then, $\mathcal{M}$ achieves full $\mathcal{B}$-implementation if the following condition holds:

$\mathcal{B}$-Limited Strategic Externalities ($\mathcal{B}$-LSE) for each agent $i$, for all $\theta_i$, for all $m \in \Delta_{\theta_i}^\mathcal{B}$, and for all $m_i, m'_i \in M_i$

\[
\left| \int_{\Theta_i \times M_{-i}} \frac{\partial^2 U_i}{\partial^2 m_i} (m'_i, m_{-i}, \theta_i, \theta_{-i}) \, d\mu \right| > \int_{\Theta_i \times M_{-i}} \sum_{j \neq i} \left| \frac{\partial^2 U_i}{\partial m_i \partial m_j} (m_i, m_{-i}, \theta_i, \theta_{-i}) \right| \, d\mu. \tag{8}
\]

Proof. (See Appendix A.)

To understand this result, consider the first-order condition of type $\theta_i$’s optimization problem, given conjectures $\mu \in \Delta_{\theta_i}^\mathcal{B}$: $\int_{\Theta_i \times M_{-i}} \frac{\partial U_i}{\partial m_i} (m^*_i, m_{-i}, \theta_i, \theta_{-i}) \, d\mu = 0$. Because of the strict concavity assumption (Def. 5), in $\mathcal{B}$-nice mechanisms this condition is both necessary and sufficient for $m^*_i \in \text{int} (\Theta_i)$ to be a best response to $\mu \in \Delta_{\theta_i}^\mathcal{B}$. The second derivative $\frac{\partial^2 U_i}{\partial m_i \partial m_j} (m_i, m_{-i}, \theta_i, \theta_{-i})$ therefore captures how $j$’s report affects $i$’s best response, hence $j$’s ‘strategic externality’ on $i$. Condition (8) requires the ‘own effect’ to be stronger than the opponents’ effects, considered jointly. This condition therefore captures the idea that strategic externalities should not be too large.

Theorem 1 generalizes a result from Moulin (1984), which obtains for the special case in which $\Theta_i$ is a singleton for every $i$, so that the game has complete information. The proof for the general case, however, requires a different argument. This is partly due to the infinite-dimensional strategy spaces, but also to the robustness requirement implicit in the belief restrictions. Unlike Moulin’s complete information case, the LSE-conditions alone do not suffice for the result. A case in point is provided by Section 5.2, in which we show that mechanisms that satisfy the LSE-condition but fail $\mathcal{B}$-incentive compatibility may have multiple rationalizable outcomes. (See Ollár and Penta (2015) for a full game theoretic generalization, beyond mechanism design.)

4 Designing Transfers for Full Implementation

Theorem 1 suggests that nice mechanisms can be used to guarantee full implementation, provided that they are incentive compatible and that the strategic externalities are adequately bounded. In the following we exploit this insight to explicitly construct transfers to achieve full implementation.

We begin by considering belief-free implementation. This is the most demanding notion of Implementation, and in many cases it is not possible. When possible, however, it is convenient to adopt a mechanism that achieves it, because it entails full robustness of the result. In Section 4.1 we introduce the canonical transfers, and show that they characterize the mechanisms that achieve belief-free implementation. Hence, if the canonical transfers induce overly strong strategic externalities, belief-free implementation is impossible. Full implementation may still be possible if information about beliefs is used. In Section 4.2 we obtain transfers for full implementation adding a belief-based term to the canonical transfers. The extra term is based on moment conditions, properly chosen in order to weaken the strategic externalities and deliver a ‘nice’ mechanism. Full implementation then follows from Theorem 1.

\[\text{With complete information, condition (8) coincides with Moulin’s (1984) ‘diagonal dominance’, which besides dominance-solvability also implies the only Nash Equilibrium is stable (Gabay and Moulin, 1980). Equilibrium stability is related to dominance solvability, and is stronger than equilibrium uniqueness, which is implied for instance by the weaker Gale and Nikaido (1965) conditions, that require the Hessians of the payoff functions to be P-matrices.}\]
4.1 Canonical Transfers and Belief-Free Implementation

Consider the following transfers: for each \( i \in I \) and \( \theta \in \Theta \), let

\[
t^*_i(\theta) = -v_i(d(\theta), \theta) + \int_{\theta_i}^{d_i} \frac{\partial v_i}{\partial \theta_i}(d(s_i, \theta_{-i}), s_i, \theta_{-i}) \cdot ds_i.
\]  

(9)

We will refer to \( t^* = (t^*_i(\cdot))_{i \in I} \) as the canonical transfers, and to \( M = (d, t^*) \) as the canonical mechanism. In the canonical mechanism, agents pay their valuation as entailed by the reports profile (treated as truthful) minus the ‘total own preference effect’. This way, agents’ payments coincide with the ‘total allocation effect’ of their report, given the opponents’.\(^{11}\)

Canonical transfers generalize several known mechanisms, such as the VCG mechanism if \( d \) is the efficient allocation rule, Myerson (1981), Laffont and Maskin (1980) and Mookherjee and Reichelstein’s (1992) mechanisms in private value settings, and Li (2013) and Roughgarden and Talgam-Cohen’s (2013) with interdependent values. Proposition 1 below shows that the canonical transfers characterize the direct mechanisms that achieve belief-free full implementation. This result follows immediately from the following Lemma, which generalizes analogous results for the above mentioned special cases. We report it here because it has intrinsic interest from the viewpoint of partial implementation (proofs are in Appendix B):

**Lemma 1** Suppose that \( M = (d, t) \) is EPIC and \( t \) is differentiable. Then, for every \( i \) and for every \( m \), there exists a function \( \tau_i : \Theta_{-i} \to \mathbb{R} \) such that \( t_i(m) = t^*_i(m) + \tau_i(m_{-i}) \).

**Proposition 1** Allocation rule \( d \) is belief-free implementable by a differentiable direct mechanism if and only if the canonical mechanism is belief-free (fully) implementable.

In many environments of economic interest the canonical mechanism induces a nice game (cf. Section 5). Hence, if in such environments ex-post incentive compatibility is possible, full implementation can only fail if the canonical mechanism induces overly strong strategic externalities. We provide next a measure of such strategic externalities. For any \( i \in I \), let \( W_i : \Theta \times \Theta \to \mathbb{R} \) be such that for any \( (m, \theta) \in \Theta \times \Theta \):

\[
W_i(m, \theta) := \left( \frac{\partial v_i(d(m), \theta)}{\partial d} - \frac{\partial v_i(d(m), m)}{\partial d} \right) \frac{\partial d(m)}{\partial \theta_i}.
\]

For every \( i \in I \), define the ‘contractivity gap’ as:

\[
CG_i := \max_{\theta, m, m'} \left( \sum_{j \neq i} \left| \frac{\partial W_i(m, \theta)}{\partial m_j} \right| - \left| \frac{\partial W_i(m', m_{-i}, \theta)}{\partial m_i} \right| \right).
\]  

(10)

\(^{11}\)Let \( \varpi_i(\theta) \equiv v_i(d(\theta), \theta) \) and consider its derivative with respect to \( \theta_i \):

\[
\frac{\partial \varpi_i(\theta)}{\partial \theta_i} = \frac{\partial v_i(d(\theta), \theta)}{\partial \theta_i} + \frac{\partial v_i(d(\theta), \theta)}{\partial x} \cdot \frac{\partial d(\theta)}{\partial \theta_i}.
\]

The first term represents the ‘own preference effect’: the variation of \( i \)'s valuation due to \( \theta_i \), holding \( d(\theta) \) constant. The second term is the ‘allocation effect’: the variation of \( i \)'s valuation at \( \theta \), when the allocation changes due to a change in the reported \( \theta_i \). Integrating both terms with respect to \( \theta_i \), we obtain that \( \varpi_i \) can be decomposed as

\[
\varpi_i(\theta) = \int_0^{d_i} \frac{\partial v_i}{\partial \theta_i}(d(s_i, \theta_{-i}), s_i, \theta_{-i}) \cdot ds_i + \int_0^{d_i} \frac{\partial v_i}{\partial x}(d(s_i, \theta_{-i}), s_i, \theta_{-i}) \cdot \frac{\partial d(\theta)}{\partial \theta_i} \cdot ds_i
\]

\[
\cdot ds_i,
\]

where the first term is the ‘total preference effect’ and the second is the ‘total allocation effect’. Rearranging terms, the canonical transfers can be seen as the negative of the total allocation effect, given the opponents’.
Suppose that the canonical mechanism is nice. Then: if the allocation rule is EPIC but not belief-free fully implementable, then $CG_i > 0$ for some $i$.

To understand this result, note that $W_i(m,\theta)$ is nothing but the derivative of the ex-post payoff function of the canonical mechanism with respect to $i$’s type, evaluated at state $\theta$, when the reported profile is $m$. The ‘contractivity gap’ therefore measures the maximal difference between the opponents’ ability to jointly affect this derivative and player $i$’s own effect, evaluated across all possible combinations of states and reports. Hence, $CG_i < 0$ means that $i$’s own effect on the first-order condition of the canonical mechanism always dominates the combined strategic externalities at all states and reports. Then result then follows from Theorem 1.

### 4.2 Full Implementation via Moment Conditions

By the results in Section 4.1, if the canonical mechanism is nice and ex-post incentive compatible, failure to achieve belief-free implementation is due to the existence of positive contractivity gaps. In these cases, information about beliefs may be useful to weaken the strategic externalities and achieve full implementation. In general, also ex-post incentive compatibility may be problematic. In that case, belief restrictions can be used to ensure both properties.

The next theorem relates the possibility of achieving full implementation to the moment conditions consistent with $\mathcal{B}$. As discussed in Section 2, the choice of the moment condition affects both the design and the degree of robustness achieved by the mechanism. Envisioning robustness as a choice of the designer, formalized by this result, is an important conceptual innovation:

**Theorem 2** Allocation rule $d : \Theta \to X$ is fully $\mathcal{B}$-Implementable if there exists a moment condition $\rho = (L_i, f_i)_{i \in I} \in \varrho(\mathcal{B})$ such that, for all $i$, $\theta_i, m_i, m'_i$ and for all $\mu \in \Delta_{\rho_i}$:

1. $\int_{\Theta_i \times M_i} \frac{\partial W_i(m_i, m_{-i}, \theta_i, \theta_{-i})}{\partial m_i} d\mu < f'_i(m_i)$, and
2. $\int_{\Theta_i \times M_i} \left[ \frac{\partial W_i(m'_i, m_{-i}, \theta_i, \theta_{-i})}{\partial m_i} - f'_i(m'_i) \right] d\mu > \sum_{j \neq i} \int_{\Theta_j \times M_j} \left[ \frac{\partial W_i(m_i, m_{-i}, \theta_i, \theta_{-i})}{\partial m_j} + \frac{\partial L_i(m_{-i})}{\partial m_j} \right] d\mu.$

Moreover, for $\rho = (L_i, f_i)_{i \in I} \in \varrho(\mathcal{B})$ that satisfies the two conditions, the following transfers guarantee Full $\mathcal{B}_0$-Implementation (hence $\mathcal{B}$-Implementation):

$$t^\rho_i(m) = \underbrace{t^*_i(m)}_{\text{canonical transfers}} + \underbrace{L_i(m_{-i}) m_i - \int_{0}^{m_i} f_i(s_i) ds_i}_{\text{moment condition-based term}}.$$ (11)

**Proof.** (See Appendix B.) ■

We also provide a stronger, ‘ex-post’ version of these conditions, which is often easier to check in applications:

**Remark 3** The conditions of Theorem 2 are satisfied if for all $i$, for all $\theta \in \Theta$, for all $m_{-i} \in M$ and for all $m_i, m'_i \in M_i$:

1. $\frac{\partial W_i(m, \theta)}{\partial m_i} < f'_i(m_i)$
2. $\left[ \frac{\partial W_i(m'_i, m_{-i}, \theta)}{\partial m_i} - f'_i(m_i) \right] > \sum_{j \neq i} \left[ \frac{\partial W_i(m_i, m_{-i}, \theta)}{\partial m_j} + \frac{\partial L_i(m_{-i})}{\partial m_j} \right].$
Theorem 2 states two properties of moment conditions that are useful to guarantee full implementation, and may thus guide the designer’s choice of a suitable moment condition. To understand what these are, let us consider the stronger versions stated in Remark 3. First, note that if the contractivity gap (10) is negative for all \( i \), then Condition 2 is satisfied by any trivial moment condition, in which \( f_i \) and \( L_i \) are constant functions. Since such trivial moment conditions are consistent with any belief restrictions, full implementation is guaranteed by the canonical mechanism in the belief-free sense. Now, suppose that the contractivity gap is positive for some agent. Condition 2 clarifies which properties of beliefs can be used to weaken the strategic externalities: a moment condition in which the derivative of \( f_i \) has the opposite sign of \( \frac{\partial W_i}{\partial m_i} \) can be used to increase the ‘own effect’, whereas the ‘external effects’ can be weakened by moment functions \( L_i \) with derivatives that contrast the strategic externality in the canonical mechanism. Condition 1 instead requires that the ‘own effect’ in the canonical mechanism is bounded above by the derivative of the \( f_i \) function.

To gain further insights on how these conditions contribute to the full implementation result, it is useful to consider the transfers that guarantee full implementation (eq. 11). With these transfers, the first-order derivative of \( \theta_i \)’s expected payoff, given \( \mu \in \Delta (\Theta_{-i} \times M_{-i}) \), is:

\[
\frac{\partial EU_{\theta_i}^B}{\partial m_i} (m_i) = \int_{\Theta_{-i} \times M_{-i}} \left( \frac{\partial v_i (d(m), \theta)}{\partial d} - \frac{\partial v_i (d(m), m)}{\partial d} \right) \frac{\partial d(m)}{\partial m_i} + L_i (m_{-i}) - f_i (m_i) \, d\mu.
\]

This shows that for any truth-telling conjecture \( \mu \in \Delta_0 \cap C_i^T \), the report \( m_i = \theta_i \) satisfies the first-order conditions. This does not necessarily result in \( B^o \)-IC, as that depends on the second-order conditions as well. Condition 1 in Theorem 2, however, guarantees that the ensuing mechanism is nice, and hence the second order conditions are satisfied. This mechanism therefore is \( B^o \)-IC. Full implementation follows from the fact that Condition 2 in Theorem 2 also guarantees the \( B^o \)-LSE condition of Theorem 1.

Theorem 2 is constructive in the sense that it pins down a precise design principle: the designer shall start out with the canonical transfers, and then add a new term which is based on suitable moment conditions. ‘Suitable’ here means that the term added to the canonical transfers ought to guarantee niceness of the mechanism and reduce the strategic externalities. In the next section we illustrate the versatility as well as further robustness properties of this design strategy.

5 Applications and Extensions

In this Section we illustrate how the general results of Theorem 2 can be applied to special cases of economic interest, under different assumptions on agents’ beliefs.

5.1 SCC-Environments: A Robustness Trade-off

For simplicity, in this subsection we maintain that \( X \subseteq \mathbb{R} \).\(^{12}\) A common assumption in applications is provided by the following single-crossing condition (SCC):

**Assumption 3 (SCC):** for every \( i \in I \), and \( (x, \theta) \), \( \partial^2 v_i (x, \theta) / \partial x \partial \theta_i \geq 0. \)

\(^{12}\)The extension to a multidimensional outcome space is straightforward, but notationally cumbersome.
The next lemma generalizes standard results on ex-post (partial) implementation:

**Lemma 2** Under the maintained assumptions 1-3, the canonical mechanism is EPIC if and only if the allocation rule is strictly increasing: \( \partial d(\theta)/\partial \theta_i > 0 \) for every \( \theta \) and every \( i \).

**Proof.** (See Appendix C). ■

Because of this result, in the following we refer to SCC-Environments as those that satisfy Assumptions 1-3 and \( \partial d(\theta)/\partial \theta_i > 0 \) for every \( \theta \) and every \( i \).

The next result, which follows immediately from Lemma 2 and Theorem 2, summarizes easy-to-check conditions for belief-free full implementation in SCC-environments:

**Proposition 2** In SCC-environments, the canonical mechanism ensures belief-free implementation whenever \( \frac{\partial W_i}{\partial \theta_i} (m, \theta) < 0 \) and \( \frac{\partial W_i}{\partial \theta_i} (m, \theta) > \sum_{j \neq i} \frac{\partial W_i}{\partial \theta_j} (m', \theta) \) for all \( i, \theta, m \) and \( m' \).

This result highlights an important feature of SCC-environments. Namely, failure to achieve belief-free full implementation is possible only if the canonical mechanism is not nice or if there are positive contractivity gaps. These environments therefore present an interesting trade-off between the robustness of the partial implementation result, obtained by the canonical mechanism in a belief-free sense, and the possibility of achieving full implementation: the latter necessarily relies on belief restrictions and therefore reduces the robustness of the partial implementation result.

To simplify the analysis, we first consider the following assumption:

**Assumption 4:** (i) For each \( i, j \), the second-order derivatives \( \frac{\partial^2 v_i}{\partial \theta_j^2} \) and \( \frac{\partial^2 v_i}{\partial \theta_j^2} \) are constant in \( \theta \); (ii) the allocation rule is linear in \( \theta : \frac{\partial^2 a(\theta)}{\partial \theta_i \partial \theta_j} = 0 \) for all \( i, j \in I \) and \( \theta \in \Theta \).

A special case of these conditions is provided by environments with quadratic valuations and linear allocation functions. Assumption 4 also accommodates more general dependence on \( x \), as long as the concavity and the cross derivatives are public information in the environment. We thus refer to the SCC-environments which also satisfy this assumption as SCC-environments with ‘public concavity’ (SCC-PC). Assumption 4, however, is not essential to our analysis. In Section 5.1.3 we discuss how the results are affected when this assumption is relaxed.

**Remark 4** In SCC-PC environments, for any \( i, j \in I \) and \( \theta, m \in \Theta \):

\[
\frac{\partial W_i}{\partial \theta_j} (m, \theta) = - \left( \frac{\partial^2 v_i}{\partial \theta_j} (d(m), m) \right) \frac{\partial d(m)}{\partial \theta_j}.
\]

For \( j = i \), this implies \( \frac{\partial W_i}{\partial \theta_i} (m, \theta) / \partial \theta_i < 0 \), hence the canonical mechanism is nice (Def. 5).

---

13 This result is closely related to Bergemann and Morris (2009a, BM), who characterized belief-free implementation via direct mechanisms. The difference between the results is mainly in focus, as we pursue an explicit design of transfers and more intuitive sufficient conditions. For ease of reference, we recall here BM’s characterization: under the assumption that valuation functions can be written as \( v_i (x, \theta) = w_i (x, h_i (\theta)) \) for some \( w_i : X \times R \rightarrow R \) and \( h_i : R^n \rightarrow R \), belief-free implementation is characterized by strict EPIC and the following property (Def.5, p.1183, ibid.): for each \( \beta \in \Theta \rightarrow 2^\Theta \) s.t. \( \beta(\theta') \neq \beta(\theta) \) for some \( \theta' \), there exists \( i, \theta_i' \) and \( \theta_i'' \in \beta_i (\theta_i) \) with \( \theta_i'' \neq \theta_i' \) such that, for all \( \theta_{-i} \) and \( \theta_{-i}' \in \beta_{-i} (\theta_{-i}) \), \( \text{sign} (\theta_{-i} - \theta_{-i}') = \text{sign} \left( h_i (\theta_{-i}) - h_i (\theta_{-i}') \right) \).

14 While rather special in the context of our paper, such quadratic-linear models are extremely common in both applied and theoretical literature, to ensure linear best-responses. This is common, for instance, in social interactions models (see, e.g., Blume et al., (2015) and references therein), finance and demand-function competition (e.g., Vives (2011), Rostek and Weretka (2012), Bergemann et al. (2015)), divisible good auctions (e.g., Wilson, 1979), public goods (e.g., Duggan and Roberts, 2002), networks (e.g., Fainmesser et al., 2015), etc.
5.1.1 Common Prior Models

Independent Types. In an independent common prior model, for any $L_i: \Theta_{-i} \rightarrow \mathbb{R}$, the condition $E(L_i(\theta_{-i})) | \theta_i = f_i(\theta_i)$ holds true with $f_i: \Theta_i \rightarrow \mathbb{R}$ s.t. $f_i^0 = 0$ (just by the definition of independence). Hence, since $L_i$ can be chosen freely in common prior models, independence leaves us huge leeway in manipulating the external effects on the RHS of Condition 2 of Theorem 2, without affecting the LHS. The ex-post condition of Remark 3 therefore can be rewritten as:

$$\left| \frac{\partial^2 v_i}{\partial x \partial \theta_i} (d(m), m) \right| > \sum_{j \neq i} \left| \frac{\partial^2 v_i}{\partial x \partial \theta_j} (d(m), m) \right| \frac{\partial d(m)}{\partial \theta_i} + \frac{\partial L_i (m_{-i})}{\partial m_j}$$

(Equations (13) and (14) are well-defined because Assumption 4 ensures that the RHS of (13) is constant in $m_i$). Hence, the following Proposition holds:

**Proposition 3** Full implementation is always possible in SCC-PC environments with independent common prior. In particular, let $B$ be an independent common prior model. For any $i \in I$, let $\hat{L}_i: \Theta_{-i} \rightarrow \mathbb{R}$ be defined as in (14) and let $\hat{f}_i(\theta_i) := E \left( \hat{L}_i (\theta_{-i}) | \theta_i \right)$. Then the transfers

$$t_i^0 (m) = t_i^c (m) + \hat{L}_i (m_{-i}) m_i - \int^{m_i} \hat{f}_i (s_i) ds_i$$

ensure $B^\rho$-DS Implementation. Since $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I} \in \varrho(B)$, full B-Implementation follows.

To understand the logic of the mechanism, first notice that the function $\hat{L}_i (m_{-i})$ constructed above is nothing but a measure of the strategic externality that other players impose on $i$. The transfers in (15) therefore are such that, starting from the canonical mechanism, player $i$ is compensated for the total strategic externality he is subject to. The last term in (15) is nothing but the expected value of such compensation, when $i$ reports $m_i$. This term is added to prevent the agent from misreporting his type, in order to inflate the implied compensation for the strategic externality. Hence, the first term eliminates the strategic externalities, and the second restores incentive compatibility. Full Implementation follows, in B-Dominant Strategies.

Affiliated Types. Under the maintained assumptions for SCC-PC environments, and if valuations are supermodular (that is, for all $i, j$ and $x$ and $\theta$, $0 < \partial^2 v_i (x, \theta) / \partial x \partial \theta_j < \infty$), the moment

\[ \text{Consider eq. (14): The term in parenthesis represents the effect of } j \text{'s report on } i \text{'s marginal utility for } x \text{, and is multiplied by the impact of } i \text{'s report on the allocation. Overall, this is the total strategic externality that player } i \text{ is subject to, for each increment of his own report.} \]
functions $\hat{L}_i : \Theta \rightarrow \mathbb{R}$ defined in (14) are such that $\frac{\partial \hat{L}_i(m)}{\partial m_j} > 0$ for all $m$ and $j \neq i$. Then, Theorem 5 in Milgrom and Weber (1982) implies that if types are affiliated (ibid., p.1098) then $\mathbb{E}(\hat{L}_i(\theta_{-i}|\theta_i)$ is an increasing function of $\theta_i$. Hence, letting $\hat{f}_i(\cdot) \equiv \mathbb{E}(\hat{L}_i(\theta_{-i}|\cdot))$, the moment condition $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I}$ is defined as in (14) and s.t. $\hat{f}_i(\theta_i) = \mathbb{E}(\hat{L}_i(\theta_{-i}|\theta_i)$ for each $\theta_i$. Since $\hat{f}_i > 0$, SCC (Assumption 3) implies that the LHS of Condition 2 in Theorem 2 is equal to zero. The next result follows:

**Proposition 4** Full implementation is always possible in SCC-PC environments with supermodular valuations and affiliated types. In particular, the transfers in (15) ensure $B$-DS Implementation, where $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I}$ is defined as in (14) and s.t. $\hat{f}_i(\theta_i) = \mathbb{E}(\hat{L}_i(\theta_{-i}|\theta_i)$ for each $\theta_i$. Since $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I} \in \mathcal{G}(B)$, $B$-DS Implementation follows.

**Equivalence of EPIC and DS-Implementation.** The construction above can also be used to derive an interesting equivalence between ex-post and (interim) dominant strategy incentive compatibility (iDSIC):

**Proposition 5** Under assumptions 1-4, in independent common prior environments, an allocation function is ex-post incentive compatible if and only if it is interim dominant strategy implementable. If valuations are supermodular, the EPIC-iDSIC equivalence extends to affiliated types.

The proof is simple (see Appendix C). First, we show that an allocation rule is iDSIC only if it is increasing. The ‘only if’ part of the proposition then follows immediately from Lemma 2. The ‘if’ direction follows from the discussion above: if the allocation rule is EPIC, Lemma 2 implies that it is increasing. Propositions 3 and 4 in turn imply that the allocation rule is iDSIC.

This result is somewhat related to important results by Manelli and Vincent (2010, MV) and Gershkov et al. (2013) which show that, in Bayesian environments with private values, for any interim-IC mechanism there is an ‘equivalent’ mechanism that is DSIC. Given the restriction to private values, one way of interpreting this result is as an equivalence between ‘partial’ and ‘full’ implementation in direct mechanisms. From this viewpoint, Proposition 5 can be seen as a generalization of that insight to Bayesian environments with interdependent values. We should point out, however, that MV’s notion of equivalence is different from ours. In particular, MV define two mechanisms as ‘equivalent’ if they deliver the same interim expected utilities for all agents and the same ex-ante expected social surplus. Here instead we maintain the traditional notion of equivalence, which requires that the mechanisms induce the same ex-post allocation. (As shown by Gershkov et al. (2013), equivalence results à la MV do not extend beyond environments with linear utilities and independent types.)

### 5.1.2 Moment Conditions without a Prior

In real-world problems of mechanism design, the designer’s information typically does not take the form of a common prior distribution on agents’ types. For instance, when the designer’s information is based on econometric estimates, the belief restrictions $B$ are naturally represented directly in terms of a set of moment conditions (cf. Section 2.2). In this section we show how Theorem 2 can be used in these non-Bayesian settings as well.

---

16We are grateful to Stephen Morris for this insight.
For instance, suppose that the only information available to the designer concerns the conditional averages $E(\theta_j|\theta_i)$. For each $i, j$, let $\varphi_{ji} : \Theta_i \rightarrow \Theta_j$ be such that, for each $\theta_i \in \Theta_i$, $\varphi_{ji}(\theta_i) := E(\theta_j|\theta_i)$. For simplicity, assume that these functions $\varphi_{ji}$ are differentiable. Then, the designer’s information is represented by belief restrictions $B_{avc} = \{(B^a_{avc})_{\theta_i \in \Theta_i} | i \in I \}$ such that $B^a_{avc} = \{\beta \in \Delta(\Theta_i) : \beta(\theta_j) = \varphi_{ji}(\theta_i)\}$, for each $i \in I$ and $\theta_i \in \Theta_i$.

Next notice that, in SCC-PC environments, the function $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$ defined in (14) is linear. Hence, if the conditional averages $E(\theta_j|\theta_i)$ are common knowledge in $B_{avc}$, so are the conditional expectations $E(\hat{L}_i(\theta_{-i})|\theta_i)$, which may thus be used as ‘moment conditions’ to weaken the strategic externalities. Formally, let $\hat{f}_i(\theta_i) := \hat{L}_i((\varphi_{ji}(\theta_i))_{j \in I \setminus \{i\}})$. Then, because of the linearity of the $E(\cdot)$ operator and of $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$, we have that $E(\hat{L}_i(\theta_{-i})|\theta_i) = \hat{f}_i(\theta_i)$ for all $i$, that is $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I} \in \varrho(B_{avc})$. Moreover, $\hat{f}_i$ is non-decreasing if so are the functions $\varphi_{ji}$. Then, the next result follows from Theorem 2 for the same reasons as Proposition 4:

**Proposition 6** Let $B_{avc}$ be such that, for each $i, j$, the functions $\varphi_{ji}$ are non-decreasing. Then, in PC-SCC environment with supermodular valuations, the mechanism defined in (15) with $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I}$ and $\hat{L}_i$ defined as in (14), ensures $B^\rho$-DS Implementation. Since $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I} \in \varrho(B_{avc})$, $B_{avc}$-DS Implementation follows.

### 5.1.3 Discussion

The logic of Propositions 3, 4 and 5 extends well beyond the cases of common prior models with independent or affiliated types. To see this, notice that for $\hat{L}_i : \Theta_{-i} \rightarrow \mathbb{R}$ defined in eq. (14), the maintained assumptions for SCC-PC environments guarantee that the RHS of Condition 2 in Theorem 2 is equal to zero. Affiliation or independence further guarantee that the conditional moment $E(\hat{L}_i(\theta_{-i})|\theta_i)$ is (weakly) increasing in $\theta_i$, hence the moment condition $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I}$ can be used with no risk of upsetting the RHS of Condition 2 in Theorem 2. This argument, however, remains valid whenever $E(\hat{L}_i(\theta_{-i})|m_i) < \frac{\partial W_i(m, \theta)}{\partial m_i}$ for all $m$, which ensures that both conditions of Theorem 2 are still satisfied by $\rho = (\hat{L}_i, \hat{f}_i)_{i \in I}$. In Proposition 6, the assumption that functions $\varphi_{ji}$ are non decreasing plays the same role as the assumptions of independence and affiliation in the common prior models, and can be weakened similarly.

Assumption 4 may also be weakened in Propositions 3, 4 and 6. In the argument above, we used that assumption to ensure that $\partial W_i/\partial m_i < 0$ and that $\hat{L}_i$ could be designed to completely neutralize the strategic externalities of the canonical mechanism. Clearly, $\partial W_i/\partial m_i < 0$ can be guaranteed by weaker conditions. If Assumption 4 is violated, however, then we may not be able to choose $\hat{L}_i$ to completely offset the strategic externalities. But if $|\partial^2d/\partial \theta_i \partial \theta_j|$ and the variations of the valuations’ concavity are small relative to $|\partial W_i/\partial m_i|$, then $\hat{L}_i$ can still be chosen so that the RHS of (12) is bounded above by $|\partial W_i/\partial m_i|$, and the argument goes through essentially unchanged. The only difference is that in this case full implementation would not occur in only one round of $\Delta^B$-Rationalizability: that is, we would have $B^\rho$-Implementation, but not in dominant strategies.

### 5.2 Sensitivity Analysis and Approximate Moment Conditions

The results in Theorem 2 ensure ‘robust’ implementation in the sense that only common knowledge of a certain moment condition is required (namely, the one used to weaken the strategic externalities). But what if such a moment condition is not exactly satisfied? How sensitive are
the implementation results to possible misspecifications of the moment condition? In this section we show that our design strategy also ensures that the mechanism is well-behaved with respect to such possible misspecifications.

**Example 8 (Sensitivity Analysis)** Consider the environment in Example 2. Adopting the design strategy of Theorem 2, we showed that the strategic externalities of the VCG mechanism could be sufficiently reduced, so as to induce contractive best responses, adopting the moment condition $\gamma \mathbb{E}(\theta_k - \theta_j | \theta_i) = 0$. But what if $\gamma \mathbb{E}(\theta_k - \theta_j | \theta_i)$ is only within $\varepsilon$ of 0, so that the moment condition is not exactly satisfied? Then, for any $\theta_i$, the set of rationalizable reports consistent with common belief that $\gamma \mathbb{E}(\theta_k - \theta_j | \theta_i) \in [0 \pm \varepsilon] := [-\varepsilon, +\varepsilon]$ is equal to $R^\Delta_i (\theta_i) = [\theta_i \pm \frac{1}{1-|\gamma + \delta|} \varepsilon]$.

Thus, small misspecifications of the moment condition induce small misreports in the mechanism, and hence (given the continuity of the allocation rule) small misallocation relative to the designer’s objective. Hence, the example suggests an ‘almost implementation’ result reminiscent of the literature on virtual implementation, but with the important difference that in this case the allocation is guaranteed to be ‘nearby’ in the allocation space, not in the space of lotteries. Second, note that the size of the misallocation depends on a measure of the allocation is guaranteed to be ‘nearby’in the allocation space, not in the space of lotteries. Second, note that the size of the misallocation depends on a measure of the **strategic externalities** in the belief-based mechanism, $|\gamma + \delta| < 1$: the smaller the strategic externality, the smaller the impact of misspecified moment conditions. Thus, the resilience to such misspecifications is improved by more contractive mechanisms, and maximally so if they achieve dominant strategy implementation (in Example 1, for instance, the rationalizable reports if the moment condition is misspecified are $R^\Delta_i (\theta_i) = [\theta_i \pm \varepsilon]$).

Our next result generalizes these insights. In particular, we show that contractive mechanisms not only ensure continuity with respect to small misspecifications of the moment conditions – which per se can be considered surprising, given the well-known sensitivity of rationalizability to perturbations of common knowledge assumptions – but we also characterize the impact of such misspecifications, relating it to the strength of the strategic externalities in the mechanism.

To this end, consider the transfers defined in (11), using moment condition $\rho = (L_i, f_i)_{i \in I}$, and suppose that the conditions of Theorem 2 are satisfied. Then, for any $\theta_i \in \Theta_i$, we define the smallest own-concavity and strongest strategic externality for $\theta_i$, respectively as:

$$OC^\rho_i (\theta_i) := \min_{(m'_i, \mu) \in M_i \times \Delta^\theta_i} \int_{\Theta_i \times M_i} \left| \frac{\partial W_i (m'_i, m_{-i}, \theta_i, \theta_{-i})}{\partial m_i} - f'_i (m'_i) \right| d\mu$$

$$SE^\rho_i (\theta_i) := \max_{(m'_i, \mu) \in M_i \times \Delta^\theta_i} \sum_{j \neq i} \int_{\Theta_i \times M_{-i}} \left| \frac{\partial W_i (m'_i, m_{-i}, \theta_i, \theta_{-i})}{\partial m_j} + \frac{\partial L_i (m_{-i})}{\partial m_j} \right| d\mu.$$

We take the measure of contractiveness to be the smallest difference (across the types of player $i$) between such own-concavity and strategic externalities:

$$MC^\rho_i := \min_{\theta_i \in \Theta_i} (OC^\rho_i (\theta_i) - SE^\rho_i (\theta_i)).$$

Under the conditions of Theorem 2, $MC^\rho_i > 0$, and the larger the value the stronger the contractiveness of $i$’s best-reply.\(^{17}\) (Note that, in Example 8, $MC^\rho_i = 1 - |\delta + \gamma| > 0$.) We next show that $MC^\rho_i$ also captures the sensitivity to misspecifications of the moment condition $\rho$:

\(^{17}\)Note that $OC^\rho_i (\theta_i)$ and $SE^\rho_i (\theta_i)$ correspond, respectively, to the LHS and RHS of Condition 2 in Theorem 2. Hence, the measure of contractiveness is closely connected to the contractivity gap introduced in Section 4.1.
**Theorem 3** Suppose that \( \rho = (L_i, f_i)_{i \in I} \) satisfies Conditions 1 and 2 of Theorem 2, but it is only approximately satisfied in \( B \): that is, for all \( i, \theta_i \) and \( b_i \in B_{\theta_i} \mathbb{E}_{b_i} (L_i (\theta_{-i} | \theta_i)) \in [f_i (\theta_i) \pm \varepsilon] \) for some \( \varepsilon > 0 \). Then, the transfers \((t_i^B)_{i \in I}\) defined in (11) achieve ‘almost truthful’ \( B \)-implementation. That is, for all \( i \) and \( \theta_i \in \Theta_i \), \( R_i^{BS} (\theta_i) \subseteq \left[ \theta_i \pm \frac{1}{MC_i^\rho} \cdot \varepsilon \right] \).

**Proof.** (See Appendix C.) □

The first implication of this result is a nice continuity property: as the mispecification of the moment condition vanishes (\( \varepsilon \to 0 \)), the mechanism approaches truthful implementation. Moreover, for given \( \varepsilon > 0 \), deviations from truthful implementation are decreasing in the strength of the contraction, measured by \( MC_i^\rho \). The measure of contractiveness in turn is directly related to the strategic externalities and to the own-concavity. These results therefore suggest novel notions of robustness and provide further reasons to pursue contractiveness of the mechanism.

Theorem 3 can also be seen as a generalization of Theorem 2 to accommodate ‘approximate’ moment conditions. In particular, for any \( B \) and \( \varepsilon \geq 0 \), let \( \varrho (B, \varepsilon) \) denote the set of moment conditions that are ‘approximately consistent’ with \( B \). Formally, \( \rho = (L_i, f_i)_{i \in I} \in \varrho (B, \varepsilon) \) if and only if for all \( i, \theta_i \) and \( b_i \in B_{\theta_i} \mathbb{E}_{b_i} (L_i (\theta_{-i} | \theta_i)) \in [f_i (\theta_i) \pm \varepsilon] \). Clearly, \( \varrho (B, 0) = \varrho (B) \) and the set \( \varrho (B, \varepsilon) \) increases with \( \varepsilon \) (for any \( B, \varepsilon' > \varepsilon \) implies \( \varrho (B, \varepsilon) \subseteq \varrho (B, \varepsilon') \)).

**Corollary 2** Let \( \rho = (L_i, f_i)_{i \in I} \in \varrho (B, \varepsilon) \) satisfy Conditions 1 and 2 of Theorem 2. Then, the transfers \((t_i^B)_{i \in I}\) defined in (11) ensure that for all \( i \) and \( \theta_i \in \Theta_i \), \( R_i^{BS} (\theta_i) \subseteq \left[ \theta_i \pm \varepsilon / MC_i^\rho \right] \).

Hence, for ‘exact’ moment conditions (\( \varepsilon = 0 \)), we obtain the truthful implementation result of Theorem 2 as a special case. As \( \varepsilon \) increases and ‘approximate’ moment conditions are included, the implementation result is ‘approximate’. As discussed above, with continuous allocation function this notion is reminiscent of virtual implementation, but with the important difference that in this case the allocation is guaranteed to be ‘nearby’ in the allocation space. This is yet another conceptual innovation which suggests novel directions of research.

### 6 Related Literature

Our work is related to several strands of the literature in game theory and mechanism design. We briefly discuss the most closely related literature.

**Solution Concept.** \( \Delta \)-Rationalizability (Battigalli (2003) and Battigalli and Siniscalchi (2003)) generalizes several versions of rationalizability for incomplete information games, including the ‘belief-free’ version of Bergemann and Morris (2009) and Dekel, Fudenberg and Morris’ (2007) ‘interim correlated rationalizability’ (ICR, also studied by Penta (2013) and Weinstein and Yildiz (2007, 2011, 2013)). Battigalli et al. (2011) provide a thorough analysis of \( \Delta \)-Rationalizability and its connections with other versions of rationalizability.

**Full Implementation.** Within the vast literature on (full) implementation, the closest papers are Bergemann and Morris (2009a) and Oury and Tercieux (2012), which study implementation in ‘belief free’ rationalizability and ICR, respectively. Both ‘belief free’ and ICR-Implementation are special cases of ours, with the proviso that Oury and Tercieux (2012) do not restrict attention to

In particular, suppose that the canonical mechanism failed full implementation (i.e., it had positive contractivity gaps). Then, \( MC_i^\rho > 0 \) represents precisely the extent by which using the moment condition \( \rho \) allowed to offset such positive contractivity gaps, so as to induce a contraction.
direct mechanisms. The restriction to direct mechanisms is also shared by Bergemann and Morris (2009a), while Bergemann and Morris (2011) study belief-free implementation in general mechanisms. Within the classical literature, Jackson (1991) and Postlewaite and Schmeidler's (1986) Bayesian Monotonicity are also connected (see Ollár and Penta (2014)), and our results clearly imply Bayes-Nash Implementation in Bayesian environments. From a conceptual viewpoint, our departure from that literature is inspired by Jackson's (1992) critique of unbounded mechanisms. We push the concern for ‘relevance’ a bit further, requiring that full implementation is achieved via simple transfer schemes.\footnote{D’Aspremont, Cremer and Gerard-Varet (2005) also studied full implementation in environments with transferable utility, but they resort to unbounded mechanisms of the kind criticized above. Duggan and Roberts (2002) fully implement the efficient allocation of pollution via transfers, but under complete information and richer reports.}

In a complete information setting with quadratic preferences, Bergemann and Morris (2007) show that an ascending auction may make implementation easier, relative to its sealed-bid counterpart, by reducing the strategic uncertainty in the mechanism (in the symmetric example with \( n \) agents, full implementation is possible if the preference interdependence parameter satisfies \( \gamma < 1 \) rather than \( \gamma < 1/(n - 1) \).) That insight, however, relies on the complete information assumption (see Penta, 2015) and is orthogonal to the reduction of strategic externalities we pursued here, which ensures implementation for all \( \gamma \).

**Robust Mechanism Design.** As already mentioned, most of the literature on robust mechanism design has focused on the belief-free case. See, for instance, Bergemann and Morris (2005, 2009 and 2011) for static mechanism design, and Mueller (2013, 2015) and Penta (2015) for dynamic mechanism design. Kim and Penta (2012) explore partial implementation with interdependent values, maintaining some restrictions on beliefs. Jehiel et al. (2012) show that, under some restrictions on preferences, minimal notions of robustness are no less demanding than the belief-free case when types are multi-dimensional. This suggests that, when \( \mathcal{B} \) is not a Bayesian type space, the one-dimensionality of \( \Theta_i \) is important for our results. As pointed out, however, our results on full implementation are novel even within the standard (non-robust) Bayesian settings. Lopomo, Rigotti and Shannon (2013) explore partial implementation with belief restrictions analogous to the ones in this paper, but focus on single-agent problems and a different notion of robustness. Artemov, Kunimoto and Serrano (2013) also maintain some restrictions on beliefs, but focus on virtual implementation. Different approaches to robust mechanism design have been put forward by Yamashita (2013a,b), Börgers and Smith (2012,2013), Carroll (2015) and Wolitzky (2014).

**Mechanism Design in TU-Environments.** TU-environments are the typical domain of the partial implementation literature. Within this area, the closest works are those that allow for interdependent values (e.g., Cremer and McLean (1985, 1988), Dasgupta and Maskin (2000), McLean and Postlewaite (2004).\footnote{McLean and Postlewaite (2002) also explore related ideas in environments without transferable utility.} In recent years, a growing literature has revisited standard results, imposing extra desiderata inspired by more practical considerations. The paper by Deb and Pai (2013) is one such example, which pursues symmetry of the mechanism. Mathevet (2010) and Mathevet and Taneva (2013) instead pursue supermodularity. In those papers, the extra desiderata are achieved by adding a belief-dependent component to some baseline payments, very much as we attain full implementation appending an extra term to the canonical transfers.\footnote{Early examples of this principle are the mechanisms of D’Aspremont and Gerard-Varet (1975) and of Cremer and McLean (1985), which append the baseline VCG mechanism with a belief-based component in order to achieve budget balance and surplus extraction, respectively.} One difference is that those papers maintain that types are independently distributed, whereas we...
allow more general correlations, as well as weaker restrictions on beliefs. At a more technical level, our design results in a contractive mechanism. Given our concern with full implementation, contractiveness is a more convenient property than supermodularity. Healy and Mathevet (2013) also pursue contractiveness of the mechanism, though in a complete information setting.

7 Conclusion

Mechanism design is concerned with identifying conditions under which there exist institutions that guarantee socially desirable outcomes through the decentralized interaction of rational agents. The relevance of the theory therefore depends on the nature of the solution concepts adopted to model such interactions, and on the properties of the mechanisms that are used to achieve the results.

In this paper we developed an approach to full implementation based on the solution concept of $\Delta$-Rationalizability (Battigalli (2003) and Battigalli and Siniscalchi (2003)). Our approach subsumes as special cases the important notions of belief-free (Bergemann and Morris, 2009a) and ICR-implementation (Oury and Tercioux, 2012), and accommodates more realistic assumptions on players’ beliefs, intermediate between the ‘belief-free’ and the classical Bayesian benchmarks. In Bayesian settings, which are standard in the applied and classical literature, our conditions also ensure Bayes-Nash Implementation (e.g., Jackson (1991)). Importantly, however, we achieve these results through mechanisms with a clear economic interpretation, and are as simple as those developed by the partial implementation literature. This is an important advance on the literature on full implementation (both Bayesian and robust), and it provides a bridge between two branches of the literature which have typically proceeded in parallel.

While largely inspired by the literature on belief-free mechanism design, we departed from it in important ways. The ability of our framework to accommodate general belief restrictions was key to go beyond the characterization results of the belief-free literature, and provide constructive results on what can still be achieved when agents’ preferences violate the conditions for belief-free implementation. The key insights of focusing on the strategic externalities rather than preferences and of using moment conditions to induce contractive best replies, deliver a clear design principle: start out with the ‘canonical transfers’, which generalize well known necessary conditions for ex-post incentive compatibility, and then add a belief-based component designed to weaken the strategic externalities which may otherwise impair the full implementation result. The resulting mechanism is contractive, and induces truthful revelation as the only rationalizable outcome. These results also show that the methodology developed by the belief-free literature can indeed be extended to address more applied problems of mechanism design, overcoming important limitations of the traditional approach to full implementation.

The basic design strategy of modifying baseline transfers adding a belief-dependent component is also shared by recent work by Mathevet (2010) and Mathevet and Taneva (2013). The difference is that here we pursue contractiveness of the best reply, rather than supermodularity of the mechanism. As shown in Section 5.1, contractiveness ensures other important properties besides uniqueness, such as small sensitivity to perturbations of the moment conditions: small misspecifications of the moment conditions result in an outcome that is proportionately close to the desired one. Importantly, the notion of ‘closeness’ here is in terms of the natural allocation space, as opposed to the probabilistic notion of the virtual implementation literature. Though beyond
the scope of this paper, this suggests that the fundamental logic of our construction can also be extended to moment conditions with inequalities. It also points at a novel notion of approximate implementation, alternative to ‘virtual’, which may be of independent interest for future research.

In Section 5 we discussed some implications of our general results for important special cases, such as environments with single-crossing preferences, with and without common priors. In common prior environments, we provided sufficient conditions for full implementation with independent and correlated types, as well as an equivalence of partially and fully implementable allocation rules that is somewhat reminiscent of Manelli and Vincent’s (2010) important result. In environments with ‘public concavity’, our construction indeed ensures that strategic externalities are completely eliminated, thereby achieving dominant strategy implementation. When this is the case, our results also imply max-min implementation (e.g., Wolitzky (2014), or Carroll (2015)). Hence, our results also show that our design strategy provides useful insights also for max-min implementation.

Appendix

A Proof of Theorem 1

Let \( l := \max_{i, \theta_i} \left\{ \max_{m_i \in R^2_\theta (\theta_i)} |m_i - \theta_i| \right\} \) denote the largest distance between the truthful and some other rationalizable report, across all types. By contradiction, suppose that \( l > 0 \), and let \( i, \theta^*_i \) and \( m^*_i \in R^2_\theta (\theta^*_i) \) be such that \( |m^*_i - \theta^*_i| = l \). Since \( m^*_i \in R^2_\theta (\theta^*_i) \), \( \exists \mu \in \Delta_{\theta^*_i} \cap \Delta \left( R^2_\theta \right) : m^*_i \in \arg \max_{m_i} E U^\mu_{\theta^*_i} (m_i) \). By B-IC we also know that \( \theta_i \in R^2_\theta (\theta_i) \) for all \( \theta_i \) and \( i \), hence \( C^T_i \subseteq R^2_\theta \). Let \( \mu^* \in C^T_i \) be s.t. \( \text{marg}_{\theta_i} \mu^* = \text{marg}_{\theta_i} \mu \). B-IC implies that \( \theta^*_i \in \arg \max_{m_i} E U^\mu_{\theta^*_i} (m_i) \). By construction, \( \mu \) and \( \mu^* \) justify, respectively, \( m^*_i \) and \( \theta^*_i \).

By the assumed B-niceness of the mechanism, best responses are unique and minimize the absolute value of the derivative of the expected utility function. We examine the difference in the first order conditions at the optimum for \( \mu \) and \( \mu^* \), for the case in which \( m^*_i > \theta^*_i \) (the proof is analogous for \( m^*_i < \theta^*_i \)): \( \partial E U^\mu_{\theta^*_i} (m^*_i) / \partial m_i - \partial E U^{\mu^*}_{\theta^*_i} (\theta^*_i) / \partial m_i \), where for any \( \theta_i, m_i, \) and \( \mu \),

\[
\frac{\partial E U^\mu_{\theta^*_i} (m^*_i)}{\partial m_i} = \int_{M_i \times \Theta_i} \frac{\partial U_i}{\partial m_i} (m_i, m_{-i}, \theta_i, \theta_{-i}) \, d\mu. \tag{16}
\]

Since, by assumption, \( E U^\mu_{\theta^*_i} (m_i) \) is strictly concave and maximized at \( m^*_i \), whereas \( E U^{\mu^*}_{\theta^*_i} (m_i) \) is strictly concave and maximized at \( \theta^*_i \), if \( m^*_i > \theta^*_i \) it follows that \( \partial E U^\mu_{\theta^*_i} (m^*_i) / \partial m_i - \partial E U^{\mu^*}_{\theta^*_i} (\theta^*_i) / \partial m_i \geq 0 \). Using (16), this can be rewritten as:

\[
\int_{M_i \times \Theta_i} \frac{\partial U_i}{\partial m_i} (m^*_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu - \int_{M_i \times \Theta_i} \frac{\partial U_i}{\partial m_i} (\theta^*_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu^* \geq 0.
\]

Next, we add and subtract \( \int_{M_i \times \Theta_i} \frac{\partial U_i}{\partial m_i} (\theta^*_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu \), and rearrange terms to obtain:

\[
A := \int_{M_i \times \Theta_i} \frac{\partial U_i}{\partial m_i} (\theta^*_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu - \int_{M_i \times \Theta_i} \frac{\partial U_i}{\partial m_i} (m^*_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu \\
\leq \int_{M_i \times \Theta_i} \frac{\partial U_i}{\partial m_i} (\theta^*_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu - \int_{M_i \times \Theta_i} \frac{\partial U_i}{\partial m_i} (\theta^*_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu^* =: B_i.
\]
By the mean value theorem, there exists \( m'_i \in [\theta^*_i, m^*_i] \) such that:

\[
A_i = \left( \int_{M_{-i} \times \Theta_{-i}} \frac{\partial^2 U_i}{\partial m_i^2} (m'_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu \right) \cdot (\theta^*_i - m^*_i).
\]

Since \( l := (m^*_i - \theta^*_i) > 0 \), and expected payoffs are strictly concave, this can be written as:

\[
A_i = \left| \int_{M_{-i} \times \Theta_{-i}} \frac{\partial^2 U_i}{\partial m_i^2} (m'_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu \right| \cdot l.
\]

Since \( \text{marg}_{\theta_{-i}} \mu^* = \text{marg}_{\theta_{-i}} \mu \) and \( \mu^* \in C^T_i \), the term \( B_i \) can be written as:

\[
B_i = \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i} (\theta^*_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu - \int_{M_{-i} \times \Theta_{-i}} \frac{\partial U_i}{\partial m_i} (\theta^*_i, \theta_{-i}, \theta^*_i, \theta_{-i}) \, d\mu
\]

which by a mean-value Cauchy-Schwarz inequality, is bounded by

\[
B_i \leq \int_{M_{-i} \times \Theta_{-i}} \sum_{j \neq i} \left( \frac{\partial^2 U_i}{\partial m_i \partial m_j} (\theta^*_i, m_{-i}, \theta^*_i, \theta_{-i}) \cdot |\theta_j - m_j| \right) \, d\mu
\]

\[
\leq \int_{M_{-i} \times \Theta_{-i}} \sum_{j \neq i} \left| \frac{\partial^2 U_i}{\partial m_i \partial m_j} U_i (m_i, m_{-i}, \theta^*_i, \theta_{-i}) \right| \, d\mu \cdot l,
\]

where the last step follows from the mean value theorem again. Since \( A_i \leq B_i \), we have that

\[
\left| \int_{M_{-i} \times \Theta_{-i}} \frac{\partial^2 U_i}{\partial m_i^2} (m'_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu \right| \leq \left| \int_{M_{-i} \times \Theta_{-i}} \sum_{j \neq i} \frac{\partial^2 U_i}{\partial m_i \partial m_j} (m_i, m_{-i}, \theta^*_i, \theta_{-i}) \, d\mu \right|
\]

which contradicts the \( B \)-LSE condition for \( i \).

**B** **Proofs from Section 4**

**Lemma 1:** Suppose that \( \mathcal{M} = (d, t) \) is EPIC and \( t \) is differentiable. Then, for every \( i \) and for every \( m \), there exists a function \( \tau_i : \Theta_{-i} \rightarrow \mathbb{R} \) such that \( t_i (m) = t^*_i (m) + \tau_i (m_{-i}) \).

**Proof:** A necessary condition for truthful revelation to be a best response to the opponent truthful revelation at every state (that is, EPIC) is that the following first-order condition is satisfied for every \( i \) and every \( \theta \):

\[
\frac{\partial v_i (d (\theta), \theta)}{\partial x} \cdot \frac{\partial d (\theta)}{\partial \theta_i} + \frac{\partial v^*_i (\theta)}{\partial \theta_i} = 0
\]

hence, \( \frac{\partial v^*_i (\theta)}{\partial \theta_i} = -\frac{\partial v_i (d (\theta), \theta)}{\partial x} \cdot \frac{\partial d (\theta)}{\partial \theta_i} \).
Integrating over $m_i$, it follows that, for any $\theta = (\theta_i, \theta_{-i})$
\begin{equation}
t^*_i(\theta_i, \theta_{-i}) = -\int_0^{\theta_i} \frac{\partial v_i(d(s, \theta_{-i}), s, \theta_{-i})}{\partial x} \cdot \frac{\partial d(s, \theta_{-i})}{\partial \theta_i} ds + K \tag{17}
\end{equation}

Now, for every $i$, define the function $\varpi_i : \Theta \rightarrow \mathbb{R}$ s.t. $\forall (\theta_i, \theta_{-i}) \in \Theta_i \times \Theta_{-i}$, $\varpi_i(\theta_i, \theta_{-i}) = v_i(d(\theta_i, \theta_{-i}), \theta_i, \theta_{-i})$, and notice that
\[ \frac{\partial \varpi_i(\theta_i, \theta_{-i})}{\partial \theta_i} = \frac{\partial v_i(d(\theta_i, \theta_{-i}), \theta_i, \theta_{-i})}{\partial x} \cdot \frac{\partial d(\theta_i, \theta_{-i})}{\partial \theta_i} + \frac{\partial v_i(d(\theta_i, \theta_{-i}), \theta_i, \theta_{-i})}{\partial \theta_i}, \]

hence (17) can be rewritten as
\begin{align}
t^*_i(\theta_i, \theta_{-i}) &= -\int_0^{\theta_i} \frac{\partial \varpi_i(s, \theta_{-i})}{\partial \theta_i} ds + \int_0^{\theta_i} \frac{\partial v_i(d(s, \theta_{-i}), s, \theta_{-i})}{\partial \theta_i} ds + K \tag{18} \\
&= -v_i(d(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) + \int_0^{\theta_i} \frac{\partial v_i(d(s, \theta_{-i}), s, \theta_{-i})}{\partial \theta_i} ds + K + v_i(d(0, \theta_{-i}), 0, \theta_{-i}) . \tag{19}
\end{align}

The result follows letting $\tau_{-i}(\theta_{-i}) = K + v_i(d(0, \theta_{-i}), 0, \theta_{-i})$ for every $\theta_{-i}$.

**Proposition 1:** Allocation rule $d$ is belief-free implementable by a differentiable direct mechanism if and only if the canonical mechanism is belief-free truthfully implementable.

**Proof:** The ‘if’ part is trivial. For the ‘only if’, suppose that $d$ is truthfully belief-free implemented by $\mathcal{M} = (d, t)$. Results in Bergemann and Morris (2009) imply that $\mathcal{M}$ is EPIC, hence by Lemma 1 transfers $t$ can be written as $t_i(m) = \tau_i(m) + \tau_i(m_{-i})$ for some $\tau_i : \Theta_{-i} \rightarrow \mathbb{R}$.

It follows that the ex-post best-responses generated by $\mathcal{M}$ and by the canonical mechanism are identical, but this implies that also the sets of (belief-free) Rationalizable strategies are identical for the two mechanisms. Hence, if $\mathcal{M}$ truthfully implements $d$, so does the canonical mechanism.

**Proof of Theorem 2:**

Consider the mechanism with transfers as in eq. (11). Observe that Condition 2 in the Theorem guarantees niceness of the mechanism. By strict concavity, truthelling is best response to any allowed conjecture concentrated on the truthelling profile, thus the mechanism is $\mathcal{B}$-IC. Condition 1 in the Theorem implies the $\mathcal{B}$-LSE Condition of Theorem 1. The result thus follows from Theorem 1.

**C Proofs from Section 5**

**C.1 Proof of Lemma 2**

**Lemma 2:** If the environment satisfies the single-crossing condition. Then: (1) The canonical mechanism is EPIC if and only if the allocation rule is strictly increasing: $\partial d(\theta) / \partial \theta_i > 0$ for every $\theta$ and every $i$.

**Proof:** To prove (1), notice that truthful revelation satisfies the (necessary) first-order conditions in the canonical mechanism, in that $W_i(\theta, \theta) = 0$ for all $\theta \in \Theta$. Taking the second order
derivative of the ex-post payoff function, and simplifying, we obtain:

$$\frac{\partial^2 U_i^*(d(\theta_i, \theta_{-i}), \theta_i, \theta_{-i})}{\partial^2 m_i} = -\frac{\partial^2 v_i(d(\theta), \theta)}{\partial x \partial \theta_i} \frac{\partial d(\theta)}{\partial \theta_i}.$$  

Since the SCC implies that $\frac{\partial^2 v_i(d(\theta), \theta)}{\partial x \partial \theta_i} > 0$, truthful revelation is uniquely optimal only if $\frac{\partial d(\theta)}{\partial \theta_i} > 0$.

### C.2 Proof of Proposition 5

**Proposition 5:** Under assumptions Q.1-2 and SCC.2, in a common prior environment with independently distributed or affiliated types, an allocation function is EPIC-Implementable if and only if it is iDSIC-implementable.

**Proof:** As explained in the text, the proof of the result follows from Lemma 2 and Propositions 3 and 4, provided that we prove the following Lemma.

**Lemma 3** Under assumptions Q.1-2 and SCC.2, if $d$ is iDSIC, then it is strictly increasing.

**Proof.** Let $\theta_i \in \Theta_i$, $\forall m \in \Theta$, define

$$U_i(m, \theta_i) := \int_{\theta_{-i}} v_i(d(m), \theta_i, \theta_{-i}) \cdot B_{\theta_i}(\theta_{-i}).$$

A necessary condition for truthful revelation to be an (interim) best response independent of the opponents’ strategies is: $\forall \theta_i \in \Theta_i$, $\forall \theta_{-i} \in \Theta_{-i}$,

$$t_i^*(\theta_i, \theta_{-i}) = \int_{0}^{\theta_i} \frac{\partial U_i(s, \theta_{-i}, \theta_i)}{\partial m_i} ds + K.$$

Substituting these transfers and taking the first order conditions of the $t_i$’s optimization problem in the resulting mechanism, it is easy to see that truthful revelation satisfies the (necessary) first-order conditions. Under the maintained assumptions Q.1 and Q.2, the second order derivative of the interim payoff, for each $\theta_i \in \Theta_i$ and $m_{-i} \in \Theta_{-i}$, simplifies to:

$$(S.O.C.) \int_{\theta_{-i}} \frac{\partial^2 U_i^*(d(m_i, m_{-i}), \theta_i, \theta_{-i})}{\partial^2 m_i} \cdot B_{\theta_i}(\theta_{-i}) = -\frac{\partial d(m)}{\partial \theta_i} \int_{\theta_{-i}} \frac{\partial^2 v_i(d(m), \theta_i, \theta_{-i})}{\partial x \partial \theta_i} \cdot B_{\theta_i}(\theta_{-i}).$$

Since SCC.2 implies that $\int_{\theta_{-i}} \frac{\partial^2 v_i(d(m), \theta)}{\partial x \partial \theta_i} B_{\theta_i}(\theta_{-i}) > 0$, truthful revelation is uniquely optimal only if $\frac{\partial d(m)}{\partial \theta_i} > 0$.  

### C.3 Proof of Theorem 3

**Proof:** Fix an arbitrary agent $i$ and a type $\theta_i$. By the definition of $t^\mu$, for any $\mu \in \Delta_{\theta_i}^{\mu}$, adding and subtracting $L_i(\theta_{-i})$, applying triangle inequality and the Newton-Leibniz formula yields the following upper bound:

$$\left| \frac{\partial EU_i^\mu}{\partial m_i} (\theta_i) \right| = \left| \int_{\theta_{-i} \times M_{-i}} \frac{\partial v_i(\theta, d)}{\partial d} - \frac{\partial v_i(\theta_i, m_{-i}, d)}{\partial d} \right| \frac{\partial d}{\partial m_i} + L_i(m_{-i}) - L_i(\theta_{-i}) + L_i(\theta_{-i}) - f_i(\theta_i) d\mu \right|$$

$$\leq \int_{\theta_{-i} \times M_{-i}} \sum_{j \neq i} |D_{ij}E \frac{\partial u_i^\mu}{\partial m_i}(\theta_j - m_j)|^2 d\mu + \epsilon \leq SE_i^\mu(\theta_i) + \epsilon.$$  

(20)
For any \( m_i^1 \in R_i^{A_{i,1}} (\theta_i) \), there exists \( \mu \in \Delta^{B_i} \), such that \( m_i^1 \in BR_{\theta_i} (\mu) \). Since \( m_i^1 \) is best reply, it minimizes the first-order partial derivative. Using (20) and by the concavity of the expected utility function, it follows that for all \( \mu \in \Delta^{B_i}, \)

\[
\left| \frac{\partial EU_{\theta_i}^\mu}{\partial m_i} (m_i^1) - \frac{\partial EU_{\theta_i}^\mu}{\partial m_i} (\theta_i) \right| \leq SE_i^\mu (\theta_i) + \varepsilon.
\]

By the mean value theorem, there exists \( s_i \in M_i \) such that \( |D_iEU_{\theta_i}^\mu (s_i)| |m_i^1 - \theta_i| \leq SE_i^\mu (\theta_i) + \varepsilon \), where the notation \( D_iEU_{\theta_i}^\mu (s_i) \) stands for the derivative \( \partial EU_{\theta_i}^\mu / \partial m_i \) at \( s_i \). Therefore,

\[
|m_i^1 - \theta_i| \leq \frac{SE_i^\mu (\theta_i) + \varepsilon}{OC_i^\mu (\theta_i)}.
\]

Then, for any \( m_i^2 \in R_i^{A_{i,2}} (\theta_i) \), there exists \( \mu \in \Delta^{B_i} \cap R_i^{A_{i,1}} (\theta_i) \) such that \( m_i^2 \in BR_{\theta_i} (\mu) \). Consider a Taylor-expansion of \( \partial EU_{\theta_i}^\mu / \partial m_i \) at \( \theta_i \) around \( m_i^2 \):

\[
\frac{\partial EU_{\theta_i}^\mu}{\partial m_i} (\theta_i) = \frac{\partial EU_{\theta_i}^\mu}{\partial m_i} (m_i^2) + D_iEU_{\theta_i}^\mu (s_i) (\theta_i - m_i^2).
\]

Since \( m_i^2 \) is best reply to \( \mu \) and \( EU_{\theta_i}^\mu (m_i) \) is strictly concave, we have that

\[
|D_iEU_{\theta_i}^\mu (s_i)| |\theta_i - m_i^2| \leq \left| \frac{\partial EU_{\theta_i}^\mu}{\partial m_i} (\theta_i) \right|.
\]

Consider the right-hand side of the above inequality. Adding and subtracting \( L_i (\theta_{-i}) \), applying triangle inequality and the Newton-Leibniz formula, we have that

\[
\left| \frac{\partial EU_{\theta_i}^\mu}{\partial s_i} (\theta_i) \right| = \left| \int_{\theta_{-i} \times M_{-i}} \left( \frac{\partial v_i (\theta, d)}{\partial d} - \frac{\partial v_i (\theta_i, m_{-i}, d)}{\partial d} \right) \frac{\partial d}{\partial m_i} + L_i (m_{-i}) - L_i (\theta_{-i}) + L_i (\theta_{-i}) - f_i (\theta_i) \, d\mu \right|
\]

\[
\leq \int_{\theta_{-i} \times M_{-i}} \sum_{j \neq i} |D_{ij}EU_{\theta_i}^\mu| |\theta_j - m_j^1| \, d\mu + \varepsilon,
\]

from which

\[
|\theta_i - m_i^2| \leq \frac{SE_i^\mu (\theta_i) SE_i^\mu (\theta_i) + \varepsilon}{OC_i^\mu (\theta_i)} + \frac{\varepsilon}{OC_i^\mu (\theta_i)}.
\]

By induction, at the \( k \)th round, we have that

\[
|\theta_i - m_i^k| \leq \left( \frac{SE_i^\mu (\theta_i)}{OC_i^\mu (\theta_i)} \right)^k + \left( \frac{SE_i^\mu (\theta_i)}{OC_i^\mu (\theta_i)} \right)^{k-1} + \frac{SE_i^\mu (\theta_i)}{OC_i^\mu (\theta_i)} + \frac{SE_i^\mu (\theta_i)}{OC_i^\mu (\theta_i)} + \ldots + \frac{SE_i^\mu (\theta_i)}{OC_i^\mu (\theta_i)} + 1 \right) \varepsilon \frac{\varepsilon}{OC_i^\mu (\theta_i)}.
\]

which implies in the limit that for all \( m_i \in R_i^{A_i} (\theta_i) \),

\[
|\theta_i - m_i| \leq \frac{1}{1 - \frac{SE_i^\mu (\theta_i)}{OC_i^\mu (\theta_i)}} \varepsilon \frac{\varepsilon}{OC_i^\mu (\theta_i)} - SE_i^\mu (\theta_i).
\]

Thus we have proven that for all \( i, \theta_i, R_i^{A_i} (\theta_i) \subseteq \left[ \theta_i \pm \frac{\varepsilon}{MC_i} \right] \).
References


