“Incentive Compatibility and Belief Restrictions”

Mariann Ollar and Antonio Penta
Incentive Compatibility and Belief Restrictions

MARIANN OLLÁR
NYU Shanghai and TSE

ANTONIO PENTA
ICREA-UPF, Dept. of Econ. and Business, BSE and TSE

We study a framework for robust mechanism design that can accommodate various degrees of robustness with respect to agents’ beliefs, and which includes both the belief-free and Bayesian settings as special cases. For general belief restrictions, we characterize the set of incentive compatible direct mechanisms in general environments with interdependent values. The necessary conditions that we identify, based on a first-order approach, provide a unified view of several known results, as well as novel ones, including a robust version of the revenue equivalence theorem that holds under a notion of generalized independence that also applies to non-Bayesian settings. Our main characterizations inform the design of belief-based terms, in pursuit of various objectives in mechanism design, including attaining incentive compatibility in environments that violate standard single-crossing and monotonicity conditions. We discuss several implications of these results. For instance, we show that, under

Mariann Ollár: mo2639@nyu.edu
Antonio Penta: antonio.penta@upf.edu

We thank thank the audiences at the Workshop on the Design of Strategic Interaction (Venice, 2023), the Inaugural Janeway Institute Microeconomic Theory Conference (Cambridge, 2024), the Conference on Mechanism and Institution Design (Budapest, 2024), the Lancaster Game Theory Conference (2023), the CUHK Workshop on Economic Theory (2023), and at seminars at UPF, Northwestern, NYU-Shanghai. Antonio Penta acknowledges the financial support of the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).
weak conditions on the belief restrictions, any allocation rule can be implemented, but full rent extraction need not follow. Information rents are generally possible, and they decrease monotonically as the robustness requirements are weakened.

**KEYWORDS.** Moment Conditions, Robust Mechanism Design, Incentive Compatibility, Interdependent Values, Belief Restrictions.

**JEL CLASSIFICATION.** D62, D82, D83.

### 1. INTRODUCTION

Mechanism design has been one of the most successful areas within economic theory. It has deepened our understanding of incentives under private information, providing several theoretical and methodological advances on the way. More broadly, it has had a dramatic impact on the design and understanding of real world mechanisms and institutions. Yet, the classical approach also features some important limitations, particularly due to the strong assumptions on agents’ beliefs that are implicit in standard models, and the key role that they play in several results. The ‘Full Surplus Extraction’ results of Crémer and McLean (1985, 1988) and McAfee and Reny (1992) are notorious examples of findings that “[...] cast doubt on the value of the current mechanism design paradigm as a model of institutional design” (McAfee and Reny (1992), p.400). But several other results, both in game theory and mechanism design, have contributed to motivating Wilson (1987)’s famous call for a “[...] repeated weakening of common knowledge assumptions [...]” in the theory.

A large literature has studied the implications of different relaxations of common knowledge assumptions, and various models of robust mechanism design have been explored. The belief-free approach, spurred by Bergemann and Morris (2005, 2009a,b), has been especially influential. In essence, it requires mechanisms to ‘perform well’, regardless of the agents’ beliefs about each other. But this approach, which voids beliefs of any role, is perhaps too extreme or at least sometimes unnecessarily demanding: in many settings, it may be the case that the designer does possess some information about agents’ beliefs, albeit not necessarily
to the extent that is entailed by the standard Bayesian paradigm. Accounting for this possibility, and providing a systematic analysis of the implications of various degrees of robustness about agents’ beliefs, is key to fulfill the ultimate objective of the Wilson doctrine, “[...] to conduct useful analyses of practical problems [...]” (Wilson, 1987).

In this paper we study a framework that can accommodate various degrees of robustness with respect to agents’ beliefs. This is modeled by means of belief restrictions, $\mathcal{B} = ( (B_{\theta_i})_{\theta_i \in \Theta_i} )_{i \in I}$, where each type $\theta_i \in \Theta_i$ of an agent is endowed with a set of beliefs about others’ types, $B_{\theta_i} \subseteq \Delta(\Theta_{-i})$, that the designer regards as possible. This way, we accommodate as special cases both the classical Bayesian framework (where all such sets are singletons), and the belief-free setting (where $B_{\theta_i} = \Delta(\Theta_{-i})$ for all $i$ and $\theta_i \in \Theta_i$). Crucially, we also accommodate the intermediate cases where the designer can rely on some, but not full, information about agents’ beliefs. Intuitively, the smaller the beliefs sets, the more the designer knows (or is willing to assume) about agents’ beliefs.\(^1\) Within these settings, and for general environments with quasilinear utilities, we characterize the set of B-incentive compatible (B-IC) direct mechanisms: that is, the set of transfers and allocation rules in which truthful revelation is a mutual best-response, for all types and for all beliefs in the belief restrictions. We then discuss several implications of these results.

We start our analysis with the introduction of the canonical transfers. These are the transfers which are pinned down by the first-order conditions that are necessary for truthful revelation to be an ex-post equilibrium of the direct mechanism.

\(^1\)The belief restrictions framework was first introduced in Ollár and Penta (2017), to study how beliefs can be used to attain full implementation, taking incentive compatibility as given (see Ollár and Penta (2022, 2023) for some special cases). Here, in contrast, we tackle the more fundamental question of how beliefs can be used for the very establishment of incentive compatibility, including when single-crossing or monotonicity conditions fail. A related exercise is pursued by Carvajal and Ely (2013), albeit in a standard Bayesian setting. Related approaches to beliefs instead include Jehiel et al. (2012), He and Li (2022), Lopomo et al. (2021, 2022), Gagnon-Bartsch et al. (2021) and Gagnon-Bartsch and Rosato (2023). The related literature is discussed in Section 6.
Thus, they only depend on the ex-post payoffs (and, hence, on agents’ preferences and the allocation rule). Under standard single-crossing conditions, the ex-post payoff functions induced by these transfers are concave at each truthful profile if and only if the allocation rule is increasing, in which case truthful revelation is an ex-post equilibrium, and incentive compatibility is attained in a belief-free sense (ex-post incentive compatibility, ep-IC). But if either single-crossing or monotonicity fail, then the second-order conditions are not met, and ep-IC is not possible. In those cases, suitable modifications of the transfers may restore incentive compatibility, but only by relying on information about beliefs. Whether this is possible, or how, it depends on the information that is available to the designer.

For any $B = (B_{\theta_i})_{\theta_i \in \Theta_i, i \in I}$, suppose that a $B$-IC transfer scheme can be obtained via an additive modification of the canonical transfers. Since, by construction, the canonical transfers ensure that truthful revelation satisfies the first-order conditions (F.O.C.) in the ex-post sense, so they do for all beliefs in $B$. Hence, if an additive modification of the canonical transfers yields a $B$-IC transfer scheme, then it must be that the added term also satisfies the F.O.C., for all beliefs in the belief sets. Theorem 1, in Section 3, shows that this intuition is general: for any belief-restrictions $B$, any $B$-IC transfer can be written as $t_i(m) = t_i^*(m) + \beta_i(m)$, where (letting $m \in M = \Theta$ denote a generic message profile in the direct mechanism) $t_i^*: M \to \mathbb{R}$ denotes the canonical transfers, and $\beta_i: M \to \mathbb{R}$ is a belief-based term that satisfies $\mathbb{E}^{b_{\theta_i}} \left[ \frac{\partial \beta_i}{\partial m_i} (\theta_i, \theta_{-i}) \right] = 0$ for all $\theta_i$ and $b_{\theta_i} \in B_{\theta_i}$.

The bite of the latter condition depends on the richness of the belief sets. It has several direct implications, which provide both a unified view on known results, as well as novel ones. One of the new results is a robust version of the revenue equivalence theorem, which we obtain under a notion of generalized independence that also applies to non-Bayesian settings (Corollary 3). Specifically, if for each agent $i$, the intersection $\bigcap_{\theta_i \in \Theta_i} B_{\theta_i}$ is non-empty, then $B$-IC is possible if and only if it is attained by the canonical transfers, and equilibrium expected payments and payoffs are all pinned down, up to a constant. Note that this condition on the belief-restrictions admits as special cases all belief restrictions in...
which the belief sets of the agents are constant in their types, which in turn include as special cases both the belief-free case, and Bayesian settings with independent types.

Theorem 2 in Section 4 shows that, in order to guarantee that the second-order conditions are satisfied, besides the condition in Theorem 1, the belief-based terms must also satisfy the following:

\[ \mathbb{E}^{b_{\theta_i}} \left[ \frac{\partial^2 \beta_i}{\partial m_i} (\theta_i, \theta_{-i}) \right] \leq -\mathbb{E}^{b_{\theta_i}} \left[ \frac{\partial^2 U^*_i}{\partial m_i} (\theta_i, \theta_{-i}) \right] \]

for all \( \theta_i \) and any \( b_{\theta_i} \in B_{\theta_i} \) (where \( U^*_i (\cdot) \) denotes the payoff function induced by the canonical transfers). A slight strengthening of this condition is also sufficient (Theorem 2). Theorem 3 instead provides a tight characterization that highlights the role of belief-based terms in overcoming failures of standard single-crossing and monotonicity conditions.

These results formalize a general design principle. The main idea is to focus on the design of belief-based terms that satisfy suitable conditions, to be added to the canonical transfers, in order to pursue specific objectives. These may include extra desiderata, beyond incentive compatibility, in settings that satisfy standard single-crossing and monotonicity conditions.\(^2\) But also more fundamental interventions, such as remedying the convexity of the payoff function when single-crossing and monotonicity conditions fail. More broadly, these results identify the scope of \( \mathcal{B} \)-IC in a general class of settings.

For instance, the ‘robust revenue equivalence’ result that we discussed earlier implies that, under generalized independence, there is no scope for improving over the canonical transfers’ ability to achieve incentive compatibility, via the design of belief-based terms. Outside of these cases, however, Proposition 1 shows that a weak responsive moment condition suffices to make any allocation rule

\(^2\) Classic examples of ‘extra desiderata’ include budget balance (d’Aspremont and Gérard-Varet, 1979) or surplus extraction (Crémer and McLean, 1985, 1988; McAfee and Reny, 1992). More recently, other properties have been pursued, such as supermodularity (Mathevet, 2010; Mathevet and Taneva, 2013), contractiveness (Healy and Mathevet, 2012) or uniqueness (Ollár and Penta, 2017, 2022, 2023). Pursuing uniqueness via ‘simple’ mechanisms (as opposed to the classical approach to full implementation (e.g., Maskin, 1999; Palfrey and Srivastava, 1989; Jackson, 1991, etc.) has been the focus of a growing literature on ‘unique implementation’ (cf., Ollár and Penta, 2017, 2022, 2023, 2024b; Winter, 2004; Bernstein and Winter, 2012; Halac et al., 2021, 2022).


\[ d : \Theta \rightarrow X \] incentive compatible, in any environment, via the suitable design of

a belief-based term. Loosely speaking, this condition requires that the designer

knows how agents’ expectations of a moment of the opponents’ types moves,

conditional on their own type, and that this is described by a function that is

nowhere constant. This condition is violated under generalized independence,

but it is very permissive otherwise, thereby showing that minimal knowledge

about agents’ beliefs may go a long way in terms of expanding the possibility of

implementation.

The ‘any \( d \) goes’ result of Proposition 1, which arises discontinuously as gen-

eralized independence is lifted, is somewhat reminiscent of the Crémer and

McLean (1985, 1988) and McAfee and Reny (1992) results on full surplus extration

(FSE), which also arise discontinuously in Bayesian environments, when mini-

mal degrees of correlation are introduced. Importantly, however, FSE does not

generally ensue in our setup. If the belief-restrictions are not Bayesian, even if

any \( d \) can be implemented under the responsive moment condition, there may

still be bounds to the surplus that can be extracted (Propositions 3 and 4). In-

formation rents generally remain, and their size depends on the joint properties

of the allocation rule, agents’ preferences, and the belief restrictions. Moreover,

information rents shrink as the belief sets get finer, and the designer relies on

more information about agents’ beliefs (Proposition 5). At the extreme, if \( B \) is a

Bayesian setting with correlated types, then FSE obtains. In fact, under a novel

‘full rank’ condition, we provide the following ‘anything goes’ result (Proposition

2): in a Bayesian setting that satisfies ‘full rank’, for any \((d, t)\), there exist transfers \( t' \)

that are both incentive compatible and that attain the same expected payments

as \( t \). This in turn implies an exact FSE result for settings with a continuum of

types.\(^3\)

\(^3\) Crémer and McLean (1985, 1988) first studied FSE with finite types. McAfee and Reny (1992) ex-

 tended the result to a continuum of types and to general mechanism design problems. Their con-

 dition does not always ensure exact FSE, but it characterizes almost FSE, in the sense that for any

\( \epsilon > 0 \), there is a mechanism in which agents’ surplus in the truthful equilibrium is less than \( \epsilon \). Our

condition, in contrast, ensures exact FSE. It is stronger than McAfee and Reny’s, but closer in spirit to

Jointly, Propositions 1-5 show that the ultimate source of FSE results is not the comovement between types and beliefs per se, but rather the information that, in standard Bayesian settings, the designer has about agents’ beliefs. This observation highlights an important feature of our framework. Specifically, since their very inception, FSE results have famously been received as disturbing. In response, mechanism design has largely shied away from studying environments with correlated or non-exclusive information. But the pervasiveness and economic relevance of these settings can hardly be underplayed:

“ [...] we should stress that in our opinion the independence assumption should be used only with great caution [...]. It does enable the derivation of results that on the surface look more ‘realistic’ (there is no full extraction of the surplus). However, the derivation of these results rely on a very ‘unrealistic’ assumption. Furthermore, [...] a small deviation from this assumption can induce fundamentally different results.” (Crémer and McLean (1988, p.1255)).

Our results show that the belief-restrictions framework is capable of expressing a meaningful notion of non-exclusive information that is useful for implementation, but without incurring into the pitfalls of FSE. This framework may thus favor mechanism design’s reappropriation of environments with non-exclusive information, in which distilling intuitive and reliable economic intuition has long appeared elusive, within the prevailing paradigm.

In Section 5 we discuss further methodological considerations. Theorem 4, in particular, provides a characterization of the equilibrium payoffs that clarifies the connection between standard envelope formulae and the belief-based terms at the center of our analysis, and to compare the relative merits of the envelope approach and of the first-order approach that we pursued in this paper. Section 6 discusses the related literature. Section 7 concludes.

4The quote from McAfee and Reny (1992) at the beginning of this introduction echos analogous remarks by Crémer and McLean (1988, p.1254): “Economic intuition and informal evidence (we know of no way to test such a proposition) suggest that this result is counterfactual, and several explanations can be suggested.” The influential critique of Neeman (2004) may also be ascribed to this view.
2. Framework

**Payoff Environments.** The payoff environment represents agents’ information about everyone’s preferences over the set of feasible allocations, and an allocation rule that maps agents’ information to the space of allocations, and which represents the designer’s objective. Formally, let $I = \{1, \ldots, n\}$ denote the (finite) set of agents, $X \subseteq \mathbb{R}^m$ the set of allocations. For each $i \in I$, we let $\Theta_i$ denote the set of player $i$’s payoff types, with typical element $\theta_i$, assumed private information. We adopt the standard notation for type profiles, and let $\theta \in \Theta := \times_{i \in I} \Theta_i$, and for each $i$, we let $\theta_{-i} \in \Theta_{-i} := \times_{j \neq i} \Theta_j$. For each $i$, the *valuation function* is denoted $v_i : X \times \Theta \rightarrow \mathbb{R}$. Note that we allow $v_i$ to depend on the entire profile of types, so as to allow the case of interdependent values. For each $i$, we let $t_i \in \mathbb{R}$ denote the monetary transfer to agent $i$, and assume that $i$’s utility for each $(x, t, \theta) \in X \times \mathbb{R}^n$, given type profile $\theta \in \Theta$, is equal to $u_i(x, t, \theta) = v_i(x, \theta) + t_i$. The model can thus accommodate both private and interdependent values, as well as general externalities in consumption, including the cases of pure private goods and public goods. An *allocation rule* is a function $d : \Theta \rightarrow X$, which assigns, to each type profile, the allocation that the designer wishes to implement. We maintain throughout the following assumptions:

**Assumption 1 (Payoff Environment).** $\mathcal{E} = ((\Theta_i, v_i)_{i \in I}, d)$ is such that $\forall i \in I$:

(i) $\Theta_i := [\theta_i^L, \theta_i^U] \subseteq \mathbb{R}$

(ii) $v_i$ is twice continuously differentiable.

(iii) $d$ is piecewise differentiable.\(^5\)

Note that these assumptions require that $d$ is only *piecewise* differentiable in types, and hence the model also accommodates discontinuous allocation rules, which are common for instance in auctions, bilateral trade and assignment.

---

\(^5\)We say that $f : S \rightarrow \mathbb{R}$ is *piecewise differentiable* on a closed and convex set $S \subseteq \mathbb{R}^n$ if there exist a collection $(S_k)_{k=1,\ldots,K}$ of pairwise disjoint convex sets such that $\cup_{k=1}^K S_k = S$, and continuously differentiable functions $g_k : S_k \rightarrow \mathbb{R}$, $k = 1 \ldots K$, such that $f = \sum_{k=1}^K f_k$ where, for each $k = 1, \ldots, K$, $f_k(x) = 1_{[x \in S_k]} \cdot g_k(x)$. 
Belief Restrictions. We model the maintained assumptions on agents’ beliefs via the belief-restrictions we first introduced in Ollár and Penta (2017). We let $\Delta(\Theta_{-i})$ denote the set of probability measures over $\Theta_{-i}$, which represent beliefs about the opponents’ types. Belief restrictions consist of a collection of sets of possible beliefs, for each type of each agent, over the set of type profiles of the other agents. Formally, a belief restriction is a collection $B = ((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$, such that, $B_{\theta_i} \subseteq \Delta(\Theta_{-i})$ is non-empty for each $i$ and $\theta_i$. Belief restrictions can be used to accommodate varying degrees of robustness. For instance:

(i) the belief-free settings of the early literature on robust mechanism design (e.g., Bergemann and Morris (2005, 2009a,b), Penta (2015), etc.) are obtained by letting $B_{\theta_i} = \Delta(\Theta_{-i})$ for all $i$ and $\theta_i \in \Theta_i$, and denoted by $B^{BF} = ((B_{\theta_i}^{BF})_{\theta_i \in \Theta_i})_{i \in I}$;

(ii) standard Bayesian settings correspond to the special case in which belief restrictions are commonly known and each belief set is a singleton for every type: $B_{\theta_i} = \{ b_{\theta_i} \}$ for all $i$ and $\theta_i \in \Theta_i$. In this case, each player’s payoff type uniquely pins down the infinite belief hierarchy, as in the interim formulation in a standard Harsanyi type space. Further, in the special case of a common prior type space, there exists $p \in \Delta(\Theta)$ s.t., for each $i$ and $\theta_i, p(\cdot | \theta_i) = b_{\theta_i} \in \Delta(\Theta_{-i})$. If, furthermore, such a common prior is independent across agents, then we also have $b_{\theta_i} = b_{\theta_i}'$ for all $\theta_i, \theta_i' \in \Theta_i$ and for all $i \in I$.

(iii) intermediate notions of robustness obtain whenever $B_{\theta_i} \subset \Delta(\Theta_{-i})$ for some $\theta_i$. Some special cases have been considered, for instance, by Ollár and Penta (2017) and Ollár and Penta (2023), respectively to model situations in which agents commonly know some moments of the distributions of the opponents’ types (common knowledge of moment conditions), or that agents commonly believe that the opponents’ types are identically distributed (common belief in identicality). The latter belief restrictions, which we denote as $B^{id} = ((B_{\theta_i}^{id})_{\theta_i \in \Theta_i})_{i \in I}$,
are defined for settings with a common set of types (i.e. $\Theta_j = \Theta_k$ for all $j, k \in I$) as follows: $B_{\theta_i}^{id} = \{b_{\theta_i} \in \Delta(\Theta_{-i}) : \text{marg}_{\Theta_j} b_{\theta_i} = \text{marg}_{\Theta_k} b_{\theta_i} \text{ for all } j, k \neq i \}$ for all $i$ and $\theta_i$.

These are just examples of some special cases, but the framework is much more general. We also stress that since the focus here is on partial implementation and incentive compatibility, the results in this paper do not require the belief restrictions to be common knowledge among the agents. Hence, they are just restrictions on the first-order beliefs.

Given belief restrictions $B = ((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$ and $B' = ((B'_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$, we write $B \subseteq B'$ to denote that $B_{\theta_i} \subseteq B'_{\theta_i}$ for all $i \in I$ and all $\theta_i \in \Theta_i$. If $B \subseteq B'$, then $B$ imposes stronger restrictions than $B'$, in that the designer can rule out more beliefs in the former than in the latter. In this sense, the belief-free model $B^{BF}$ is minimal in the information that the designer has, as any model $B$ is such that $B \subseteq B^{BF}$.

At the opposite extreme, any Bayesian setting $B^\circ$ is maximal, as no distinct belief restriction $B$ is such that $B \subseteq B^\circ$. Belief restrictions $B^{id}$ are an example of an intermediate robustness requirement, $B^\circ \subseteq B^{id} \subseteq B^{BF}$.

**Mechanisms.** A mechanism is a tuple $\mathcal{M} = ((M_i)_i, g)$, where $M_i$ denotes the set of messages of player $i$, and $g : M \to X \times \mathbb{R}^n$ is the outcome function, that assigns to each profile of messages, $m \in M := \times_{i \in I} M_i$, an allocation and a profile of payments, $g(m) = (x, t) \in X \times \mathbb{R}^n$. We consider direct mechanisms, in which agents report their type (i.e., $M_i = \Theta_i$ for all $i$) and the allocation is chosen according to $d$ (i.e. $g(m) = (d(m), t(m))$). A direct mechanism therefore is completely pinned down by the transfer scheme $t = (t_i)_{i \in I}$, where for each $i \in I$, $t_i : M \to \mathbb{R}$ specifies the transfer to agent $i$ for all profile of reports $m \in M \equiv \Theta$. Notice that, by definition, each $t_i$ is bounded.

Each (direct) mechanism $(d, t)$ induces a game with incomplete information, with ex-post payoff functions $U^t_i(m; \theta) = v_i(d(m), \theta) + t_i(m)$, which are bounded functions under the maintained assumptions. We adopt the following notation: For any $\theta_i \in \Theta_i$, $b \in \Delta(\Theta_{-i})$ and $m_i \in M_i$, we let $E^b U^t_i(m_i; \theta_i) := \int_{\Theta_{-i}} U^t_i(m_i, \theta_{-i}; \theta_i, \theta_{-i}) \, db$, and for any $f : \Theta \to \mathbb{R}$, $\theta_i \in \Theta_i$ and $b \in B_{\theta_i}$, we let $E^b[f(\theta_i; \theta_{-i})] := \int_{\Theta_{-i}} f(\theta_i, \theta_{-i}) \, db$. 
**Incentive Compatibility.** Incentive compatibility requires that truthtelling is a mutual best response for the agents, for all beliefs that are consistent with the belief restrictions $B$.

**Definition 1.** A direct mechanism $(d, t)$ is $B$-incentive compatible ($B$-IC) if for all $i \in I$, $\theta_i \in \Theta_i$, $m_i \in M_i$, $\mathbb{E}^b U^t_i(m_i; \theta_i) \leq \mathbb{E}^b U^t_i(\theta_i; \theta_i)$ for all $b \in B_{\theta_i}$.

When $d$ is clear from the context, we say that the transfer scheme $t$ is $B$-IC.

Note that in a Bayesian environment, $B$-IC is equivalent to interim (or Bayesian) incentive compatibility (IIC). At the opposite extreme, in belief-free settings it is equivalent to ex-post incentive compatibility (ep-IC). For intermediate belief restrictions, i.e. such that there exists at least some type $\theta_i$ of some agent $i$ for which $B_{\theta_i}$ is a strict subset of $\Delta(\Theta_{-i})$, but not a singleton, then $B$-IC is weaker than ep-IC (since truthful revelation need not be optimal for all beliefs about $\Theta_{-i}$) but it is stronger than IIC (in that it requires truthful revelation to be optimal for all beliefs in $B_{\theta_i}$, not just for one). More generally:

**Remark 1.** If $B \subseteq B'$, and $(d, t)$ is $B'$-IC, then it is also $B$-IC.

### 2.1 Leading Example and Preview of Results

**Example 1 (IIC without Monotonicity (Interdependent Values)).** Two agents, with sets of types $\Theta_i = [0, 1]$ and valuation functions $v_i(x, \theta) = (\theta_i + \gamma \theta_j) x$, for each $i$ and $j \neq i$, where $x \geq 0$ denotes the quantity of a public good, and $\gamma$ is a parameter of preference interdependence. These preferences satisfy the following Single-Crossing Conditions:

\[ \text{(ep-SCC:)} \quad \frac{\partial^2 v_i}{\partial x \partial \theta_i} (x, \theta) > 0 \quad (1) \]

Agents’ types are such that $\theta_i = \theta_0 + \eta_i$, where $\theta_0$ is a (unobserved) common value component, uniformly distributed over $[0, 1/2]$, and $\eta_i$ is an idiosyncratic component, also uniformly distributed over $[0, 1/2]$, independently from $\theta_0$ and $\eta_j$. Agents only observe $\theta_i$. Clearly, this is a standard Bayesian setting (hence,
\[ B_{\theta_i} = \{b_{\theta_i}\} \] for each \( \theta_i \in \Theta_i \), and given the distributional assumptions, the following conditional expectations hold for all \( \theta_i \in \Theta_i \) and \( i \): \( \mathbb{E}^{b_{\theta_i}}(\theta_j) = \mathbb{E}(\theta_j | \theta_i) = \frac{\theta_i}{2} + \frac{1}{4} \).

With cost of production \( c(x) = x^2/2 \), the efficient allocation is \( d^*(\theta) = (1 + \gamma)(\theta_1 + \theta_2) \). As it is well-known, under the single-crossing condition above, an allocation rule is implementable if and only if it is increasing in agents’ types, which is clearly not the case for the efficient allocation rule, if \( \gamma = -2 \). In fact, let us consider the generalized VCG transfers in this setting, and the ex-post payoff functions they induce:

\[
\begin{align*}
t_i^{VCG}(m) &= -(1 + \gamma) \left( \frac{1}{2} m_i^2 + \gamma m_i m_j + \gamma m_j^2 \right), \\
U_i^{VCG}(m, \theta) &= (1 + \gamma)(m_i + m_j)(\theta_i + \gamma \theta_j) - (1 + \gamma) \left( \frac{1}{2} m_i^2 + \gamma m_i m_j + \gamma m_j^2 \right)
\end{align*}
\]

It is easy to check that while truthful revelation satisfies the first-order conditions of the ex-post payoff function, it violates the second order conditions: with \( \gamma = -2 \), \( \partial^2 U_i^{VCG}(\theta, \theta) / \partial^2 m_i = -(1 + \gamma) > 0 \). Thus, due to the combination of the ep-SCC and of the decreasing allocation rule, if the opponents report truthfully, the payoff function induced by the VCG transfers is globally convex, and hence truthful revelation is a local minimum. Ex-post incentive compatibility therefore is impossible in this setting. Furthermore, the VCG transfers are not IIC either: with these transfers, truthful revelation fails the second-order conditions also from the viewpoint of the interim payoffs.

We illustrate next how the VCG transfers may be modified to solve this problem, using information about agents’ beliefs. For example, consider the following modified transfers,

\[
t_i^{\text{mod}}(m) = t_i^{VCG}(m) + (1 + \gamma) \left( m_i^2 + m_i - 4 m_i m_j \right),
\]

which induce the following payoff functions:

\[
\begin{align*}
U_i^{\text{mod}}(m; \theta) &= U_i^{VCG}(m; \theta) + (1 + \gamma) \left( m_i^2 + m_i - 4 m_i m_j \right) = \\
&= (1 + \gamma) \left( (\theta_i + \gamma \theta_j) - (m_i + \gamma m_j) \right) \left( m_i + m_j \right) + \frac{3}{2} m_i^2 + m_i - 3 m_i m_j.
\end{align*}
\]
Taking the first order conditions from the interim payoff function, and evaluating it at the truthful profile, we obtain:

$$\frac{\partial E[b_i|U_{mod}^i(\theta; \theta)]}{\partial m_i} = E[b_i] \left( (1 + \gamma) \left( 2\theta_i + 1 - 4\theta_j \right) \right)$$

$$= (1 + \gamma) \left( 2\theta_i + 1 - 4E[b_i|\theta_j|\theta_i] \right) = 0.$$

Hence, truthful revelation does satisfy the first-order conditions, particularly thanks to the simplification in the last equality, which used the property we highlighted above, that $E[b_i|\theta_j|\theta_i] = E(\theta_j|\theta_i) = \theta_i/2 + 1/4$ for all $\theta_i$. To check the second order conditions, since $\gamma = -2$, we have $\frac{\partial^2 U_{mod}^i}{\partial^2 m_i}(m; \theta) = -1 < 0$. Truthful revelation therefore is a best response to the opponents’ truthful strategy, and hence these modified transfers are IIC. □

Note that the transfers in (2) can be written as $t_{mod}^i(m) = t_{VCG}^i(m) + \beta_i(m)$, where $\beta_i : M \to \mathbb{R}$ is a belief-based term that satisfies $E[b_i]\left[ \frac{\partial \beta_i}{\partial m_i}(\theta_i, \theta_{-i}) \right] = 0$ for all $\theta_i$ and $b_{\theta_i} \in B_{\theta_i}$. Theorem 1 in Section 3 shows that this holds in general: for any belief-restrictions $B$, any $B$-IC transfers must be of this form, provided that $t_{VCG}$ is replaced with a suitable generalization of the VCG mechanism, which we call canonical transfers. Section 3.2 discusses several implications of this result, including a robust version of the revenue equivalence theorem, which we obtain under a notion of generalized independence that also applies to non-Bayesian settings (i.e., the $B_{\theta_i}$ are not all singletons).

The above, however, are not the only IIC transfers in this setting. For instance, if some $t = t_{VCG} + \beta$ is incentive compatible, then truthful revelation satisfies the first-order conditions also for the transfers $t_{VCG} + \alpha\beta$, for any $\alpha \in \mathbb{R}^n$. Incentive compatibility, however, may hold for some $\alpha$ but fail for others.

Example 1 (continued): In the setting of Ex. 1, consider transfers of the form $t_{mod,\alpha}^i(m) = t_{VCG}^i(m) + \alpha_i(1 + \gamma)(m_i^2 + m_i - 4m_i m_j)$. With these transfers, truthful revelation satisfies the second-order conditions if and only if $(1 + \gamma)(2\alpha_i - 1) < 0$. Hence, despite the allocation being decreasing when $\gamma < -1$, IIC is possible here for any $\gamma \in \mathbb{R}$. □
Extending this logic, Theorem 2 in Section 4 implies that, in order to guarantee that the second-order conditions are satisfied, besides the necessary condition above the belief-based terms should also be such that $E^b \left[ \frac{\partial^2 U_{VCG}}{\partial^2 m_i} (\theta_i, \theta_{-i}) \right] < -E^b \left[ \frac{\partial^2 \beta_i}{\partial^2 m_i} (\theta_i, \theta_{-i}) \right]$ for all $\theta_i$ and $b \in B_{\theta_i} \subseteq \Delta (\Theta_{-i})$. Theorem 2 generalizes this insight beyond efficient allocation rules, provided that the VCG transfers are replaced by their suitable generalization. Theorem 3 provides a characterization that highlights the role of belief-based terms in overcoming failures of standard single-crossing and monotonicity conditions. Theorem 4 in Section 5 characterizes the equilibrium payoffs, vis-à-vis standard envelope formulae.

We used Ex. 1 to illustrate the basic logic of our first-order approach, within a standard Bayesian environment and with standard single-crossing conditions. As we discuss in Section 4.3, a lot more can be achieved in this setting. Proposition 2, for instance, implies that, within the context of this example, any allocation rule could be implemented, and inducing any expected payments, including those that extract the full surplus. Outside of Bayesian settings, however, even if weak conditions on beliefs suffice to obtain very permissive implementation results (Proposition 1), informational rents generally remain (Propositions 3 and 4), and they get larger as the robustness requirements get stronger (Proposition 5).

### 3. Generalized Incentive Compatibility: Necessity

In this section we derive necessary conditions for $B$-IC transfers. We first introduce the canonical transfers, $t^* = (t^*_i (\cdot))_{i \in I}$, which are defined as follows: for each $i$ and $m$,

$$t^*_i (m) = -v_i (d (m), m) + \int_{\theta_i}^{m_i} \frac{\partial v_i}{\partial \theta_i} (d (s_i, m_{-i}), s_i, m_{-i}) ds_i. \quad (3)$$
These transfers are pinned down by the necessary conditions for ep-IC, up to an additive term that is constant in own report.\footnote{The ‘canonical transfers’, and the associated \textit{canonical direct mechanism} \((d, t^*)\), should not be confused with the ‘canonical mechanism’, which traditionally refers to Maskin’s (non-direct) mechanism \textit{for full} implementation. Special instances of the canonical direct mechanism have appeared throughout the literature on \textit{partial} implementation, e.g. in the auction mechanisms of Myerson (1981), Dasgupta and Maskin (2000), and Segal (2003), the pivot mechanisms of Milgrom (2004) and Jehiel and Lamy (2018), the public goods mechanisms of Green and Laffont (1977) and Laffont and Maskin (1980), and the one-dimensional results of Jehiel and Moldovanu (2001)).} This characterization of the ep-IC transfers can be obtained both by inverting the \textit{envelope formula} for the ex-post payoff function (Milgrom and Segal, 2002), or directly from the \textit{first-order approach}, which derives the (necessary) local incentive constraints for ep-IC from the first-order conditions of the ex-post payoff function. In this section we provide an analogous result for \(B\)-IC transfers based on a first-order approach. An envelope formulation is discussed in Section 5.2.

\textbf{3.1 A first-order approach}

The main result in this section derives necessary conditions for \(B\)-IC transfers, for general belief restrictions. In our result, we provide a generalization of the classical \textit{first-order approach} that identifies necessary conditions for \textit{local} incentive compatibility constraints (cf. Rogerson (1985); Jewitt (1988)). Compared to the classical results, the main difference is that, instead of focusing on the ex-post payoff function, we take an interim perspective and consider the expected payoff function of every type \(\theta_i\), for all beliefs in the set \(B_{\theta_i}\).

\textbf{THEOREM 1 (\(B\)-IC Transfers (Necessity))}. \textit{Under the maintained assumptions, if \(t\) is piecewise differentiable and \((d, t)\) is \(B\)-IC, then for all \(i\), and for all \(m \in M \equiv \Theta\),}

\[ t_i (m) = t_i^* (m) + \beta_i (m), \tag{4} \]

\textit{where} \(\beta_i : M \rightarrow \mathbb{R}\) \textit{is piecewise differentiable and such that, for all} \(\theta_i\) \textit{and for all beliefs} \(b \in B_{\theta_i}\) \textit{that have a piecewise differentiable pdf, at all points of differentiability,}
\frac{\partial \mathbb{E}^b [\beta_i (m_i, \theta_{-i})]}{\partial m_i} \bigg|_{m_i = \theta_i} = 0. \quad (5)

The result in Equation (4) shows that, in order to design a \( B \)-IC transfer scheme, it is without loss to restrict attention to additive modifications of the canonical transfers, provided that the added terms satisfy the expectation condition in Equation (5). We refer to the functions \( \beta_i : M \to \mathbb{R} \) that satisfy Equation (5) as the belief-based terms that are consistent with \( B \) (or simply belief-based terms, when \( B \) is clear from the context).

### 3.2 Some Direct Implications of Theorem 1

Theorem 1 implies that identifying the set of belief-based terms is crucial to understand the limits of incentive compatibility. For some belief-restrictions, identifying this set, or some of its key properties, is relatively straightforward and delivers immediately interesting insights on the incentive compatible transfers. We discuss a few cases:

#### 3.2.1 Belief-Free Settings

In belief-free settings, \( B^{BF} \), the condition in (5) is required to hold for all beliefs about \( \Theta_{-i} \), including degenerate ones, which is only possible if \( \beta_i \) is constant in \( m_i \). Hence, a transfer scheme is \( B^{BF} \)-IC (that is, ep-IC) only if it coincides with the canonical transfers, up to a function that is constant in agents’ own reports. Thus, when all beliefs are allowed, there are no non-trivial belief-based terms. In this sense, the classical result discussed above obtains as a special case of Theorem 1:

**Corollary 1.** If \( t \) is \( B^{BF} \)-IC, then, \( \forall i, \beta_i (m) := t_i (m) - t^*_i (m) \) is constant in \( m_i \).

#### 3.2.2 Bayesian Settings

In a Bayesian setting, \( B^\circ \), for any agent \( i \) and for any function \( G_i : M \to \mathbb{R} \) that is Lebesgue-integrable with respect to \( m_i \), the term 
\[
f_i (\theta_i) := \mathbb{E}^{b^\circ}_i G_i (\theta_i, \theta_{-i})
\]
is uniquely pinned down by the collection \( (b^\circ_i)_{\theta_i \in \Theta_i} \) of agent \( i \)'s beliefs. Hence, letting
\[
\beta_i (m) := \int_{\theta_i}^{m_i} G_i (s, m_{-i}) \, ds - \int_{\theta_i}^{m_i} f_i (s) \, ds,
\]

we obtain a belief-based term, since $\beta_i$ thus defined satisfies the condition in eq. (5).

In this sense, Bayesian settings are maximal in the set of belief-based terms they admit, since they can be generated starting from any arbitrary $G_i : M \to \mathbb{R}$. This is in stark contrast with the belief-free case, which as seen admits no non-trivial belief-based terms, and hence essentially no incentive compatible transfers other than the canonical ones. Here, the richness of belief-based terms gives rise to a multitude of IIC transfers, which may be used to attain different objectives beyond incentive compatibility. Some of this richness has been exploited by the literature, for instance to pursue budget balance, surplus extraction, supermodularity, contractiveness, or uniqueness (see references in footnote 2). By identifying the key condition on the belief-based terms, Theorem 1 unifies these results and lays the ground to a systematic understanding of the possibilities, and particularly the limits, of IIC.

3.2.3 Independent Types  In Bayesian settings with independent types, the belief sets not only are all singletons, but also contain the same distribution for all types of a player: for each $i$, $B_{\theta_i}^\diamond = \{b_{\theta_i}^\diamond\}$ for all $\theta_i \in \Theta_i$. Then, the condition in eq. (5) implies that, for any belief-based term, its expected value at the truthful profile is constant in the agent’s own type. This is stated formally in point 1 of the next Corollary. In turn, it also implies the following two points:

**Corollary 2.** Let $B^\diamond$ be a Bayesian environment with independent types, and let $b_i^\diamond \in \Delta(\Theta_{-i})$ denote agent $i$’s beliefs, regardless of his type. Then:

(i) If $t$ is $B^\diamond$-IC, then for each $i$, there exists $\kappa_i \in \mathbb{R}$ s.t. $\mathbb{E}^{b_i^\diamond}[\beta_i(m_i, \theta_{-i})] = \kappa_i$ for all $m_i$.

(ii) If $t$ is $B^\diamond$-IC, then for each $i$, there is a $\kappa_i \in \mathbb{R}$ such that, $\mathbb{E}^{b_i^\diamond} t_i(\theta_i, \theta_{-i}) = \mathbb{E}^{b_i^\diamond} [t^*_i(\theta_i, \theta_{-i})] + \kappa_i$ for all $\theta_i \in \Theta_i$.

(iii) $(d, t)$ is $B^\diamond$-IC for some $t$ if and only if $(d, t^*)$ is $B^\diamond$-IC.

Point (ii) is Myerson’s (1981) revenue equivalence, here stated for general environments with interdependent values and independently distributed types. Point (iii) says that an allocation rule is partially implementable, in the sense
of interim (or Bayes-Nash) equilibrium, if and only if it is implemented by the canonical transfers. Intuitively, since all types of an agent share the same beliefs, beliefs are not helpful to screen types, beyond what can be achieved based on the ex-post payoffs. Note that this is not to say that IIC is as demanding as ep-IC: for instance, if single-crossing conditions hold in the interim sense, but not ex-post, then it may be that \( t^\ast \) is IIC, but not ep-IC. Nonetheless, to verify whether some transfers are IIC, it suffices to check whether IIC holds for such transfers: if \( t^\ast \) is not IIC, then no belief-dependent term could recover incentive compatibility.

### 3.2.4 Generalized Independence

The logic above points to another interesting implication of Theorem 1, which suggests introducing the following notion of generalized independence for non-Bayesian settings:

**Definition 2.** \( \mathcal{B} \) satisfies **generalized independence** if, for each \( i \in I, \bigcap_{\theta_i \in \Theta_i} B_{\theta_i} \neq \emptyset \).

This condition is weaker than requiring that the belief sets are constant across types (i.e., \( \forall i \in I, B_{\theta_i} = B_{\theta_i}', \) for all \( \theta, \theta_i' \in \Theta_i \)), which in turn holds in any of the following special cases: (i) belief-free settings; (ii) Bayesian models with independent types; (iii) the \( \mathcal{B}^{\text{id}} \)-restrictions, for common belief in identicality. With this, we obtain the following:

**Corollary 3.** Let \( \mathcal{B} \) satisfy generalized independence, and let \( p_i \in \cap_{\theta_i \in \Theta_i} B_{\theta_i} \). Then:

(i) For any belief-based term \( \beta_i : M \to \mathbb{R}, \exists \kappa_i \in \mathbb{R} \) s.t. \( \mathbb{E}^{p_i}[\beta_i (m_i, \theta_{-i})] = \kappa_i \) for all \( m_i \).

(ii) If \( (d,t) \) is \( \mathcal{B} \)-IC, then for each \( i \), there is a \( \kappa_i \) such that, \( \mathbb{E}^{p_i}t_i (\theta_i, \theta_{-i}) = \mathbb{E}^{p_i}[t^\ast_i (\theta_i, \theta_{-i})] + \kappa_i \) for all \( \theta_i \in \Theta_i \).

(iii) \( (d,t) \) is \( \mathcal{B} \)-IC for some \( t \) if and only if \( (d,t^\ast) \) is \( \mathcal{B} \)-IC.

The discussion that follows Corollary 2 therefore applies to any belief-restrictions that satisfy generalized independence. Point (ii), in particular, extends revenue
equivalence to such non-Bayesian settings as well. All these results follow directly from Theorem 1.8

4. GENERALIZED INCENTIVE COMPATIBILITY: A DESIGN PRINCIPLE

By design, the transfers that satisfy the conditions in Theorem 1 are such that truthful-revelation satisfies the \textit{first-order conditions} of the interim payoff functions, for all beliefs consistent with the belief restrictions for every type. In this sense, these restrictions only reflect local requirements of incentive compatibility. But just like the canonical transfers may fail to be incentive compatible, so may the transfers that satisfy the conditions in Theorem 1. This may be either because truth-telling is a local minimum (e.g., if the payoff function is locally convex) or if it is a local but not a global maximum (which may be the case if the payoff function is not globally concave). Fully understanding incentive compatibility therefore requires exploring what conditions ensure that the payoff function has the right curvature. This is typically what single-crossing and monotonicity conditions do.

In this Section we discuss how the belief-based terms can be used to induce the concavity of the payoff function that is needed to ensure incentive compatibility. In Section 4.1 we first consider the special case of environments with differentiable allocation rules, where Theorem 1 readily delivers tractable necessary and sufficient conditions (Theorem 2). Then, in Section 4.2 we relax the differentiability assumption, and provide a general characterization of the $B$-IC transfers that sheds further light on the role that the belief-based terms have in relation with standard single-crossing and monotonicity conditions (Theorem 3).

\footnote{This Corollary is related to some of the results in Lopomo et al. (2021), who showed that under standard ep-SCC and Monotonicity assumptions, a “full dimensionality” condition on the overlap of the belief sets implies that there is no gap between the possibility of ep-IC and $B$-IC. As we explain in Section 5.1.3, and also using the characterization in Theorem 3, such an equivalence of $B$-IC and ep-IC follows from Corollary 3 and Theorem 3 under standard ep-SCC and Monotonicity conditions, but not necessarily otherwise.}
4.1 B-IC in the differentiable case: a second-order approach

First we consider the special case in which all functions are differentiable. In these settings, Theorem 1 readily delivers the following simple conditions for B-IC:

**Theorem 2 (Conditions under Differentiability).** Assume that \(v_i, t_i, d\) are all twice differentiable, and for each \(i\), let

\[
\beta_i := t_i - t_i^*.
\]

[Necessity:] Transfers \(t = (t_i)_{i \in I}\) are B-IC only if, for all \(i\) and \(\theta_i \in \Theta_i\), for all \(b \in B_{\theta_i}\):

1. \(E^b[\partial_1 \beta_i (\theta_i, \theta_{-i})] = 0\) and
2. there exists an open neighborhood of \(\theta_i, N_{\theta_i}\), s.t. for all \(m_i \in N_{\theta_i}\):

\[
E^b[\partial_2 \beta_i (m_i, \theta_{-i}; \theta_i, \theta_{-i})] \leq -E^b[\partial_2 \beta_i (m_i, \theta_{-i})].
\]

(Sufficiency:) Transfers \(t = (t_i)_{i \in I}\) are B-IC if, for all \(i\) and \(\theta_i \in \Theta_i\), for all \(b \in B_{\theta_i}\), Condition (i) holds and Inequality (6) holds for all \(m_i \in M_i\).

Condition (i) states the necessary condition from Theorem 1, for the differentiable case; Condition (ii) states the necessary second order condition instead, it relates the curvature of the payoff function of the canonical direct mechanism to the belief-based term.

**Example 1 (redux):** In terms of the decomposition from Theorem 1, the belief-based terms in the transfers in eq. (2) are such that

\[
\beta_i(m) = (1 + \gamma)(m_i^2 + m_i - 4m_i m_j),
\]

with first- and second-order derivatives, respectively, \(\partial_1 \beta_i(m) = (1 + \gamma)(2m_i + 1 - 4m_j)\) and \(\partial_2 \beta_i(m) = (1 + \gamma)2\). The expected payoffs of the canonical transfers instead are such that, for all beliefs consistent with the belief-restrictions, \(\partial_2 \mathbb{E}^b_{\theta_i}[U_i^b (m; \theta)] = -(1 + \gamma)\). Hence, \(\beta_i\) satisfies Condition (i) of Theorem 2, since it holds in that setting that \(\mathbb{E}^b_{\theta_i}[2\theta_i + 1 - 4\theta_j] = 0\). Moreover, since with \(\gamma = -2\) the VCG transfers induce convex payoffs, the left-hand side of Condition (ii) is larger than 0, but \(\beta_i\) is concave enough that Condition (ii) holds, so that \(\mathbb{E}^b_{\theta_i}[U_i^{mod}]\) overall is indeed concave in \(m_i\) for all \(\theta_i\) and \(b_{\theta_i} \in B_{\theta_i}\). \(\square\)
Theorem 2 distills a general design principle. To see this, note that the canonical transfers are ep-IC if the term on the left-hand side of (6) is less than zero, i.e. if $U^*_i$ is itself concave. When this is not the case, the belief-based term can be used to relax this constraint: if belief-based terms exist that satisfy Condition (i), and that are sufficiently concave so as to make (6) hold for all $m_i$, then $B$-IC can be attained. The general idea therefore is to identify sufficiently concave belief-based terms, subject to Condition (i) being satisfied. This is useful both to recover incentive compatibility when the canonical transfers do not achieve it, but also to identify the limits of $B$-IC. We illustrate these points with the next example, that exhibits a perhaps starker violation of standard SCM conditions than Ex. 1.

Example 2 (Opposing Interests and Belief Restrictions). A government is deciding on the quantity $x$ of spending in pollution reduction activities. For simplicity, society consists of two agents, and the government’s desired level of expenditure is $d(\theta) = K(\theta_1 + \theta_2)$, where $K > 0$, and $\theta_i \in [0, 1]$ denotes the productivity of agent $i$, which is their private information. Agents work in different sectors, with opposing preferences over pollution reduction, as a function of their productivity: their valuation functions are $v_1(\theta, x) = \theta_1 x$ and $v_2(\theta, x) = -\theta_2 x$, respectively. Clearly, the government’s policy is not efficient in this case. This may be due to political or institutional considerations, which may lead the government to favor a particular agenda, despite the opposite preferences of certain social groups.

The belief restrictions are such that $B_{\theta_i} = \{ b \in \Delta(\Theta_j) : \mathbb{E}^b(\theta_j) = \theta_i/2 \}$, for each $\theta_i$ and $i$. In words, the designer knows that both agents’ expect the opponent’s type, on average, to be half of their own. But beyond this, the actual distributions that describe their beliefs are not known to the designer.

The canonical transfers (eq. (3)) in this problem are such that:

$$t_1^*(m) = -m_1 K(m_1 + m_2) + K \int_0^{m_1} (s + m_2) \, ds = -K \frac{1}{2} m_1^2,$$

and

$$t_2^*(m) = +m_2 K(m_1 + m_2) - K \int_0^{m_2} (m_1 + s) \, ds = K \frac{1}{2} m_2^2.$$
which induce the following payoff functions:

\[ U_1^*(m, \theta) = \theta_1 K (m_1 + m_2) - K \frac{1}{2} m_1^2, \]

\[ U_2^*(m, \theta) = -\theta_2 K (m_1 + m_2) + K \frac{1}{2} m_2^2. \]

Due to the agents’ opposing interests, standard single crossing and monotonicity conditions fail in this setting, and it can be checked that the optimal strategies in \((d, t^*)\) have agent 2 always report extremal messages, either 0 or 1. The canonical transfers therefore are neither ep-IC nor B-IC. The reason is that while truthful revelation satisfies the F.O.C. for both agents, since the allocation rule moves with \(\theta_2\) in the opposite direction of 2’s marginal utility for \(x\), \(U_2^*\) is convex in \(m_2\) and hence the S.O.C. fail for agent 2.

To characterize the set of B-IC transfers, first we identify the set of belief-based terms that satisfy the necessary condition in part 1 of Theorem 2. (We maintain in this example that the lowest type of each agent always pays 0.) In this setting, it can be shown that \(\beta_i : M \to \mathbb{R}\) satisfies such condition if and only if

\[ \partial_i \beta_i (m_i, m_j) = (m_i - 2m_j) H_i (m_i) \]

where \(H_i\) is a real function on \(M_i \equiv \Theta_i\). (It is easy to see that for such \(\beta_i\) function, \(\partial_i \mathbb{E}^b \beta_i (\theta_i) = 0\). The only-if part is less straightforward, and we leave it to the Appendix.) Hence, belief-based terms in this setting must necessarily take the following form:

\[ \beta_i(m) = \int_0^{m_i} (s - 2m_j) H_i(s) ds \]

Notice that, since for each \(\theta_i\) and \(b \in B_{\theta_i}\) we have \(\mathbb{E}^b[\theta_j] = \theta_i/2\) the following simplification occurs for all such beliefs:

\[ \partial_i^2 \mathbb{E}^b[\beta_i (\theta_1, \theta_2)] = H_i(\theta_i) + \left( \theta_i - 2\mathbb{E}^b[\theta_j | \theta_i] \right) H'_i(\theta_i) = H_i(\theta_i) \]

Given this, for agent 1 part 2 of Theorem 2 holds if and only if, for all beliefs consistent with the belief-restrictions, \(-K + \partial_{11}^2 \mathbb{E}^b[\beta_1(\theta_1, \theta_2)] \leq 0\). Exploiting the condition above, this simplifies to \(H_1(\theta_1) \leq K\) for all \(\theta_1\). Similarly, for agent 2 we obtain \(H_2(\theta_2) \leq -K\) for all \(\theta_2\). Hence, a transfer scheme is B-IC if and only if it
takes the form
\begin{align*}
t_1 (m_1, m_2) &= -\frac{1}{2} m_1^2 + \int_0^{m_1} (s - 2m_2) H_1 (s) \, ds, \text{ and} \\
t_2 (m_1, m_2) &= \frac{1}{2} m_2^2 + \int_0^{m_2} (s - 2m_1) H_2 (s) \, ds,
\end{align*}
subject to the restriction on the $H_i$ functions above. Exploiting again the fact that, for each $\theta_i$ and $b \in B_{\theta_i}$, $E^b [\theta_j] = \theta_i / 2$, the expected transfers at the truth-telling profile are:
\begin{align*}
E^b [t_1 (\theta) | \theta_1] &= -\frac{1}{2} \theta_1^2 + \int_0^{\theta_1} (s - \theta_1) H_1 (s) \, ds, \text{ and} \\
E^b [t_2 (\theta) | \theta_2] &= \frac{1}{2} \theta_2^2 + \int_0^{\theta_2} (s - \theta_2) H_2 (s) \, ds,
\end{align*}
from which we can see that they are minimized by setting each $H_i (\theta_i)$ at the corresponding upper bound, that is $H_1 \equiv K$ and $H_2 \equiv -K$. The resulting transfers, $t_1^{\min} (m_1, m_2) = \frac{m_1^2}{2} (K - 1) - 2K m_2 m_1$, and $t_2^{\min} (m_1, m_2) = \frac{m_2^2}{2} (1 - K) + 2K m_1 m_2$, therefore attain the lowest expected transfers to each agent pointwise, for each type realization $\theta \in \Theta$ and regardless of agents' true beliefs within $B_{\theta_i}$. □

4.2 B-IC transfers in the general case: A Full Characterization

We provide next a characterization of the B-IC transfers in general environments, that highlights the role that belief-based terms may play in overcoming failures of standard single-crossing and monotonicity conditions, as it was the case in the previous example.

**Theorem 3 (B-IC: Characterization).** **Under the maintained assumptions of Theorem 1, for each $i$, let $\beta_i := t_i^* - t_i$. Then, $(d, t)$ is B-IC if and only if for all $i, \theta_i, b \in B_{\theta_i}$ and $m_i$:
\begin{align*}
E^b \left[ \int_{m_i}^{\theta_i} \left( \frac{\partial v_i}{\partial \theta_i} (d (s, \theta_{-i}), s, \theta_{-i}) - \frac{\partial v_i}{\partial \theta_i} (d (m_i, \theta_{-i}), s, \theta_{-i}) \right) \, ds \right] &\geq E^b \left[ \beta_i (m_i, \theta_{-i}) - \beta_i (\theta) \right].
\end{align*}
To understand this result, let us first consider the belief-free case, where $\mathcal{B}$-IC coincides with ep-IC. First, as this condition must hold for all beliefs, it must also hold in the ex-post sense, and hence we can just focus on the terms inside the square brackets. Second, as discussed, in belief-free settings the necessary condition in Theorem 1 implies that the belief-based terms are constant in own message, and hence the right-hand side of the conditions in Theorem 3 are equal to zero. Thus, for belief-free settings, the following holds:

**Corollary 4 (ep-IC and ep-SCM).** Under the maintained assumptions of Theorem 1, $(d, t^*)$ is ep-IC if and only if for all $\theta, \theta', \theta_{-i}$:

$$
\frac{\partial v_i}{\partial \theta_i} \left( d\left(\theta_i', \theta_{-i}\right), \theta_i, \theta_{-i}\right) - \frac{\partial v_i}{\partial \theta_i} \left( d\left(\theta_i, \theta_{-i}\right), \theta_i, \theta_{-i}\right) \right) \cdot (\theta_i' - \theta_i) \geq 0.
$$

This condition entails joint restrictions on the single-crossing properties of the valuation functions, and on the monotonicity of the allocation rule. To see this, consider for instance the special case where $(v_i)_{i \in I}$ and $d$ are all everywhere differentiable, and suppose that the valuation functions also satisfy the ep-SCC in eq. (1). Then, the condition in Corollary 4 holds if and only if $\frac{\partial d}{\partial \theta_i}(\theta) \geq 0$ for all $\theta \in \Theta$ and $i \in I$. That is, with ep-SCC, an allocation rule is ex-post partially implementable if and only if it is increasing. Conversely, if the allocation rule is decreasing in all types (i.e., $\frac{\partial d}{\partial \theta_i}(\theta) \leq 0$ for all $\theta \in \Theta$ and $i \in I$), then $(d, t^*)$ is ep-IC if and only if the condition in eq. (1) holds with the reversed inequality, which is exactly what is needed for the conditions in this Corollary to hold. For these reasons, we refer to this condition as **ex-post Single-Crossing and Monotonicity (ep-SCM)**.

Analogously, in a Bayesian setting with independent types, the same logic implies that IIC is possible if and only if a suitable interim-SCM condition is satisfied:

---

9This Corollary generalizes known results on single-crossing and monotonicity conditions to our setting, which allows for not-everywhere differentiable allocation rules.
Corollary 5 (IIC with Independent Types). Let $B^\Theta$ be a Bayesian environment with independent types, and let $b^\Theta_i \in \Delta(\Theta_{-i})$ denote agent $i$'s beliefs, regardless of his type. Then, under the maintained assumptions of Theorem 1, an IIC transfer scheme exists if and only if for all $i$, and for almost all pairs of $\theta_i, \theta'_i$,

$$
\mathbb{E}^{b^\Theta_i} \left[ \frac{\partial v_i}{\partial \theta_i} (d(\theta'_i, \theta_{-i}), \theta_i, \theta_{-i}) - \frac{\partial v_i}{\partial \theta_i} (d(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \right] \cdot (\theta'_i - \theta_i) \geq 0.
$$

Corollaries 4 and 5 provide single-crossing and monotonicity conditions that are 'standard' in the sense that overall they prescribe agents’ marginal valuations and allocations to increase with each agent’s type (either in the ex-post sense, or ‘in expectation’ with respect to $b^\Theta$). Compared to these, the condition in Theorem 3 is more relaxed in the sense that, if the belief restrictions admit non-trivial belief-based terms, then they may be used to ‘fill’ what the environment lacks in terms of the SCM conditions on the left-hand side, by relaxing the constraints on the right-hand sides of the inequality.

The belief-based terms can thus be seen as additional tools to shape agents’ incentives, when standard SCM conditions are not met. The extent to which this is possible depends on the flexibility of the belief-based terms that are available to the designer, depending on the belief-restrictions. As we discussed, these are minimal in settings in which the belief sets do not vary with the type (as in belief-free settings, or in Bayesian settings with independent types, etc.), but they get larger in other cases, and more so as the belief sets get smaller.

4.3 Comovement of Types and Incentive Compatibility

The condition in Theorem 3 entails a certain discontinuity between settings that satisfy generalized independence (Def. 2), and those that do not. In the former, the only available belief-based terms are constant in $m_i$ (cf. Corollary 3.1), and hence they cannot be used to make up for failures of the SCM conditions, since the right-hand side of the condition in Theorem 3 is zero. But as soon as beliefs vary with agents’ types, the possibility of using belief-based terms to recover incentive compatibility suddenly expands.
EXAMPLE 3 (Comovement of types and belief-based terms). Consider the setting of Ex. 2, and replace the belief restrictions with the following, (more general) formulation: $B_{\theta_i} = \{ b \in \Delta(\Theta_j) : \mathbb{E}^b (\theta_j) = \gamma \frac{\theta_i}{2} + (1 - \gamma) \frac{1}{2} \}$, where $\gamma \in [0, 1]$ is a fixed parameter, known to the designer, that captures the degree of comovement between agents’ beliefs and their types: for $\gamma = 1$ we obtain the baseline model from Ex. 2; for $\gamma = 0$ instead the belief restrictions satisfy generalized independence. Since the payoff environment is the same as in Ex. 2, ep-IC is still impossible. In fact, the canonical transfers in this setting are not $B$-IC either, for any $\gamma$, and Corollary 3 and Theorem 3 jointly imply that no transfers are $B$-IC when $\gamma = 0$. Next, consider the following transfers:

$$t_{2 \text{mod}}^m(m) = t_2^*(m) - A \left( \frac{\gamma m_2^2}{2} + (1 - \gamma)m_2 - m_1m_2 \right).$$

(7)

Under these belief restrictions, truthful revelation satisfies the first-order conditions, and $\frac{\partial^2 U_{2 \text{mod}}(m, \theta)}{\partial^2 m_2} = K - A\gamma/2$. Hence, $m_2 = \theta_2$ is optimal for agent 2 whenever $A > 2K/\gamma$, and hence $B$-IC is possible for any $\gamma \in (0, 1]$: an arbitrarily small level of comovement is enough to recover incentive compatibility via the design of a suitable belief-based term. □.

The insight from this example is very general, and goes beyond private values. It extends to a large class of belief restrictions, regardless of the valuation functions and of the allocation rule. The following property of the belief restrictions is key:

**Definition 3.** We say that $B$ admits a responsive moment condition if for each $i$ there exist $L_i : \Theta_{-i} \to \mathbb{R}$ and $f_i : \Theta_i \to \mathbb{R}$ s.t. for all $\theta_i$ and $b \in B_{\theta_i}$, $\mathbb{E}^b L_i (\theta_{-i}) = f_i (\theta_i)$ where $f_i$ is cont. diff. and $f_i'$ is bounded away from 0.

If, furthermore, $B$ is such that, for each $i$ and $\theta_i$, $B_{\theta_i}$ consists of all the beliefs $b_i \in \Delta(\Theta_{-i})$ such that $\mathbb{E}^{b_i} L_i (\theta_{-i}) = f_i (\theta_i)$, then we say that $B$ is maximal with respect to the moment condition $(L_i, f_i)_{i \in I}$.

In words, $B$ admits a moment condition if, for every $i$, there exists a function of the opponents’ types whose expectation given $\theta_i$ is known to the designer (i.e.,
for each $\theta_i$, it is the same for all beliefs in $B_{\theta_i}$). If such expectations are strictly monotonic in $\theta_i$, then we say that the moment condition is responsive. Moment conditions can be seen as pieces of information that the designer may have about agents’ beliefs. In belief-free settings, for instance, only trivial moment conditions (where all $L_i$ and $f_i$ are constant) satisfy the restrictions above, and hence the designer has effectively no information about beliefs. At the opposite extreme, in a Bayesian setting, for any $L_i$ there is a $f_i$ such that $E^{b_i}L_i(\theta_{-i}) = f_i(\theta_i)$ (albeit with $f_i' = 0$ if types are independent, not necessarily otherwise). More broadly, the stricter the belief restrictions, the larger the set of admissible moment conditions, and hence the more information the designer has about agents’ beliefs. The case when $B$ is maximal with respect to some $(L_i, f_i)_{i \in I}$ represents the idea that the specific moment condition is essentially the only information about beliefs that the designer can (or is willing to) rely on.

**Proposition 1.** Fix $v$, and let the belief restrictions admit a responsive moment condition. Then, for any $d$, there exist transfers $t$ such that $(d, t)$ is $B$-IC.

**Proof:** For each agent $i$, let $t_i := t_i^* - A_i \left( \int^{m_i} f_i(s) \, ds - L_i(m_{-i}) m_i \right)$. By the smoothness and implied boundedness assumptions on $v$ and $d$, the left-hand side of the inequality in Theorem 3 is bounded, and hence there exists $A_i$ large (resp., small) enough if $f_i$ is increasing (resp., decreasing) such that the inequality in Theorem 3 holds for $\beta_i(m) = -A_i \left( \int^{m_i} f_i(s) \, ds - L_i(m_{-i}) m_i \right)$. ■

Hence, as long as the belief restrictions admit a responsive moment condition, then any allocation rule can be made $B$-IC by some $t$. (In Ex.3, $L_i(\theta_{-i}) = \theta_j$, and $f_i(\theta_i) = \gamma \theta_i + (1 - \gamma) \frac{m_i}{2}$, which satisfies the condition of the proposition if and only if $\gamma > 0$.)

The discontinuity we illustrated with Ex.3 is reminiscent of another well-known discontinuity in the literature, between Bayesian settings with independent and correlated types, namely Crémer and McLean (1985, 1988) and McAfee and Reny (1992) full-surplus extraction (FSE) results.\(^{10}\) We provide next a novel

\(^{10}\)In Bayesian settings, the result in Proposition 1 can be strengthened: under suitable restrictions, the results in McAfee and Reny (1992) imply that not only any allocation rule is implementable, but
version of FSE, that highlights more clearly how the difference between Bayesian and non-Bayesian settings affects the design of the mechanism.\textsuperscript{11} Our result is based on the following conditions:

**Definition 4.** Let \( B^\circ \) be a Bayesian setting (i.e., \( B^\circ_{\theta_i} = \{ b^\circ_{\theta_i} \} \) for each \( i \) and \( \theta_i \)).

(i) We say that \( B^\circ \) is differentiable if for each \( i \), and for any differentiable \( G : \Theta \to \mathbb{R} \), the function \( f_i : \Theta_i \to \mathbb{R} \), defined as \( f_i(\theta_i) = \mathbb{E}^{b^\circ_{\theta_i}}[G(\theta_i, \theta_{-i})] \), is differentiable.

(ii) We say that \( B^\circ \) satisfies the full rank condition if, for each \( i \), it holds that for any differentiable \( g_i : \Theta_i \to \mathbb{R} \), there exists a Borel-measurable function \( \kappa_i : \Theta_{-i} \to \mathbb{R} \) such that \( \int_{\Theta_{-i}} \kappa_i(\theta_{-i}) \, db^\circ_{\theta_i} = g_i(\theta_i) \) for all \( \theta_i \).

The next proposition shows that, in Bayesian settings that satisfy these conditions, the result in Proposition 1 can be strengthened in the sense that not only any allocation rule can be made IIC, but also the transfers can be chosen so as to match any target for the equilibrium expected payments:

**Proposition 2.** Fix \( v \), and let \( B^\circ \) be a differentiable Bayesian setting that satisfies the full rank condition. Then, for any \( d \) and for any differentiable \( t \), there exist transfers \( t' \) such that: (i) \((d, t')\) is IIC; and (ii) for each \( i \) and \( \theta_i \), \( \mathbb{E}^{b^\circ_{\theta_i}}[t'_i(\theta_i, \theta_{-i})] = \mathbb{E}^{b^\circ_{\theta_i}}[t_i(\theta_i, \theta_{-i})] \).

**Proof:** First note that if \( B^\circ \) is differentiable and satisfies the full rank condition, then there exist functions \((L_i, f_i)_{i \in I}\) that satisfy the condition of Prop. 1. Then, for each \( i \), consider \( \hat{t}_i := \hat{t}^*_i - A_i \left( \int_{m_i} f_i(s) \, ds - L_i(m_{-i}) m_i \right) \). From the proof of Prop. 1, \((d, \hat{t})\) is IIC for \( A_i \) large (small) enough if \( f_i \) is increasing (decreasing). Next, let \( g_i : \Theta_i \to \mathbb{R} \) be defined as \( g_i(\theta_i) := \int_{\Theta_{-i}} [t_i(\theta_i, s) - \hat{t}_i(\theta_i, s)] \, db^\circ_{\theta_i} \) and note that, by construction and Def. 4, \( g_i \) is differentiable in \( \theta_i \). Using the full rank condition, let

\textsuperscript{11}In contrast with the papers in the previous footnote, the sufficient condition we provide for exact FSE next is stronger than McAfee and Reny (1992)'s, but closer in spirit to Crémer and McLean (1988) full rank condition.
\[ \kappa_i : \Theta_{-i} \to \mathbb{R} \] be s.t. \( \int_{\Theta_{-i}} \kappa_i(\theta_{-i}) db_{\theta_i} = g_i(\theta_i) \) for each \( \theta_i \). Then, letting \( t_i' \) be defined as \( t_i'(\theta_i, \theta_{-i}) := \hat{t}_i(\theta_i, \theta_{-i}) + \kappa_i(\theta_{-i}) \), the direct mechanism \( (d, t') \) is both IIC and such that \( \mathbb{E}^{b_{\theta_i}}[t_i'(\theta_i, \theta_{-i})] = \mathbb{E}^{b_{\theta_i}}[t_i(\theta_i, \theta_{-i})] \). \]

The ‘anything goes’ result in this proposition stems from the joint combination of the ‘comovement’ of beliefs and payoff-types and of the environment being Bayesian: In a non-Bayesian setting, such as that in Ex. 3, arbitrary interim payment functions are generally not possible, due to the limited information about agents’ beliefs. The next proposition formalizes this insight: if the designer’s information about agents’ beliefs is limited, albeit still rich enough so as to make any allocation rule implementable, there are restrictions on the incentive compatible transfers.

**Proposition 3.** Consider a differentiable \((v, d)\) and a \( \mathcal{B} \) that is maximal with respect to a responsive moment condition \((L_i, f_i)_{i \in I}\). Then, if \((t_i)_{i \in I}\) is a \( \mathcal{B}\)-IC transfer scheme, for each \( i \) there exist a function \( H_i : M_i \to \mathbb{R} \) such that \( t_i \) can be decomposed as follows:

\[
t_i(m) = t_i^*(m) + \int_{\theta_i}^{m_i} (L_i(m_{-i}) - f_i(s)) H_i(s) \, ds + \tau_i(m_{-i}).
\]

Moreover, there exists a continuous lower bound \( K_i : \Theta_i \to \mathbb{R} \) such that, for any \( \mathcal{B}\)-IC transfer scheme, \( \mathbb{E}^{\mathcal{B}}[\int_{\Theta_i}^{t_i} (L_i(\theta_{-i}) - f_i(s)) H_i(s) \, ds] \geq K_i(\theta_i) \) for all \( \theta_i \) and \( b \in B_{\theta_i} \).

For the next proposition, we say that a function \( g : \Theta \to \mathbb{R} \) is \( L_i \)-linearly separable if it can be written in the form \( g(\theta) = \delta_1(\theta_i) L_i(\theta_{-i}) + \delta_2(\theta_i) \). Additionally, we say that a mechanism \( (d, t) \) is \( \mathcal{B}\)-individually rational (\( \mathcal{B}\)-IR) if, for each \( i \) and \( \theta_i \), \( \mathbb{E}^{b_i} U_i^t(\theta_i; \theta_i) \geq 0 \) for all \( b \in B_{\theta_i} \). Finally, we say that a mechanism extracts the full surplus if the individual rationality constraints hold with equality for all \( i, \theta_i \), and

---

12Recall that, for any \( b \in \Delta(\Theta_{-i}) \), we defined \( \mathbb{E}^{b_i} U_i^t(m_i; \theta_i) := \int_{\Theta_{-i}} U_i^t(m_i, \theta_{-i}; \theta_i, \theta_{-i}) \, db \). Also, in this section we set the outside option to 0 for simplicity, but the extension to type-dependent outside options is easy.
PROPOSITION 4. Consider a differentiable \((v, d)\) and let \(B\) be maximal with respect to a responsive moment condition \((L_i, f_i)_{i \in I}\). Unless for all \(i\), \(\frac{\partial v}{\partial \theta_i} (d(\theta), \theta)\) is \(L_i\)-linearly separable, no transfers \(t\) can extract the full surplus.

The two results together draw a line between the ‘any \(d\) goes’ result for general belief restrictions (Prop. 1), and the ‘anything goes’ result for Bayesian settings (Prop. 2): while, in the latter, any interim payment functions are achievable, the extra robustness requirement in non-Bayesian settings does restrict the possible payments. The next example illustrates the results of Propositions 1-4 and some of the restrictions on the interim payments:

EXAMPLE 3 (continued): Consider again the setting of Ex. 3, with belief restrictions \(B_{\theta_i} = \{b \in \Delta(\Theta_j) : \mathbb{E}^b[\theta_j] = \gamma \theta_j^2 + (1 - \gamma) \frac{1}{2}\}\). For simplicity, let us consider the case where \(\gamma \in [0, 1/2]\). As we already discussed, the conditions of Prop. 1 hold, and \(B\)-IC is attained by the transfers in eq. (7), as long as \(A > 2K/\gamma\) and for any \(\gamma > 0\).

\(\text{Figure 1. Possible Expected Payments to the Agents in Ex. 3: } B\text{-IC under } t_i(0, \theta_{-i}) \equiv 0.\) The thick black line, in both figures, is the expected canonical transfer to each agent (feasible for agent 1 but infeasible for agent 2). The gray area represents the possible interim payments under partial implementation (resulting from possibly different transfer schemes, with the restriction that the lowest type pays zero).
Figure 1 plots the range of expected payments (as a function of $\theta_i$, for any $b \in B_\theta$) that are associated with $B$-IC transfers and the condition that the lowest type pays 0. If, however, the designer's model consists of a Bayesian setting that also satisfies the conditions of Prop. 2, then any expected payments can be induced in an incentive compatible way. For instance, let $B^\circ$ be such that, for each $\theta_i$, $b^\circ_\theta_i$ consists of a mixture of two independent uniform distributions, over $[0, \theta_i]$ and $[0, 1]$, respectively with weights $\gamma$ and $(1 - \gamma)$. Then, mimicking the proof of Prop. 2, we can consider for surplus extraction our 'target' transfers to be $t_i(\theta) = -v_i(d(\theta), \theta)$, which would attain FSE, and obtain the expected difference $g_i(\theta_i) = \int_{\Theta_j} (t_j - \hat{t}_j) \, db_\theta_j$, where $\hat{t}_i$ is a suitable IIC transfer.

For agent 1, the canonical transfers are IIC, and hence they can be used in the role of $\hat{t}_1$. The integral equation $\int_{\Theta_2} \kappa_1(\theta_2) \, db_{\theta_1} = -K \left[ \gamma \frac{\theta_2^2}{2} + (1 - \gamma) \theta_2 \right]$ solved for $\kappa_1(\cdot)$ gives $\kappa_1(\theta_2) = \frac{K(1 + \gamma)}{\gamma} \left[ \theta_2(2 + \gamma) + (1 - \gamma) \right]$ if $\theta_2 \in [0, \gamma]$ and $\kappa_1(\theta_2) = 0$ otherwise. (See Appendix B for the solution of this class of integral equations.) For agent 2, we can take $\hat{t}_2(\theta) = t_2^*(\theta) - A \left( \gamma \frac{\theta_2^2}{2} + (1 - \gamma) \theta_2 \right)$ from eq. (7), which is IIC for $A > 2K/\gamma$. The integral equation $\int_{\Theta_1} \kappa_2(\theta_1) \, db_{\theta_2} = \frac{\theta_2^2}{2} \left[ K(1 + \gamma) - \gamma \frac{A}{2} \right] + K(1 - \gamma) \theta_2$ solved for $\kappa_2(\cdot)$ gives $\kappa_2(\theta_1) = -\frac{1(1 - \gamma)}{\gamma} \left[ \theta_1(2 + \gamma) \left( K(1 + \gamma) - \gamma \frac{A}{2} \right) + (1 - \gamma) K \right]$ if $\theta_1 \in [0, \gamma]$ and $\kappa_2(\theta_1) = 0$ otherwise. The resulting transfers, $t_i^* = \hat{t}_i + \kappa_i$, preserve IIC and at the same time extract all the surplus from both agents. Moreover, any other differentiable $t_i$ payments can be matched by constructing transfers this way. □

Hence, information rents remain, even within models where agents’ beliefs might play a role in facilitating the implementation task. If the belief-restrictions are not Bayesian, even if any $d$ can be implemented under the condition of Proposition 1, there may still be bounds to the surplus that can be extracted. The size of the information rents depends on the joint properties of the allocation rule, agents’ preferences, and the belief restrictions, and they get get larger as the robustness requirement strengthens (i.e., as the belief sets get larger).

To formalize these statements, for any $(v, d)$, and for any belief restrictions $B$, let $F(B)$ denote the set of transfer schemes that are both $B$-IC and $B$-individually
rational, and let \( V(B) \) denote the set of all triplets \((i, \theta_i, b)\) such that \( i \in I, \theta_i \in \Theta_i \) and \( b \in B_{\theta_i} \). Then, define:

\[
\tau(B) := \inf_{t \in F(B)} \sup_{(i, \theta_i, b) \in V(B)} \mathbb{E}^b U^t_i(\theta_i; \theta_i)
\]

if \( F(B) \) is non-empty, and \( \tau(B) := \infty \) otherwise.

First note that, with this notation, FSE obtains if and only if there exists \( t \in F(B) \) such that the constraint for \( B - IR \) holds with equality for all types of all agents, i.e. if \( \tau(B) = 0 \). If \( \infty > \tau(B) > 0 \), in contrast, in each incentive compatible and individually rational mechanism there is at least some type that enjoys strictly positive rents. This bound to the designer’s ability to extract surplus, however, decreases monotonically as belief restrictions get finer. At the extreme, if \( B \) is a Bayesian setting with correlated types, then FSE obtains.

**Proposition 5.** For any \((v, d)\), and for any \( B: B' \subseteq B \) implies \( \tau(B') \leq \tau(B) \). Moreover, if \( \tau(B^{BF}) > 0 \), then there exist \( B \) and \( B' \) such that\(^{13}\) (i) \( B \) admits a responsive moment condition (Def. 3) and is such that \( 0 < \tau(B) < \infty \); (ii) \( B' \subset B \) and is such that \( \tau(B') = 0 \).

The weak monotonicity of \( \tau(\cdot) \) with respect to set inclusion follows directly from the definition of \( B - IC \). The rest of the proposition states that – unless the environment is trivial – there always exist belief restrictions \( B \) in which FSE is not possible, despite \( B \) already granting maximal flexibility in implementing any allocation rule via belief-based terms. FSE can be achieved, but only by relying on extra information \( B' \subset B \) about beliefs. Hence, in essentially any environment beliefs can play a meaningful role to expand the possibility of implementation, without entailing FSE.

\(^{13}\)Note that \( \tau(B^{BF}) = 0 \) only holds in trivial environments, in which each \( v_i \) is constant in own type.
5. DISCUSSION

5.1 Implications of Theorem 1

5.1.1 On the Richness of Belief-based terms in Bayesian Settings

As we mentioned in Section 3.2.2, in a Bayesian setting, $B^o$, for any $i \in I$ and for any $G_i : M \rightarrow \mathbb{R}$ that is Lebesgue-integrable with respect to $m_i$, the function $f_i(\theta_i) := \mathbb{E}^{b_i} G_i(\theta_i, \theta_{-i})$ is uniquely pinned down by agent $i$’s beliefs. Hence, letting $\beta_i(m) := \int_{\theta_i} G_i(s, m_{-i}) ds - \int_{\theta_i} f_i(s) ds$, we obtain a viable belief-based term, since $\beta_i$ thus defined satisfies condition (5) in Theorem 1. The results in the previous section showed how this richness, and the associated freedom to choose such functions, can be used to obtain full-surplus extraction. Other results in the literature have also exploited this richness, to obtain various results (cf. footnote 2). We will return to this point throughout this Section.

5.1.2 On Bayesian Settings with Independent Types

The result in point 1 of Corollary 2 formalizes why with independent types it is with no essential loss of generality to study incentive compatibility as if there were a single agent. When this condition does not hold, however, the heterogeneity of beliefs across a player’s types may indeed expand the set of feasible interim payments and implementable allocation rules, and hence the reduction to a single-agent setting is not without loss.

Note, however, that even with independence, and notwithstanding the payoff-equivalence of all IIC transfers, there may still be a value in characterizing the full set, beyond the canonical transfers. That is if the designer has other objectives, beyond mere incentive compatibility. In these cases, the single-agent approach does entail a loss of generality, even with independent types.

EXAMPLE 4 (Independence and Multiplicity). Consider the environment from Ex. 1, but now assume that types are i.i.d. draws from the uniform distribution over $[0, 1]$. Then, Corollary 2 implies that IIC is possible if and only if the VCG transfers are IIC. In turn, Corollary 5 ensures that this is the case if and only if $\gamma \geq -1$. 


Next, suppose that $\gamma = 3/2$, and consider the following transfers:

$$t^\text{full}_i = t^\text{VCG}_i + \alpha_i \left( m_j - \frac{1}{2} \right) (1 + \gamma) m_i$$

With $\gamma = 3/2$, the VCG transfers are IIC. Furthermore, since $\mathbb{E}^b[\theta_j|\theta_i] = 1/2$ for all $\theta_i$, these modified transfers satisfy both conditions in Theorem 2 for any $\alpha_i$. While this richness of transfers is redundant from the viewpoint of IIC alone, it may still be useful for other purposes. For instance, if one also cares about unique implementation, with $\gamma = 3/2$ the VCG transfers induce too strong strategic externalities, and hence multiplicity of equilibria. The results from Ollár and Penta (2017) ensure that truthful revelation is the only rationalizable strategy (and, hence, also the unique equilibrium) for $\alpha_i \in (1/2, 5/2)$. In fact, for $\alpha_i = \gamma$, truthful revelation is an interim dominant strategy. □

5.1.3 On Generalized Independence Corollary 3 generalizes Theorem 1 in Ollár and Penta (2023), which only focused on the $\mathcal{B}^{\text{id}}$-restrictions (i.e., under common belief in identicality), and it sheds light on some influential results in Lopomo et al. (2021) and in Jehiel et al. (2012)).

Lopomo et al. (2021) showed that, under standard single-crossing and monotonicity assumptions, a “full dimensionality” condition on the overlap of the belief sets implies that there is no gap between the possibility of $\mathcal{B}$-IC and ep-IC. First note that our notion of generalized independence is weaker than the analogous condition in Lopomo et al. (2022), as it does not impose any form of full-dimensionality on the overlap of the belief sets. Furthermore, under generalized independence, $\mathcal{B}$-IC is possible if and only if it is achieved by the canonical transfers (Corollary 3). Under standard ex-post SCM conditions, the canonical transfers are ep-IC (Corollary 4), and hence our results also imply that—under generalized independence—there is no gap between the possibility of ep-IC and $\mathcal{B}$-IC.
But without ep-SCC, as in our general setting, the canonical transfers may be $B$-IC without necessarily being ep-IC. Then, it would not be the case that $B$-IC and ep-IC coincide, although revenue equivalence would still hold (Corollary 3.2).

5.2 Equilibrium Payoffs: An Envelope Formulation

Theorem 3 implies the following characterization of the equilibrium payoffs of $B$-IC mechanisms:

**Theorem 4 (Payoff Characterization).** Fix belief restrictions $B$ and allocation rule $d$. For each $i$, let $D_i \subseteq \mathbb{R}^\Theta$ denote the set of all belief-based terms that satisfy the conditions of Theorem 3. Then, $(U_i)_{i \in I} \in \times_{i \in I} \mathbb{R}^\Theta$ is a feasible payoff-function in the truthful equilibrium of a $B$-IC mechanism if and only if, for each $i$, there exists $\beta_i \in D_i$ such that

$$U_i(\theta_i, \theta_{-i}; \theta) = \int_{\theta_i}^{\theta_i} \frac{\partial v_i}{\partial \theta_i}(d(s, \theta_{-i}), s, \theta_{-i}) \, ds + \beta_i(\theta_i, \theta_{-i}).$$

This formulation of the equilibrium payoffs resembles well-known envelope conditions that characterize the equilibrium payoffs of incentive compatible transfers. In fact, Theorem 4 generalizes several such results along different dimensions. It also highlights the limitations of pursuing an envelope approach either when beliefs do not fall within certain special cases, or when the designer has other objectives beyond mere incentive compatibility.

To see this, first suppose that the environment is belief-free. Then, by Corollary 1, the set $D_i$ only contains $\beta_i : \Theta \to \mathbb{R}$ that are constant in $m_i$, and hence (8) boils down to the standard envelope condition (3) in Milgrom and Segal (2002).

More generally, for belief-restrictions that satisfy generalized independence (cf. Def. 2), and letting $b \in \cap_{\theta_i \in \Theta_i} B_{\theta_i}$, then all $\beta_i \in D_i$ are such that $E^b(\beta_i)$ is constant in $m_i$ (Corollary 3), and hence also in this case the formula in (8) delivers the

---

14 Ollár and Penta (2023) provide an example of this possibility within the context of the $B^{id}$-restrictions.
submitted 'integral condition' for the interim expected payoffs, $E^b(U_i)$, here general-
ized to accommodate both the possibility of interdependent values as well as non-Bayesian settings with generalized independence.

Thus, when $E^b(\beta_i)$ is constant in $m_i$ for all $\beta_i \in D_i$, the interim expected equi-
librium payoffs under incentive compatibility are effectively pinned down, up to a constant in own message, and hence this formula can be used to obtain the incentive compatible transfers, by inverting the integral condition and using the fact that $U_i(m, \theta) = v_i(d(m), \theta) + t_i(m)$. But when the set $D_i$ is richer than that, then there is a non-trivial multiplicity of payoff functions, each with its own envelope condition. In these cases, which include for instance Bayesian settings with correlated types, the payoff function is only determined once the transfers are fixed, and hence the envelope formula cannot be used to recover the incentive compatible transfers. The multiplicity of transfers determines a family of envelope conditions, for distinct belief-dependent terms in $D_i$.

Finally, even when the envelope approach can be used to recover the incentive compatible transfers (as under generalized independence), it still overlooks the richness of the set of incentive compatible transfers, which may be useful for other purposes beyond incentive compatibility. For instance, in Bayesian settings with independent types, the expected payments for all IIC transfers only differ up to a constant in own message. Such transfers, however, may induce different payoffs at non-equilibrium profiles, and hence exhibit different properties with respect to other objectives, such as uniqueness, budget balance, etc. (see, e.g., Ex. 4 above). In this sense, also in such settings the envelope approach is more limited than the first-order approach that we pursue in this paper.

6. RELATED LITERATURE

This paper contributes to the literature on robust mechanism design, particularly following the approach in Bergemann and Morris (2005), that is to achieve implementa-
tion of a given allocation rule for a large set of beliefs. The first wave of this literature focused on belief-free environments. More specifically, Bergemann and
Morris (2005, 2009a,b) study belief-free implementation in static settings, respectively in the partial, full and virtual implementation sense. The belief-free approach has been extended to dynamic settings by Müller (2016) and Penta (2015). Penta (2015) considers environments in which agents may obtain information over time, and applies a dynamic version of rationalizability based on a backward induction logic (cf. Penta (2011) and Catonini and Penta (2022)). Müller (2016) instead studies virtual implementation via dynamic mechanisms, in a static belief-free environment, using a stronger version of rationalizability with forward induction.

Belief restrictions as a way to introduce intermediate notions of robustness (as well as unify also the belief-free and Bayesian benchmarks) were first introduced in Ollár and Penta (2017), and some special cases are analyzed in Ollár and Penta (2022, 2023, 2024b), with the objective of studying how information about beliefs could be used to obtain unique implementations in settings in which incentive compatibility followed directly from standard assumptions. In this paper, in contrast, we focused on the more fundamental question of how beliefs can be used for the very establishment of incentive compatibility.

From a methodological viewpoint, we pursued a generalization of the classical first-order approach that identifies necessary conditions for local incentive compatibility constraints (cf. Rogerson (1985); Jewitt (1988)), and then studies sufficient conditions for global optimality. This methodological shift is necessary to account for the general belief restrictions we consider, and particularly for those that do not satisfy ‘generalized independence’, where the envelope formula cannot be used. But it also brings to the forefront a hitherto neglected richness of incentive compatible transfers also when the conditions for the envelope theorems hold (including, as discussed, Bayesian settings with independent types). Carvajal and Ely (2013) also studied the design of incentive compatible mechanisms in settings in which the envelope formula cannot be used, due to non-convexity or non-differentiability of the valuations, but only within standard Bayesian settings. Related ways of modeling robustness have been explored instead by He and Li (2022), Lopomo et al. (2021, 2022), Gagnon-Bartsch et al. (2021), and Gagnon-Bartsch and Rosato (2023).
Several papers have used special cases of belief restrictions to model robustness with respect to local perturbations around a given Bayesian belief-setting. For instance, Jehiel et al. (2012) show that, under certain restrictions on preferences, minimal notions of robustness are as demanding as the belief-free case. A similar result is proven in Lopomo et al. (2021), for overlapping beliefs, and in Lopomo et al. (2022), within an auction setting. As discussed, these results are in line with those we obtain under generalized independence (cf. Corollary 3). The exact connections between our results and those of these papers are discussed in Sections 3 and 5. In terms of the framework, the belief-restrictions that we consider encompass the belief sets studied by the above papers. In contrast to those papers, we develop a first-order approach and also provide several possibility results for transfer design under various degrees of robustness. Lopomo et al. (2021), on the other hand, also consider more general preferences, which are beyond the scope of our work (notably, their model allows for preferences that are not necessarily quasilinear in transfers, as well as the possibility of incomplete preferences due to Knightian uncertainty).

Several alternative approaches to robustness have been put forward. For instance, Börgers and Smith (2012, 2014), focus on the role of eliciting beliefs to weakly implement a correspondence in a belief-free setting. Börgers and Li (2019) provide a more systematic analysis of implementation relying on first-order beliefs. Other approaches model robustness with respect to certain behavioral concerns directly in the implementation concept. These include criteria such as credibility of the designer (Akbarpour and Li (2020)), a behavioral notion of strong strategy proofness (Li (2017)), safety considerations with respect to model misspecification (Gavan and Penta (2023)), convergence of best response dynamics (Mathevet (2010); Mathevet and Taneva (2013); Healy and Mathevet (2012), and Sandholm (2002, 2005, 2007)), etc.

Yet another approach is based on maxmin criteria, as pursued for example by Chung and Ely (2007); Chassang (2013); Carroll (2015); Yamashita (2015); He and Li (2022). The aim here is typically to explore whether ‘natural’ mechanisms can be justified as worst-case optimal, within a suitable robustness set (see Carroll (2019) for a survey of this literature). In this paper, in contrast, we fix an allocation
rule and require implementation not only for the worst-case beliefs, but for all beliefs in the robustness set. In this sense, our approach is closer to the original belief-free approach of Bergemann and Morris (2005, 2009a,b).

7. Conclusions

We studied incentive compatibility in a general framework for robust mechanism design, that can accommodate various degrees of robustness with respect to agents’ beliefs, and which includes as special cases both belief-free (e.g., Bergemann and Morris (2005, 2009a,b)) and standard Bayesian settings. For general belief restrictions, we characterized the set of incentive compatible direct mechanisms in general environments with interdependent values. The necessary conditions that we identified, based on a first-order approach, provide a unified view of several known results, as well as novel ones, including a robust version of the revenue equivalence theorem that holds under a notion of generalized independence that also applies to non-Bayesian settings.

From a methodological perspective, we showed that, in spite of its simplicity, a suitable generalization of the classical first-order approach (e.g., Laffont and Maskin (1980); Rogerson (1985); Jewitt (1988), etc.), allows a wealth of novel results: (i) on the one hand, it identifies the class of incentive compatible transfers in settings which cannot be handled with the standard envelope approach (such as in Bayesian settings with correlated types, or with general belief restrictions); (ii) on the other hand, even in settings where the the equilibrium payoffs are pinned down by the envelope approach (e.g., under generalized independence – cf. Corollary 3 and Theorem 4), it identifies the richness of incentive compatible transfers that may serve purposes beyond incentive compatibility (such as budget balance (d’Aspremont and Gérard-Varet, 1979), stability (Mathevet (2010); Mathevet and Taneva (2013); Healy and Mathevet (2012), and Sandholm (2002, 2005, 2007)), uniqueness (Ollár and Penta, 2017, 2022, 2023), etc.), which has hitherto escaped a unified, systematic analysis. Both of these features allow several directions for possible future research.
Our main results inform the design of belief-based terms, in pursuit of various objectives in mechanism design, including attaining incentive compatibility in environments that violate standard single-crossing and monotonicity conditions. Outside of environments with generalized independence, we showed that minimal information about agents’ beliefs may suffice to implement any allocation rule. Yet, if the setting is non-Bayesian, information rents are generally possible, and they get larger the less information the designer has about agents’ beliefs. Our belief restrictions may thus capture a meaningful notion of ‘comovement’ of beliefs and types that is useful for implementation, but without incurring into the pitfalls of ‘full-surplus extraction’ results (cf. Crémer and McLean, 1985, 1988). This framework may thus favor mechanism design’s reappropriation of environments with non-exclusive information, in which distilling intuitive and reliable economic intuition has long appeared elusive, within the prevailing paradigm. We believe that this is a valuable feature of our framework, which enables exploring several novel questions.

REFERENCES


Carroll, G. (2019). Robustness in mechanism design and contracting. *Annual Review of Economics* 11, 139–166. [38]


Jackson, M. O. Bayesian implementation *Econometrica* 59, 461–477. [5]


Appendix

APPENDIX A: PROOFS

Proof of Theorem 1. Fix an agent \( i \). First, we show that \( t^*_i(m) \) is well-defined since the allocation rule \( d \) is p.diff. \(^{15}\) Since \( v_i \) is twice continuously differentiable, \( \frac{\partial v_i}{\partial \theta_i} \) is continuously differentiable over \( X \times \Theta \). Now, for fixed \( m_{-i} \), \( \frac{\partial v_i}{\partial \theta_i}(d(\cdot, m_{-i}), \cdot, m_{-i}) \) – a function from \( M_i \) to \( \mathbb{R} \) – is a composite function of \( d \) and \( \frac{\partial v_i}{\partial \theta_i} \) and since \( d \) is piecewise differentiable over \( \Theta_i \), we have that for all \( m_{-i}, \frac{\partial v_i}{\partial \theta_i}(d(\cdot, m_{-i}), \cdot, m_{-i}) \), a function from \( M_i \) to \( \mathbb{R} \), is piecewise continuous, therefore integrable, over \( M_i \).

Claim 1: \( t^*_i \) is p.diff over \( M \).

Proof of Claim 1: Recall that \( t^*_i(m) = -v_i(d(m), m) + \int_{\theta_i}^{m_i} \frac{\partial v_i}{\partial \theta_i}(d(s, m_{-i}), s, m_{-i}) ds \).

Since \( d \) is p.diff, restricted to its pieces, \( \frac{\partial v_i}{\partial \theta_i}(d(\cdot), \cdot) : M \rightarrow \mathbb{R} \) is continuously differentiable over the same pieces as \( v_i \) is twice cont.diff. Therefore \( \int_{\theta_i}^{m_i} \frac{\partial v_i}{\partial \theta_i} \) is p.diff over \( M \), and thus \( t^*_i \) is p.diff over \( M \).

Now, consider a piecewise differentiable B-IC \( t_i \), and we let \( \beta_i := t_i - t^*_i \). Then, by Claim 1, \( \beta_i \) is p.diff over \( M \). Next, since \( t_i \) is B-IC, for all \( \theta_i, b \in B_{\theta_i} \), we have that, when the derivative exists, \( [\partial_i \mathbb{E}^b(v_i(d(m_i, \theta_{-i}), \theta) + t_i(m_i, \theta_{-i}))]_{m_i=\theta_i} = 0. \)

\(^{15}\) For example, consider two agents. The single item allocation rule given by the allocation probabilities \( d_1(\theta) = 1 - d_2(\theta) = \{1 \text{ if } \theta_1 > \theta_2; 1/2 \text{ if } \theta_1 = \theta_2; 0 \text{ otherwise} \} \) satisfies our definition of piecewise differentiability.
Since the canonical transfer \( t^* \) by its construction satisfies the ex-post FOC, the above statement holds for \( t^*_i \) too. Now, from this, for \( t_i - t^*_i \), for all \( \theta_i \) and \( b \in B_{\theta_i} \) for which both derivatives exist, we have \( \left[ \partial_i \mathbb{E}^b (t_i - t^*_i)(m_i) \right]_{m_i = \theta_i} = 0 \). Next, we use the following claim to extend this result to all differentiability points of \( \mathbb{E}^b \beta_i \), beyond the joint differentiability points of \( \mathbb{E}^b t_i \) and \( \mathbb{E}^b t^*_i \). □

**Claim 2:** For a p.diff \( f : M \to \mathbb{R} \) and \( b \in \Delta (\Theta_{\neg i}) \) with p.diff cdf, \( \mathbb{E}^b f : M_i \to \mathbb{R} \) is p.diff.

**Proof of Claim 2:** Consider \( b \)'s cdf, which has finitely many pieces: \( S^b_1, \ldots, S^b_K \).

Write \( \mathbb{E}^b f (m_i) = \int_{\Theta_{\neg i}} f (m_i, \theta_{\neg i}) \, db = \sum_{j=1}^K \int_{\text{int } S^b_j} f (m_i, \theta_{\neg i}) \, db \). For each \( j \), let \( A_j (m_i) := \int_{\text{int } S^b_j} f (m_i, \theta_{\neg i}) \, db \). Since \( f \) is p.diff over \( M \), it is p.diff over each \( S^b_j \) and it has finitely many pieces of \( S^b_j, S^b_{j,1}, \ldots, S^b_{j,J}, \ldots, S^b_{j,L_j} \). Rewrite \( A_j \) such that \( A_j (m_i) = \sum_{l=1}^{L_j} \int_{\text{int } S^b_{j,l}} f (m_i, \theta_{\neg i}) \, db \), and note that \( f \) is continuous over \( \text{int } S^b_{j,l} \).

Therefore \( A_j : M_i \to \mathbb{R} \) is p.diff over \( M_i \) for each \( j \). Since \( \mathbb{E}^b f \) is a sum of \( K \) such functions, it is p.diff over \( M_i \) (that is, it has at most finitely many jumps). □

Note that by Claim 2, if \( b \) has a p.diff cdf, then \( \mathbb{E}^b \nu_i \) is p.diff and thus \( \mathbb{E}^b t^*_i \) is p.diff, which also means that \( \mathbb{E}^b (t_i - t^*_i) \) is p.diff, moreover, it is differentiable in the joint differentiability points of \( \mathbb{E}^b t_i \) and \( \mathbb{E}^b t^*_i \), that is, over \( M_i \) with the exception of at most finitely many points. Therefore, if \( \mathbb{E}^b \beta_i (\cdot) \) has further differentiability points, then the expected value condition must extend to these as well, and hence the Theorem follows.

**Remark.** As this is clear from the last part of the proof above, for a belief \( b \in B_{\theta_i} \) which has a p.diff cdf,\(^{16} \mathbb{E}^b \beta_i \) is almost everywhere differentiable on \( M_i \).

Thus the expected value condition of Theorem 1, for typically considered belief-restrictions, implies substantial restrictions on what form the function \( \beta_i \) can take.

**Proof of Corollary 1.** By Theorem 1, for every \( b \in \Delta (\Theta_{\neg i}) \), at each point of differentiability, \( \partial_i \mathbb{E}^b \beta_i (m_{\neg i}, \theta_{\neg i}) = 0 \). In particular, this holds for all point-beliefs, and thus for all fixed \( m_{\neg i} \), in all points of differentiability of \( \beta_i (\cdot, m_{\neg i}) \), we have \( \partial_i \beta_i (m_{\neg i}, \theta_{\neg i}) = 0 \). Thus for each fixed \( m_{\neg i} \), the function \( \beta_i (\cdot, m_{\neg i}) \) can jump at

\(^{16}\)Note that for example, discrete distributions, full support continuous distributions, as well as their convex combinations have piecewise differentiable cdfs and are Borel-measures.
most finitely many times, and on its pieces, the derivative is 0, therefore on its
pieces, it must be constant. However, if it had a jumping point, then by the
smoothness properties of \(v_i\), it would violate incentive compatibility. Therefore
\(\beta_i\) must be constant everywhere in \(m_i\). □

Proof of Corollary 2. Let \(B^\circ\) be a Bayesian environment with independent types,
and note that by independence the belief does not change with the type, so
let \(b^\circ_i \in \Delta(\Theta_{-i})\) denote agent \(i\)'s beliefs, regardless of his type. First, recall that
\(\mathbb{E}^B[\beta_i(\cdot, \theta_{-i})]\) is a function over \(M_i\) that can jump at most finitely many times. In
its points of differentiability, the derivative is 0, thus the function is constant. If
the function itself would jump, it would violate incentive compatibility, hence it
is the same constant \(\kappa_i\) over \(M_i\), which proves (1) of this corollary. By the charac-
terization in Theorem 1, (2) and (3) follow. □

Proof of Corollary 3. The proof of Corollary 2 applies to belief \(p_i \in \cap_{\theta_i \in \Theta_i} \Delta(\Theta_{-i})\).

Proof of Theorem 2. By the assumed differentiability, \(\beta_i\) is also twice contin-
uously differentiable and as the functions have compact domains, by the Leibniz
rule, (1) obtains from Theorem 1. Further, under \(t_i\), reporting \(\theta_i\) is locally optimal
and thus (2) obtains from the decomposition of the payoff function into \(U_i^*\) and
\(\beta_i\). In the other direction, if (2) holds strictly for all \(m_i\), then the expected payoff
function is strictly concave, and by the decomposition and (1), the FOC holds at
\(\theta_i\), hence \(t_i\) is \(B\)-IC. □

Characterization of Belief-based Terms in Ex. 2. Claim: Consider the belief-
restrictions \(B^\gamma\); for all \(i \in \{1, 2\}\) and for all \(\theta_i\), \(B^\gamma_{\theta_i} = \{b \in \Delta(\Theta_j) : \mathbb{E}^b\theta_j = \gamma_i\theta_i\}\). In the
special case of \(\gamma_i = 1/2\), this is the setting considered in Ex. 2. Recall that \(\theta_i \in [0, 1]\)
and we assume that \(0 < \gamma_i < 1\). Then a function \(\beta_i : M \rightarrow \mathbb{R}\) which is differentiable
in \(m_i\) is a belief-based term if and only if for some real functions \(H_i\) on \(M\) and \(\tau_i\)
on \(M_{-i}\), it takes the form
\[
\beta_i(m) = \int_{m_i}^0 \left(s - \frac{m_j}{\gamma_i}\right) H_i(s) ds + \tau_i(m_{-i}).
\]

Proof of the Claim. First, if \(\beta_i\) is of the given form, then
\[
\partial_i \beta_i(m_i, m_j) = \left(m_i - \frac{m_j}{\gamma_i}\right) H_i'(m_i)
\]
which for all \(\theta_i\), at the truth-telling profile for all beliefs in \(B_{\theta_i}\), satisfies the ex-
pected value condition, thus it is a belief-based term. Second, in the other di-
rection, if \(\beta_i\) is a differentiable belief-based term, then by the point-beliefs in
$B^\gamma_{\theta_i}$, we have that (i) $\partial_i \beta_i (\theta_i; \gamma_i \theta_i) = 0$ for all $\theta_i$. Next, we show that $\partial_i \beta_i : M \to \mathbb{R}$ is linear in $m_j$. This is so, as $B^\gamma_{\theta_i}$ contains beliefs that place non-zero probabilities on two points $x$ and $y$ which give a splitting of $\gamma_i \theta_i$: there is a probability $\alpha$ such that $\alpha x + (1 - \alpha) y = \gamma_i \theta_i$. Note that such $\alpha$ exists for any points that are such that $x \leq \gamma_i \theta_i \leq y$. Each of these beliefs imply, by the expected value condition, that $\alpha \partial_i \beta_i (\theta_i, x) + (1 - \alpha) \partial_i \beta_i (\theta_i, y) = 0$ as well. Hence for any fixed $m_i$, $\partial_i \beta_i$ is linear in $m_j$. Hence, there are functions $f_1$ and $f_2$ on $M_i$ for which $\partial_i \beta_i (m) = f_1 (m_i) m_j + f_2 (m_i)$. At the same time, as by (i) above, these functions must be such that for all $\theta_i$, $f_1 (\theta_i) \gamma_i \theta_i + f_2 (\theta_i) = 0$. From this and by change of notation for the functions, $\beta_i (m)$ has the form as claimed. Finally, the initial condition of "0 type pays 0" of this example implies that $\tau_i \equiv 0$ and so $\beta_i$ takes the form as stated in Ex. 2. □

**Proof of Theorem 3.** The payoffs $U_i = v_i + t^*_i + \beta_i$, by using (3) and adding and subtracting $\int_{m_i}^{\theta_i} \beta_i (\theta_i, m_i)^i d (s, m_i) s) ds + \beta_i (\theta_i, m_i) - \beta_i (\theta_i)$. Hence, these beliefs imply, by the expected value condition, that $\alpha \partial_i \beta_i (\theta_i, x) + (1 - \alpha) \partial_i \beta_i (\theta_i, y) = 0$ as well. Hence for any fixed $m_i$, $\partial_i \beta_i$ is linear in $m_j$. Hence, there are functions $f_1$ and $f_2$ on $M_i$ for which $\partial_i \beta_i (m) = f_1 (m_i) m_j + f_2 (m_i)$. At the same time, as by (i) above, these functions must be such that for all $\theta_i$, $f_1 (\theta_i) \gamma_i \theta_i + f_2 (\theta_i) = 0$. From this and by change of notation for the functions, $\beta_i (m)$ has the form as claimed. Finally, the initial condition of "0 type pays 0" of this example implies that $\tau_i \equiv 0$ and so $\beta_i$ takes the form as stated in Ex. 2. □

**Proof of Proposition 3.** Fix agent $i$. It can be shown, by generalizing the Claim used in the Characterization of Belief-based terms in Ex. 2., that if $B$ is maximal with respect to $(L_i, f_i)_{i \in I}$, then any belief-based term $\beta_i$ satisfies the necessary condition of Theorem 1 if and only if $\partial_i \beta_i = (L_i (m_i) - f_i (m_i)) H_i (m_i)$, where $H_i$ is a real function over $M_i$. Then, if $t_i$ is $B$-IC, by Theorem 1, it can be written as,

$$t_i (m) = t^*_i (m) + \int_{m_i}^{m_i} (L_i (m_i) - f_i (s)) H_i (s) ds + \tau_i (m_i)$$
Next, we need to check when the SOCP at the truthful profile holds. To this end, we need to study when it is the case that for all \( b_{\theta_i} \in B_{\theta_i} \),

\[
\partial^2_{ij} E^{b_{\theta_i}} U_i^* (m_i, \theta_{-i}, \theta) \bigg|_{m_i=\theta_i} + \partial^2_{ij} E^{b_{\theta_i}} \beta_i (m_i, \theta_{-i}) \bigg|_{m_i=\theta_i} \leq 0
\]

\[
- E^{b_{\theta_i}} \left( \frac{\partial^2 v_i (d (\theta), \theta) \partial d (\theta)}{\partial x \partial \theta_i} \right) \leq f'_i (\theta_i) H_i (\theta_i)
\]

Let us set

\[
SCM_i (\theta_i) := \sup_{b_{\theta_i} \in B_{\theta_i}} E^{b_{\theta_i}} \left( - \frac{\partial^2 v_i (d (\theta), \theta) \partial d (\theta)}{\partial x \partial \theta_i} \right).
\]

With this notation, if \( f'_i > 0 \), then \( SCM_i / f'_i \) is a lower bound on \( H_i \) and if \( f'_i < 0 \), then \( SCM_i / f'_i \) is an upper bound on \( H_i \). Next, consider the modification of the interim payments and notice that the order of integration can be exchanged:

\[
E^{b_{\theta_i}} \beta_i (\theta) = E^{b_{\theta_i}} \int_{t_{\theta_i}} E^{b_{\theta_i}} (L_i (\theta_{-i}) - f_i (s)) H_i (s) \, ds
\]

\[
= \int_{t_{\theta_i}} (E^{b_{\theta_i}} L_i (\theta_{-i}) - f_i (s)) H_i (s) \, ds = \int_{t_{\theta_i}} (f_i (\theta_i) - f_i (s)) H_i (s) \, ds.
\]

First, if \( f'_i > 0 \), then the weights on \( H_i \) are positive, and the lower bound on \( H_i \) gives a lower bound on the second term. Therefore \( E^{b_{\theta_i}} \beta_i (\theta) \geq \int_{t_{\theta_i}} (f_i (\theta_i) - f_i (s)) [SCM_i / f'_i] (s) \, ds \).

Second, if \( f'_i < 0 \), then the upper bound on \( H_i \) gives a lower bound on the second term, hence, in this case too, the same inequality holds.

**Proof of Proposition 4.** By way of contradiction, assume that \( t \) is \( B \)-IC and extracts the surplus. By Theorem 1, \( t_i \) can be written as \( t_i (m) = t^*_i (m) + \int_{m_i}^{m_i} (L_i (m_{-i}) - 2f_i (s)) H_i (s) \, ds \) \( \tau_i (m_{-i}) \). Moreover, for all \( \theta_i \) and \( b \in B_{\theta_i} \), \( E^{b} U_i^* (\theta; \theta) = 0 \). Using the formula in 3, and the calculation for \( E^{b_{\theta_i}} \int_{t_{\theta_i}} (L_i (\theta_{-i}) - f_i (s)) H_i (s) \, ds = \int_{t_{\theta_i}} (f_i (\theta_i) - f_i (s)) H_i (s) \, ds \)

\[
\left( \frac{\partial^2 v_i (\theta, m (m))}{\partial \theta_i \partial m_i} - \frac{\partial^2 v_i (\theta, m (m))}{\partial \theta_i \partial m_i} \right) \frac{\partial^2 d (\theta, m (m))}{\partial m_i \partial \theta_i} + \frac{\partial^2 d (\theta, m (m))}{\partial \theta_i \partial m_i}
\]
as in the Proof of Prop. 3, these imply that
\[
\mathbb{E}^b \left(\int_{\theta_i}^{\theta_i} \frac{\partial v_i}{\partial \theta_i} (d(s, \theta_{-i}) s, \theta_{-i}) \, ds + \tau_i (\theta_{-i})\right) = - \int_{\theta_i}^{\theta_i} \left(f_i (\theta_i) - f_i (s)\right) H_i (s) \, ds.
\]

Note that the RHS of this expression depends on \(\theta_i\) but not on \(b\), therefore the LHS must also be the same for all \(b \in B_{\theta_i}\). By \(B\) being maximal wrt \((L_i, f_i)_{i \in I}\), by the generalization of the proof of the Characterization of the Belief Based Terms in Ex. 2, we have that the function \(\int_{\theta_i}^{\theta_i} \frac{\partial v_i}{\partial \theta_i} (d(s, \theta_{-i}) s, \theta_{-i}) \, ds + \tau_i (\theta_{-i})\) must take a form which is \(L_i\)-linear. This function is differentiable in \(\theta_i\) and so, its derivative \(\frac{\partial v_i}{\partial \theta_i} (d(\theta), \theta)\) must also be \(L_i\)-linear. In summary, unless \(\frac{\partial v_i}{\partial \theta_i} (d(\theta), \theta)\) is \(L_i\)-linear, \(B\)-IC and FSE lead to a contradiction. ■

**Proof of Proposition 5.** Fix \((v, d)\). The first inequality follows from the relaxed robustness requirement. The rest of the proposition requires the construction of the two belief-restrictions \(B\) and \(B'\). Note that for each \(i\), there is a function \(L_i : M_{-i} \to \mathbb{R}\) such that \(\frac{\partial v_i}{\partial \theta_i} (d(\theta), \theta)\) is not \(L_i\)-linear. For each \(i\) fix \(\gamma_i \in (0, 1)\), and let the belief-restrictions \(B\) be maximal with respect to the responsive moment condition \((L_i, \gamma_i \theta_i)_{i \in I}\). Prop. 1 implies that \(B\)-IC transfers exist, thus \(F (B)\) is non-empty and \(\infty > \tau (B)\). Yet, as a consequence of Prop. 4, FSE is not possible, that is, \(\tau (B) > 0\). Next, let \(B'\) be s.t. \(B'_{\theta_i} = \{p_{\theta_i}\}\) and s.t. (i) \(p_{\theta_i}\) has a pdf that is continuous and non-zero over the support \(\times j \neq i \left[\bar{\theta}_j, \underline{\theta}_j + (\theta_i - \theta_i) (l_i / l_i)\right]\), where for each \(i, l_i := \bar{\theta}_i - \underline{\theta}_i\), and (ii) for all \(\theta_i\), \(\mathbb{P}^{p_{\theta_i}} L_i (\theta_{-i}) = \gamma_i \theta_i\). (Note that for each \(\theta_i\), matching the fixed first moment is possible.) For \(B'\) thus constructed, the construction in Ex. 3 shows that a \(t\) exists which ensured FSE and is \(B\)-IC and hence \(B'\)-IC as well. ■

**Proof of Theorem 4.** Consider the payoff equation of the Proof of Theorem 3. By setting \(m_i = \theta_i\), the theorem follows. ■

**Appendix B: On Example 3: Beliefs and the Inverse Problem**

Consider an agent with type \(\theta_i\) and beliefs given such that \(\theta_j | \theta_i = \gamma \nu_{\theta_i} + (1 - \gamma) \eta_{ij}\) where \(\nu_{\theta_i}\) is \(U [0, \theta_i]\) and, independently of this, \(\eta_{ij}\) is \(U [0, 1]\). Let us examine the solvability of \(\int_0^1 \alpha_i (\theta_j) p (\theta_j | \theta_i) d\theta_j = f (\theta_i)\). (For a thorough mathematical treatment on the solvability of integral equations we recommend the book Hochstadt (1989).) The pdf of the conditional random variable is such that:
if $1 - \gamma > \gamma \theta_i$, 

$$p (\theta_j | \theta_i) = \begin{cases} \frac{1}{\gamma \theta_i (1 - \gamma)} \theta_j & \text{if } \theta_j \in (0, \gamma \theta_i) \\ \frac{1}{1 - \gamma} & \text{if } \theta_j \in [\gamma \theta_i, 1 - \gamma) \\ \frac{1 - \gamma + \theta_i - \theta_j}{\gamma \theta_i (1 - \gamma)} & \text{if } \theta_j \in [1 - \gamma, 1 - \gamma + \gamma \theta_i) \\ 0 & \text{otherwise} \end{cases}$$

and if $1 - \gamma < \gamma \theta_i$

$$p (\theta_j | \theta_i) = \begin{cases} \frac{1}{(1 - \gamma) \gamma \theta_i} \theta_j & \text{if } \theta_j \in (0, 1 - \gamma) \\ \frac{1}{\gamma \theta_i} & \text{if } \theta_j \in [1 - \gamma, \gamma \theta_i) \\ \frac{1 - \gamma + \theta_i - \theta_j}{(1 - \gamma) \gamma \theta_i} & \text{if } \theta_j \in [\gamma \theta_i, 1 - \gamma + \gamma \theta_i) \\ 0 & \text{otherwise} \end{cases}$$

There are two cases to be considered: either $\gamma \leq 1/2$ or $\gamma > 1/2$.

**Part 1:** If $\gamma \leq 1/2$, then for all $\theta_i$, $1 - \gamma > \gamma \theta_i$. Let us look for solutions of the form such that $\alpha_i (\theta_j)$ is 0 outside of $\theta_j \in [0, \gamma]$. In this case, since $\theta_i < \frac{1 - \gamma}{\gamma}$ for all $\theta_i$, $\int_0^1 \alpha_i (\theta_j) p (\theta_j | \theta_i) d\theta_j = f (\theta_i)$ can be written as

$$\int_0^{\gamma \theta_i} \alpha (\theta_j) \frac{\theta_j}{(1 - \gamma) \gamma \theta_i} d\theta_j + \int_{\gamma \theta_i}^\gamma \alpha (\theta_j) \frac{1}{1 - \gamma} d\theta_j = f (\theta_i).$$

Starting from this expression, in the following three lines, (1) we change variable to $s := \gamma \theta_i$ and differentiate and simplify, (2) reorganize and differentiate for a second time, (3) reorganize:

$$\int_0^s \alpha (\theta_j) \frac{-\theta_j (1 - \gamma)}{(1 - \gamma)^2 s^2} d\theta_j = f' \left( \frac{s}{\gamma} \right) \frac{1}{\gamma}$$

$$\alpha (s) s = - (1 - \gamma) \left( f'' \left( \frac{s}{\gamma} \right) \frac{s^2}{\gamma} + 2 f' \left( \frac{s}{\gamma} \right) \frac{s}{\gamma} \right)$$

$$\alpha (s) = - (1 - \gamma) \left( f'' \left( \frac{s}{\gamma} \right) \frac{s}{\gamma} + 2 f' \left( \frac{s}{\gamma} \right) \frac{1}{\gamma} \right).$$
to, finally, introduce notation $L_\gamma(s) := f''\left(\frac{s}{\gamma}\right)\frac{s}{\gamma} + 2f'\left(\frac{s}{\gamma}\right)\frac{1}{\gamma}$ and change variables to get the solution which is: for all $\theta_j \in [0, \gamma]$, $\alpha(\theta_j) = -(1 - \gamma) L_\gamma(\theta_j)$, and 0 otherwise.18

**Part 2:** If $\gamma > 1/2$, then there are two cases to be considered: either $1 - \gamma > \gamma \theta_i$ or $1 - \gamma \leq \gamma \theta_i$. Eitherways, let us look for solutions of the form such that $\alpha_i(\theta_j)$ is 0 outside of $[\gamma, 1]$.

**Case (A):** $1 - \gamma > \gamma \theta_i$. In this case, $\int_0^1 \alpha_i(\theta_j) p(\theta_j | \theta_i) \, d\theta_j = f(\theta_i)$ can be written as

$$\int_\gamma^{1-\gamma+\theta_i} \frac{1 - \gamma + \gamma \theta_i - \theta_j}{(1 - \gamma) \gamma \theta_i} \alpha(\theta_j) \, d\theta_j = f(\theta_i).$$

Starting from this expression, we change variable to $s := \gamma \theta_i$ and simplify and differentiate, differentiate for a second time,

$$\alpha(1 - \gamma + s) = (1 - \gamma) \left( f''\left(\frac{s}{\gamma}\right) s \frac{1}{\gamma} + 2f'\left(\frac{s}{\gamma}\right)\frac{1}{\gamma} \right),$$

to, finally, change variables, use the notation $L_\gamma$ and get the solution which is: for all $\theta_j \in [\gamma, 1]$, $\alpha(\theta_j) = -(1 - \gamma) L_\gamma(\theta_j - (1 - \gamma))$, and 0 otherwise.

**Case (B):** $1 - \gamma \leq \gamma \theta_i$. In this case, $\int_0^1 \alpha_i(\theta_j) p(\theta_j | \theta_i) \, d\theta_j = f(\theta_i)$ can be written as

$$\int_\gamma^{\gamma \theta_i} \frac{1}{\gamma \theta_i} \alpha(\theta_j) \, d\theta_j + \int_{\gamma \theta_i}^{1-\gamma+\theta_i} \frac{1 - \gamma + \gamma \theta_i - \theta_j}{(1 - \gamma) \gamma \theta_i} \alpha(\theta_j) \, d\theta_j = f(\theta_i).$$

Starting from this expression, we change variable to $s := \gamma \theta_i$ and simplify and differentiate, differentiate for a second time,

$$\alpha(s) + 0 - \alpha(s) + \int_s^{1-\gamma+s} \frac{1}{1 - \gamma} \alpha(\theta_j) \, d\theta_j = \left( f\left(\frac{s}{\gamma}\right) s \frac{1}{\gamma} + 2f'\left(\frac{s}{\gamma}\right)\frac{1}{\gamma} \right).$$

18Note that $L_\gamma(s) = f\left(\frac{s}{\gamma}\right) s''$. 
Finally, change variables, use the notation $L_\gamma$, and the assumption on the format such that $\alpha(s)$ is 0 for all $s < \gamma$ and get the solution which is: for all $\theta_j \in [\gamma, 1]$, 

$$\alpha(\theta_j) = 0 + (1 - \gamma) L_\gamma (\theta_j - (1 - \gamma)),$$

and 0 otherwise.

In summary, in Part 2, differentiating the integral equation twice implies a unique candidate solution since the solution suggested for Case (B) is the same as in Case (A). The candidate solution, when checked against the domain restrictions, works indeed and hence is the solution of the integral equation. □