Submitted to *Unknown Journal*

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 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 NYU, HKUST, 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', regarldess 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 $\mathcal{B}$-*incentive compatible* ($\mathcal{B}$-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 $\mathcal{B} = ((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$, suppose that a $\mathcal{B}$-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 $\mathcal{B}$. Hence, if an additive modification of the canonical transfers yields a $\mathcal{B}$-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 $\mathcal{B}$, any $\mathcal{B}$-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 θ_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 $\mathcal{B}$-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 contstant. 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: $\mathbb{E}^{b_{\theta_i}}\left[\frac{\partial^2\beta_i}{\partial^2 m_i}(\theta_i,\theta_{-i})\right] \leq -\mathbb{E}^{b_{\theta_i}}\left[\frac{\partial^2 U_i^*}{\partial^2 m_i}(\theta_i,\theta_{-i})\right]$ for all θ_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; **?**, 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).

$d : \Theta \to X$ 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 generalized 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 minimal 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). Information 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 $\mathcal{B}$ is a Bayesian setting with correlated types, then FSE obtains. In fact, under a novel 'full rank' condition, we provide the following 'any*thing* 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) extended the result to a continuum of types and to general mechanism design problems. Their condition 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 ϵ. Our condition, in contrast, ensures *exact* FSE. It is stronger than McAfee and Reny's, but closer in spirit to Crémer and McLean (1985, 1988)'s *full rank* condition.

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.[4] 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.

---

[4] The 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 $I = \{1, ..., n\}$ denote the (finite) set of agents, $X \subseteq \mathbb{R}^m$ the set of allocations. For each $i \in I$, we let Θ_i denote the set of player i's payoff types, with typical element θ_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 $v_i : X \times \Theta \to \mathbb{R}$. 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 $(x, t) \in X \times \mathbb{R}^n$, 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 $d : \Theta \to X$, 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) $\Theta_i := [\underline{\theta}_i, \overline{\theta}_i] \subset \mathbb{R}$

(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 assignement

[5]We say that $f : S \to \mathbb{R}$ is *piecewise differentiable* on a closed and convex set $S \subset \mathbb{R}^n$ if there exist a collection $(S_k)_{k=1,...,K}$ of pairwise disjoint convex sets such that $\cup_{k=1}^K S_k = S$, and continuously differentiable functions $g_k : S \to \mathbb{R}$, $k = 1 \ldots K$, such that $f = \sum_{k=1}^K f_k$ where, for each $k = 1, ..., K$, $f_k(x) = \mathbf{1}_{[x \in S_k]} \cdot g_k(x)$.

problems. The main substantial restriction is the one-dimensionality of the payoff types.[6]

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 Θ_{-i}, which represent beliefs about the opponents' types. A belief restriction is 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 $\mathcal{B} = ((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$, such that, $B_{\theta_i} \subseteq \Delta(\Theta_{-i})$ is non-empty for each i and θ_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 $\mathcal{B}^{BF} = ((B^{BF}_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$.

(ii) The standard *Bayesian settings* are obtained if the belief restrictions are commonly known and each belief set is a singleton for every type: $B^{\diamond}_{\theta_i} = \{b^{\diamond}_{\theta_i}\}$ 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 $\theta_i, p(\cdot|\theta_i) = b^{\diamond}_{\theta_i} \in \Delta(\Theta_{-i})$. If, furthermore, such a common prior is *independent* across agents, then we also have $b^{\diamond}_{\theta_i} = b^{\diamond}_{\theta'_i}$ 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 θ_i. Special cases have been considered, for instance, by Ollár and Penta (2017) and Ollár and Penta (2023) to model that agents commonly know some moments of the distributions of the opponents' types (*common knowledge of moment conditions*) and that agents commonly believe that the opponents' types are identically distributed (*common belief in identicality*), respectively. The latter belief restrictions, which we denote as $\mathcal{B}^{id} = ((B^{id}_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$, are defined for

[6] It is well known that incentive compatibility is significantly more problematic outside of this domain, as multidimensionality of types severaly limits its possibility (Jehiel and Moldovanu (2001) and Jehiel et al. (2006)). We extend our approach to the multidimensional case in Ollár and Penta (2024a).

settings with a common set of types (i.e. $\Theta_j = \Theta_k$ for all $j, k \in I$) as follows: $B^{id}_{\theta_i} = \{b_{\theta_i} \in \Delta(\Theta_{-i}) : \text{marg}_{\Theta_j} b_{\theta_i} = \text{marg}_{\Theta_k} b_{\theta_i} \text{ for all } j, k \neq i\}$ for all i and θ_i.

These are examples of special instances from the mechanism design literature, but the framework is more general. We 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: they depend only on the *first-order beliefs*.

Given belief restrictions $\mathcal{B} = ((B_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$ and $\mathcal{B}' = ((B'_{\theta_i})_{\theta_i \in \Theta_i})_{i \in I}$, we write $\mathcal{B} \subseteq \mathcal{B}'$ to denote that $B_{\theta_i} \subseteq B'_{\theta_i}$ for all $i \in I$ and all $\theta_i \in \Theta_i$. If $\mathcal{B} \subseteq \mathcal{B}'$, then $\mathcal{B}$ imposes stronger restrictions than $\mathcal{B}'$, in that the designer can rule out more beliefs in the former than in the latter. In this sense, the belief-free model $\mathcal{B}^{BF}$ is minimal in the information that the designer has, as any model $\mathcal{B}$ is such that $\mathcal{B} \subseteq \mathcal{B}^{BF}$. At the opposite extreme, any Bayesian setting $\mathcal{B}^{\diamond}$ is maximal, as no distinct belief restriction $\mathcal{B}$ is such that $\mathcal{B} \subseteq \mathcal{B}^{\diamond}$. Belief restrictions $\mathcal{B}^{id}$ are an example of an intermediate robustness requirement, $\mathcal{B}^{\diamond} \subseteq \mathcal{B}^{id} \subseteq \mathcal{B}^{BF}$.

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 $\mathbb{E}^b U^t_i(m_i; \theta_i) := \int_{\Theta_{-i}} U^t_i(m_i, \theta_{-i}; \theta_i, \theta_{-i})\, db$, and for any $f : \Theta \to \mathbb{R}$, $\theta_i \in \Theta_i$ and $b \in B_{\theta_i}$, we let $\mathbb{E}^b[f(\theta_i, \theta_{-i})] := \int_{\Theta_{-i}} f(\theta_i, \theta_{-i})\, db$.

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 $\mathcal{B}$.

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

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

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

REMARK 1. *If $\mathcal{B} \subseteq \mathcal{B}'$, and (d,t) is $\mathcal{B}'$-IC, then it is also $\mathcal{B}$-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 γ is a parameter of preference interdependence. These preferences satisfy the following *Single-Crossing Conditions*:

$$\textbf{(ep-SCC:)}\ \text{for all } i \text{ and } (x,\theta),\ \frac{\partial^2 v_i}{\partial x \partial \theta_i}(x,\theta) > 0 \tag{1}$$

Agents' types are such that $\theta_i = \theta_0 + \eta_i$, where θ_0 is a (unobserved) common value component, uniformly distributed over $[0,1/2]$, and η_i is an idiosyncratic component, also uniformly distributed over $[0,1/2]$, independently from θ_0 and η_j. Agents only observe θ_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: $\mathbb{E}^{b_{\theta_i}}(\theta_j) = \mathbb{E}(\theta_j|\theta_i) = \theta_i/2 + 1/4$.

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:

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

$$U_i^{VCG}(m,\theta) = (1+\gamma)(m_i + m_j)(\theta_i + \gamma\theta_j) - (1+\gamma)\left(\frac{1}{2}m_i^2 + \gamma m_i m_j + \gamma m_j^2\right)$$

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,

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

which induce the following payoff functions:

$$\begin{aligned} U_i^{mod}(m;\theta) =& U_i^{VCG}(m;\theta) + (1+\gamma)\left(m_i^2 + m_i - 4m_i m_j\right) = \\ &= (1+\gamma)\left(\left((\theta_i + \gamma\theta_j) - (m_i + \gamma m_j)\right)(m_i + m_j) + \frac{3}{2}m_i^2 + m_i - 3m_i m_j\right). \end{aligned}$$

Taking the first order conditions from the interim payoff function, and evaluating it at the truthful profile, we obtain:

$$\begin{aligned}\frac{\partial \mathbb{E}^{b_{\theta_i}}[U_i^{mod}(\theta;\theta)]}{\partial m_i} =& \mathbb{E}^{b_{\theta_i}}\Bigg((1+\gamma)(2\theta_i+1-4\theta_j)\Bigg)\\ =& (1+\gamma)\Big(2\theta_i+1-4\mathbb{E}^{b_{\theta_i}}(\theta_j|\theta_i)\Big)=0.\end{aligned}$$

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 $\mathbb{E}^{b_{\theta_i}}(\theta_j)=\mathbb{E}(\theta_j|\theta_i)=\theta_i/2+1/4$ for all θ_i. To check the second order conditions, since $\gamma=-2$, we have $\frac{\partial^2 U_i^{mod}}{\partial^2 m_i}(m;\theta)=-1<0$. 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 $t_i^{mod}(m)=t_i^{VCG}(m)+\beta_i(m)$, where $\beta_i: M\to\mathbb{R}$ 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 θ_i and $b_{\theta_i}\in B_{\theta_i}$. Theorem 1 in Section 3 shows that this holds in general: for any belief-restrictions $\mathcal{B}$, any $\mathcal{B}$-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_{θ_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 α but fail for others.

EXAMPLE 1 (continued): In the setting of Ex. 1, consider transfers of the form $t_i^{mod,\alpha}(m)=t_i^{VCG}(m)+\alpha_i(1+\gamma)(m_i^2+m_i-4m_im_j)$. 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 $\mathbb{E}^b\left[\frac{\partial^2 U_i^{VCG}}{\partial^2 m_i}(\theta_i,\theta_{-i})\right] < -\mathbb{E}^b\left[\frac{\partial^2 \beta_i}{\partial^2 m_i}(\theta_i,\theta_{-i})\right]$ for all θ_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 $\mathcal{B}$-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,

$$t_i^*(m) = -v_i(d(m), m) + \int_{\underline{\theta}_i}^{m_i} \frac{\partial v_i}{\partial \theta_i}(d(s_i, m_{-i}), s_i, m_{-i})\, ds_i. \tag{3}$$

These transfers are pinned down by the necessary conditions for ep-IC, up to an additive term that is constant in own report.[7] This characterization of the ep-IC transfers can be obtained both by inverting the *envelope formula* for the ex-post payoff function (Milgrom and Segal, 2002), or directly from the *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 $\mathcal{B}$-IC transfers based on a first-order approach. An envelope formulation is discussed in Section 5.2.

3.1 *A first-order approach*

The main result in this section derives necessary conditions for $\mathcal{B}$-IC transfers, for general belief restrictions. In our result, we provide a generalization of the classical *first-order approach* that identifies necessary conditions for *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 θ_i, for all beliefs in the set B_{θ_i}.

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

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

where $\beta_i : M \to \mathbb{R}$ is piecewise differentiable and such that, for all θ_i and for all beliefs $b \in B_{\theta_i}$ that have a piecewise differentiable pdf, at all points of differentiability,

---

[7] The 'canonical transfers', and the associated *canonical direct mechanism* (d,t^*), should not be confused with the 'canonical mechanism', which traditionally refers to Maskin's (non-direct) mechanism for *full* implementation. Special instances of the canonical direct mechanism have appeared throughout the literature on *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)).

$$\left.\frac{\partial \mathbb{E}^{b}\left[\beta_i\left(m_i, \theta_{-i}\right)\right]}{\partial m_i}\right|_{m_i=\theta_i}=0. \tag{5}$$

The result in Equation (4) shows that, in order to design a $\mathcal{B}$-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* $\mathcal{B}$ (or simply *belief-based terms*, when $\mathcal{B}$ 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, $\mathcal{B}^{BF}$, the condition in (5) is required to hold for all beliefs about Θ_{-i}, including degenerate ones, which is only possible if β_i is constant in m_i. Hence, a transfer scheme is $\mathcal{B}^{BF}$-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 $\mathcal{B}^{BF}$-IC, then, $\forall i$, $\beta_i\left(m\right) := t_i(m) - t_i^*(m)$ is constant in m_i.*

3.2.2 *Bayesian Settings* In a *Bayesian setting*, $\mathcal{B}^{\diamond}$, 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 $f_i\left(\theta_i\right) := \mathbb{E}^{b^{\diamond}_{\theta_i}} G_i\left(\theta_i, \theta_{-i}\right)$ is uniquely pinned down by the collection $(b^{\diamond}_{\theta_i})_{\theta_i \in \Theta_i}$ of agent i's beliefs. Hence, letting

$$\beta_i\left(m\right) := \int_{\underline{\theta}_i}^{m_i} G_i\left(s, m_{-i}\right) ds - \int_{\underline{\theta}_i}^{m_i} f_i\left(s\right) ds,$$

we obtain a belief-based term, since β_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, $\mathcal{B}^{\diamond}_{\theta_i} = \{b^{\diamond}_i\}$ 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 $\mathcal{B}^{\diamond}$ be a Bayesian environment with independent types, and let $b^{\diamond}_i \in \Delta(\Theta_{-i})$ denote agent i's beliefs, regardless of his type. Then:*

(i) If t_i is $\mathcal{B}^{\diamond}$-IC, then $\exists \kappa_i \in \mathbb{R}$ s.t. $\mathbb{E}^{b^{\diamond}_i}\beta_i(m_i, \theta_{-i}) = \kappa_i$ for all m_i.

(ii) If t_i is $\mathcal{B}^{\diamond}$-IC, $\exists \kappa_i \in \mathbb{R}$ s.t. $\mathbb{E}^{b^{\diamond}_i} t_i(\theta_i, \theta_{-i}) = \mathbb{E}^{b^{\diamond}_i} t^*_i(\theta_i, \theta_{-i}) + \kappa_i$ for all θ_i.

(iii) (d, t) is $\mathcal{B}^{\diamond}$-IC for some t if and only if (d, t^*) is $\mathcal{B}^{\diamond}$-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^* 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^* 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 $\theta_i, \theta_i' \in \Theta_i$), which in turn holds in any of the following special cases: *belief-free* settings; Bayesian models with *independent types*; the $\mathcal{B}^{id}$-restrictions, for *common belief in identicality*. With this for each i we obtain the following:

COROLLARY 3. *Let $\mathcal{B}$ satisfy* generalized indepence*, 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 t_i is $\mathcal{B}$-IC, then $\exists \kappa_i \in \mathbb{R}$ s.t. $\mathbb{E}^{p_i} t_i (\theta_i, \theta_{-i}) = \mathbb{E}^{p_i} t_i^* (\theta_i, \theta_{-i}) + \kappa_i$ for all θ_i.

(iii) Assume $B_{\theta_i} = B_{\theta_i'}$ for all θ_i, θ_i'. Then, there exists a $\mathcal{B}$-IC t_i if and only if t_i^* 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 the beliefs in the intersection in such non-Bayesian settings as well. These results follow directly from Theorem 1.[8]

[8] 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 interim IC (wrt. the

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 *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 $\mathcal{B}$-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).

4.1 *$\mathcal{B}$-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 $\mathcal{B}$-IC:

THEOREM 2 (Conditions under Differentiability). *Assume that v_i, t_i, d are all twice differentiable, and for each i, let $\beta_i := t_i - t_i^*$.*

overlapping beliefs). As we explain in Section 5.1.3 such an equivalence of $\mathcal{B}$-IC and ep-IC follows, using the characterization in Theorem 3, from Corollary 3 (ii)'s application to overlapping beliefs.

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

(i) $\mathbb{E}^b[\partial_i \beta_i(\theta_i, \theta_{-i})] = 0$ *and*

(ii) there exists an open neighborhood of θ_i*,* $\mathcal{N}_{\theta_i}$*, s.t. for all* $m_i \in \mathcal{N}_{\theta_i}$*:*

$$\mathbb{E}^b[\partial_{ii}^2 U_i^*(m_i, \theta_{-i}; \theta_i, \theta_{-i})] \leq -\mathbb{E}^b[\partial_{ii}^2 \beta_i(m_i, \theta_{-i})]. \tag{6}$$

[Sufficiency:]*: Transfers* $t = (t_i)_{i \in I}$ *are* $\mathcal{B}$*-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 nessecary 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, $\partial_i \beta_i(m) = (1+\gamma)(2m_i + 1 - 4m_j)$ and $\partial_{ii}^2 \beta_i(m) = (1+\gamma)2$. The expected payoffs of the canonical transfers instead are such that, for all beliefs consistent with the belief-restrictions, $\partial_{ii}^2 \mathbb{E}^{b_{\theta_i}}[U_i^*(m;\theta)] = -(1+\gamma)$. Hence, β_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 β_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 θ_i and $b_{\theta_i} \in B_{\theta_i}$. □

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 $\mathcal{B}$-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 $\mathcal{B}$-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 θ_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:

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

$$\text{and } t_2^*(m) = +m_2 K(m_1 + m_2) - K\int_0^{m_2} (m_1 + s)\, ds = K\frac{1}{2}m_2^2,$$

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 $\mathcal{B}$-IC. The reason is that while truthful revelation satisfies the F.O.C. for both agents, since the allocation rule moves with θ_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 $\mathcal{B}$-IC transfers, first we identify the set of belief-based terms that satisfy the necessary condition in part 1 of Theorem 2. (We mainitain in this example that the lowest type of each agent always pays 0.) In this setting, $\beta_i : M \to \mathbb{R}$ 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 $M_i \equiv \Theta_i$, see the Appendix. (One direction is easy: for such β_i function, $\partial_i \mathbb{E}^b \beta_i (\theta_i) = 0$.) 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 θ_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:

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

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 θ_1. Similarly, for agent 2 we obtain $H_2 (\theta_2) \leq -K$ for all θ_2. Hence, a transfer scheme is $\mathcal{B}$-IC if and only if it takes the form

$$t_1 (m_1, m_2) = -\frac{1}{2} m_1^2 + \int_0^{m_1} (s - 2m_2) H_1 (s) \, ds, \text{ and}$$
$$t_2 (m_1, m_2) = \frac{1}{2} m_2^2 + \int_0^{m_2} (s - 2m_1) H_2 (s) \, ds,$$

subject to the restriction on the H_i functions above. Exploiting again the fact that, for each θ_i and $b \in B_{\theta_i}$, $\mathbb{E}^b[\theta_j] = \theta_i/2$, the expected transfers at the truth-telling profile are:

$$\mathbb{E}^b[t_1 (\theta) |\theta_1] = -\frac{1}{2} \theta_1^2 + \int_0^{\theta_1} (s - \theta_1) H_1 (s) \, ds, \text{ and}$$

$$\mathbb{E}^b[t_2(\theta)|\theta_2] = \frac{1}{2}\theta_2^2 + \int_0^{\theta_2}(s-\theta_2)H_2(s)\,ds,$$

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, $t_1^{Cmin}(m_1,m_2) = \frac{m_1^2}{2}(K-1) - 2Km_2m_1$, and $t_2^{Cmin}(m_1,m_2) = \frac{m_2^2}{2}(1-K) + 2Km_1m_2$, 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_{θ_i}. $\square$

4.2 *$\mathcal{B}$-IC transfers in the general case: A Full Characterization*

We provide next a characterization of the $\mathcal{B}$-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 ($\mathcal{B}$-IC: Characterization). *Under the maintained assumptions of Theorem 1, for each i, let $\beta_i := t_i^* - t_i$. Then, (d,t) is $\mathcal{B}$-IC if and only if for all i, θ_i, $b \in B_{\theta_i}$ and m_i:*

$$\mathbb{E}^b\left[\int_{m_i}^{\theta_i}\left(\frac{\partial v_i}{\partial\theta_i}(d(s,\theta_{-i}),s,\theta_{-i}) - \frac{\partial v_i}{\partial\theta_i}(d(m_i,\theta_{-i}),s,\theta_{-i})\right)ds\right] \geq \mathbb{E}^b\Big[\beta_i(m_i,\theta_{-i}) - \beta_i(\theta)\Big].$$

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 θ_i, θ_i' and for all θ_{-i}:*[9]

[9] This Corollary generalizes known results on single-crossing and monotonicity conditions to our setting, which allows for not-everywhere differentiable allocation rules.

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

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:

COROLLARY 5 (IIC with Independent Types). *Let $\mathcal{B}^\diamond$ 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, 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 θ_i, θ_i',*

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

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 $b^\diamond$). 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 $\mathcal{B}$-IC either, for any γ, and Corollary 3 and Theorem 3 jointly imply that no transfers are $\mathcal{B}$-IC when $\gamma = 0$. Next, consider the following transfers:

$$t_2^{mod}(m) = t_2^*(m) - A\left(\frac{\gamma m_2^2/2 + (1-\gamma)m_2}{2} - m_1 m_2\right). \tag{7}$$

Under these belief restrictions, truthful revelation satisfies the first-order conditions, and $\frac{\partial^2 U_2^{mod}(m;\theta)}{\partial^2 m_2} = K - A\gamma/2$. Hence, $m_2 = \theta_2$ is optimal for agent 2 whenever $A > 2K/\gamma$, and hence $\mathcal{B}$-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 $\mathcal{B}$ 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 θ_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, $\mathcal{B}$ is such that, for each i and θ_i, B_{θ_i} consists of all *the beliefs $b \in \Delta(\Theta_{-i})$ such that $\mathbb{E}^b L_i(\theta_{-i}) = f_i(\theta_i)$, then we say that $\mathcal{B}$ is* maximal *with respect to the moment condition $(L_i, f_i)_{i \in I}$.*

In words, $\mathcal{B}$ admits a *moment condition* if, for every i, there exists a function of the opponents' types whose expectation given θ_i is known to the designer (i.e., for each θ_i, it is the same for all beliefs in B_{θ_i}). If such expectations are strictly monotonic in θ_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 oppositve extreme, in a Bayesian setting, for *any* L_i there is a f_i such that $\mathbb{E}^{b_i^\diamond} L_i(\theta_{-i}) = f_i(\theta_i)$ (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 $\mathcal{B}$ 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 $\mathcal{B}$-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 $\mathcal{B}$-IC by some t. (In Ex.3, $L_i(\theta_{-i}) = \theta_j$, and $f_i(\theta_i) = \frac{\gamma\theta_i + (1-\gamma)}{2}$, 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 version of FSE, that highlights more clearly how the difference between Bayesian and non-Bayesian settings affects the design of the mechanism.[11] Our result is based on the following conditions:

DEFINITION 4. *Let $\mathcal{B}^{\diamond}$ be a Bayesian setting (i.e., $B^{\diamond}_{\theta_i} = \{b^{\diamond}_{\theta_i}\}$ for each i and θ_i).*

(i) *We say that $\mathcal{B}^{\diamond}$ 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 $f_i(\theta_i) = \mathbb{E}^{b^{\diamond}_{\theta_i}}[G(\theta_i, \theta_{-i})]$, is differentiable.*

---

[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 that this can be done so that agents' surplus is *almost* fully extracted (cf. footnote 3). Chen and Xiong (2013) further showed that this form of FSE holds generically in the space of Bayesian models. More recent results are provided by Hu et al. (2021) and Lopomo et al. (2022), who consider alternative approaches to FSE.

[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.

(ii) *We say that $\mathcal{B}^{\diamond}$ 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 $\int_{\Theta_{-i}} \kappa_i(\theta_{-i})\, db^{\diamond}_{\theta_i} = g_i(\theta_i)$ for all θ_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 $\mathcal{B}^{\diamond}$ 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 θ_i, $\mathbb{E}^{b^{\diamond}_{\theta_i}}[t'_i(\theta_i, \theta_{-i})] = \mathbb{E}^{b^{\diamond}_{\theta_i}}[t_i(\theta_i, \theta_{-i})]$.*

Proof: First note that if $\mathcal{B}^{\diamond}$ 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 $\hat{t}_i := t^*_i - A_i\left(\int^{m_i} f_i(s)\, ds - L_i(m_{-i})\, m_i\right)$. 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 $g_i(\theta_i) := \int_{\Theta_{-i}} [t_i(\theta_i, s) - \hat{t}_i(\theta_i, s)] db^{\diamond}_{\theta_i}$ and note that, by construction and Def. 4, g_i is differentiable in θ_i. Using the full rank condition, let $\kappa_i : \Theta_{-i} \to \mathbb{R}$ be s.t. $\int_{\Theta_{-i}} \kappa_i(\theta_{-i}) db^{\diamond}_{\theta_i} = g_i(\theta_i)$ for each θ_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 $\mathbb{E}^{b^{\diamond}_{\theta_i}}[t'_i(\theta_i, \theta_{-i})] = \mathbb{E}^{b^{\diamond}_{\theta_i}}[t_i(\theta_i, \theta_{-i})]$. ■

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:*

$$t_i\left(m\right)=t_i^*\left(m\right)+\int_{\underline{\theta}_i}^{m_i}\left(L_i\left(m_{-i}\right)-f_i\left(s\right)\right)H_i\left(s\right)\,ds+\tau_i\left(m_{-i}\right).$$

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

For the next proposition, we say that a function $g : \Theta \to \mathbb{R}$ is L_i-linear if it can be written in the form $g\left(\theta\right)=\delta_1\left(\theta_i\right)L_i\left(\theta_{-i}\right)+\delta_2\left(\theta_i\right)$. Additionally, we say that a mechanism (d, t) is $\mathcal{B}$-*individually rational* ($\mathcal{B}$-IR) if, for each i and θ_i, $\mathbb{E}^bU_i^t(\theta_i;\theta_i)\geq 0$ for all $b\in B_{\theta_i}$.[12] Finally, we say that a mechanism *extracts the full surplus* if the individual rationality constraints hold with equality for all i, θ_i, and $b\in B_{\theta_i}$

PROPOSITION 4. *Fix v and d, and let $\mathcal{B}$ be maximal with respect to a responsive moment condition $(L_i, f_i)_{i\in I}$. Unless for all i, $\frac{\partial v_i}{\partial\theta_i}\left(d\left(\theta\right),\theta\right)$ is L_i-linear, 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 'any*thing* 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 $B_{\theta_i}=\{b\in\Delta(\Theta_j):\mathbb{E}^b[\theta_j]=\gamma\frac{\theta_i}{2}+(1-\gamma)\frac{1}{2}\}$. For simplicity, let us consider the

[12] Recall that, for any $b\in\Delta\left(\Theta_{-i}\right)$, we defined $\mathbb{E}^bU_i^t(m_i;\theta_i):=\int_{\Theta_{-i}}U_i^t\left(m_i,\theta_{-i};\theta_i,\theta_{-i}\right)db$. Also, in this section we set the outside option to 0 for simplicity, but the extension to type-dependent outside options is easy.

case where $\gamma \in [0, 1/2]$. As we already discussed, the conditions of Prop. 1 hold, and $\mathcal{B}$-IC is attained by the transfers in eq. (7), as long as $A > 2K/\gamma$ and for any $\gamma > 0$.

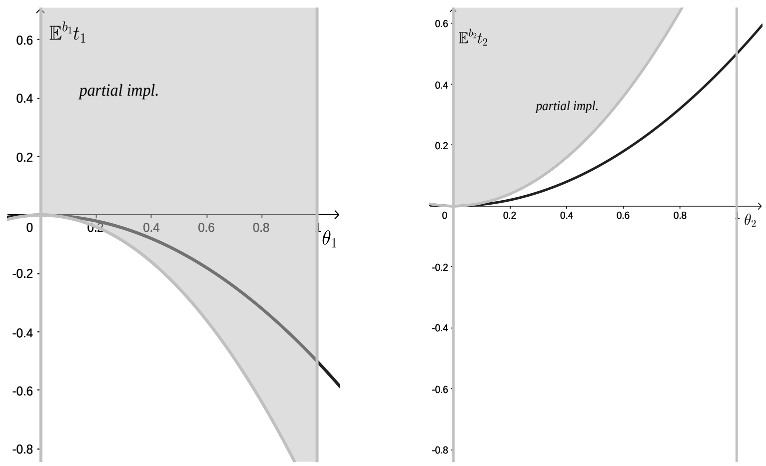

FIGURE 1. **Possible Expected Payments to the Agents in Ex. 3: $\mathcal{B}$-IC under $t_i(0, \theta_{-i}) \equiv 0$.** The thick black line, in both figures, is the expected canonical transfer to each agent (feasible for agent 1 but infeasible for agent 2). The gray area represents the possible interim 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 θ_i, for any $b \in B_{\theta_i}$) that are associated with $\mathcal{B}$-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 $\mathcal{B}^{\diamond}$ be such that, for each θ_i, $b^{\diamond}_{\theta_i}$ consists of a mixture of two independent uniform distributions, over $[0, \theta_i]$ and $[0, 1]$, respectively with weights γ 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 $g_i(\theta_i) = \int_{\Theta_j} \left(t_i - \hat{t}_i\right) db_{\theta_i}$, 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 $\int_{\Theta_2} \kappa_1(\theta_2)\, db_{\theta_1} = -K\left[\gamma \frac{\theta_1^2}{2} + (1-\gamma)\frac{\theta_1}{2}\right]$ 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 $\hat{t}_2(\theta) = t_2^*(\theta) - A\left(\frac{\gamma\theta_2^2/2+(1-\gamma)\theta_2}{2} - \theta_1\theta_2\right)$ from eq. (7), which is IIC for $A > 2K/\gamma$. The integral equation $\int_{\Theta_1} \kappa_2(\theta_1)\, db_{\theta_2} = \frac{\theta_2^2}{2}\left[K(1+\gamma) - \gamma\frac{A}{2}\right] + K(1-\gamma)\frac{\theta_2}{2}$ solved for $\kappa_2(\cdot)$ gives $\kappa_2(\theta_1) = -\frac{(1-\gamma)}{\gamma}\left[\theta_1\frac{(2+\gamma)}{\gamma}\left(K(1+\gamma)-\gamma\frac{A}{2}\right)+(1-\gamma)K\right]$ 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 strenghtens (i.e., as the belief sets get larger).

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

$$\tau(\mathcal{B}) := \inf_{t\in F(\mathcal{B})} \sup_{(i,\theta_i,b)\in\mathcal{V}(\mathcal{B})} \mathbb{E}^b U_i^t(\theta_i;\theta_i)$$

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

First note that, with this notation, FSE obtains if and only if there exists $t \in F(\mathcal{B})$ such that the constraint for $\mathcal{B}$-IR holds with equality for all types of all agents, i.e. if $\tau(\mathcal{B}) = 0$. If $\infty > \tau(\mathcal{B}) > 0$, 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 $\mathcal{B}$ is a Bayesian setting with correlated types, then FSE obtains.

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

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

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*, $\mathcal{B}^{\diamond}$, for any $i \in I$ and for any $G_i : M \to \mathbb{R}$ that is Lebesgue-integrable with respect to m_i, the function $f_i(\theta_i) := \mathbb{E}^{b^{\diamond}_{\theta_i}} G_i(\theta_i, \theta_{-i})$ is uniquely pinned down by agent i's beliefs. Hence, letting $\beta_i(m) := \int_{\underline{\theta}_i}^{m_i} G_i(s, m_{-i})\, ds - \int_{\underline{\theta}_i}^{m_i} f_i(s)\, ds$, we obtain a viable belief-based term, since β_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

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

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:

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

With $\gamma = 3/2$, the VCG transfers are IIC. Furthermore, since $\mathbb{E}^b[\theta_j|\theta_i] = 1/2$ for all θ_i, these modified transfers satisfy both conditions in Theorem 2 for any α_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 (iii) generalizes Theorem 1 in Ollár and Penta (2023), which only focused on the $\mathcal{B}^{id}$-restrictions (i.e., under *common belief in identicality*). Corollary 3 (iii) can be extended to settings where belief restrictions are full dimensional and overlap to shed light on some influential results in Lopomo et al. (2021) and in Jehiel et al. (2012)).

COROLLARY 6. *Let $\mathcal{B}$ be "full dimensional and independent" that is for each i and θ_i (1) B_{θ_i} is full dimensional (2) for all $p_i \in B_{\theta_i}$ $\exists \mathcal{N}_{\theta_i} \subseteq \Theta_i$ open interval such that $p_i \in B_{\theta_i'}$ for all $\theta_i' \in \mathcal{N}_{\theta_i}$. Then, there exists a $\mathcal{B}$-IC t_i if and only if t_i^* is $\mathcal{B}$-IC.*

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. However, Jehiel et al. (2012) gave an example of local robust implementability in a non-standard single-crossing environment, where ex post implementability is impossible.

Under Corollary 6's "full dimensional and independent" beliefs, $\mathcal{B}$-IC is possible if and only if it is achieved by the canonical transfers. Under standard ex-post SCM conditions, the canonical transfers are ep-IC (Corollary 4), and hence our results also imply that – under constant belief restrictions as in Corollary 3 (iii) and under "full dimensional and independent" beliefs as above – there is no gap between the possibility of ep-IC and $\mathcal{B}$-IC. However, it is important to note that without ep-SCC, as in our general setting, the canonical transfers may be $\mathcal{B}$-IC without necessarily being ep-IC.[14] That is, outside of standard single-corssing and monotonicity, $\mathcal{B}$-IC and ep-IC might not coincide, while *revenue equivalence* still holds (Corollary 3 (ii)).

5.2 *Equilibrium Payoffs: An Envelope Formulation*

Theorem 3 implies the following characterization of the equilibrium payoffs of $\mathcal{B}$-IC mechanisms:

THEOREM 4 (Payoff Characterization). *Fix belief restrictions $\mathcal{B}$ and allocation rule d. For each i, let $D_i \subseteq \mathbb{R}^M$ denote the set of all belief-based terms that satisfy the inequalities 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 $\mathcal{B}$-IC mechanism if and only if, for each i, there exists*

[14] Ollár and Penta (2023) provide an example of this possibility within the context of the $\mathcal{B}^{id}$-restrictions.

$\beta_i \in D_i$ *such that*

$$U_i(\theta_i, \theta_{-i}; \theta) = \int_{\underline{\theta_i}}^{\theta_i} \frac{\partial v_i}{\partial \theta_i} (d(s, \theta_{-i}), s, \theta_{-i}) \, ds + \beta_i(\theta_i, \theta_{-i}). \tag{8}$$

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 $\beta_i : M \to \mathbb{R}$ 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 $\mathbb{E}^b(\beta_i)$ is constant in m_i (Corollary 3), and thus the formula in (8) again delivers the standard condition for the interim expected payoffs, $\mathbb{E}^b(U_i)$, here generalized to accommodate both the possibility of interdependent values as well as non-Bayesian settings with *generalized independence*.

Thus, when $\mathbb{E}^b(\beta_i)$ is constant in m_i for all $\beta_i \in D_i$, the interim expected equilibrium payoffs under incentive compatibility are effectively pinned down, up to a constant in own message, and hence the formula gives the incentive compatible transfers as well, by 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 envelopes, one for each distinct belief-dependent term 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 implementation 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 may not be used. But it also brings to the forefront a hiterto neglacted 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 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

Akbarpour, M. and S. Li (2020). Credible auctions: A trilemma. *Econometrica 88*(2), 425–467. [38]

Bergemann, D. and S. Morris (2005). Robust mechanism design. *Econometrica*, 1771–1813. [2, 9, 36, 38]

Bergemann, D. and S. Morris (2009a). Robust implementation in direct mechanisms. *The Review of Economic Studies 76*(4), 1175–1204. [2, 9, 36, 38]

Bergemann, D. and S. Morris (2009b). Robust virtual implementation. *Theoretical Economics 4*(1), 45–88. [2, 9, 36, 38]

Bernstein, S. and E. Winter (2012). Contracting with heterogeneous externalities. *American Economic Journal: Microeconomics 4*(2), 50–76. [5]

Börgers, T. and J. Li (2019). Strategically simple mechanisms. *Econometrica 87*(6), 2003–2035. [38]

Börgers, T. and D. Smith (2012). Robustly ranking mechanisms. *American Economic Review 102*(3), 325–329. [38]

Börgers, T. and D. Smith (2014). Robust mechanism design and dominant strategy voting rules. *Theoretical Economics 9*(2), 339–360. [38]

Carroll, G. (2015). Robustness and linear contracts. *American Economic Review 105*(2), 536–563. [38]

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

Carvajal, J. C. and J. C. Ely (2013). Mechanism design without revenue equivalence. *Journal of Economic Theory 148*, 104–133. [3, 37]

Catonini, E. and A. Penta (2022). Backward induction reasoning beyond backward induction. *TSE Working Paper*. [36]

Chassang, S. (2013). Calibrated incentive contracts. *Econometrica 81*(5), 1935–1971. [38]

Chen, Y.-C. and S. Xiong (2013). Genericity and robustness of full surplus extraction results. *Econometrica 81*(1), 825–847. [27]

Chung, K.-S. and J. C. Ely (2007). Foundations of dominant-strategy mechanisms. *The Review of Economic Studies 74*(2), 447–476. [38]

Crémer, J. and R. P. McLean (1985). Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica 53*(2), 345–361. [2, 5, 6, 27, 39]

Crémer, J. and R. P. McLean (1988). Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, 1247–1257. [2, 5, 6, 7, 27, 39]

Dasgupta, P. and E. Maskin (2000). Efficient auctions. *The Quarterly Journal of Economics 115*(2), 341–388. [15]

d'Aspremont, C. and L.-A. Gérard-Varet (1979). Incentives and incomplete information. *Journal of Public Economics 11*(1), 25–45. [5, 39]

Gagnon-Bartsch, T., M. Pagnozzi, and A. Rosato (2021). Projection of private values in auctions. *American Economic Review 111*(10), 3256–3298. [3, 37]

Gagnon-Bartsch, T. and A. Rosato (2023). Quality is in the eye of the beholder: taste projection in markets with observational learning. *working paper*. [3, 37]

Gavan, M. J. and A. Penta (2023). Safe implementation. *BSE working paper*. [38]

Green, J. and J.-J. Laffont (1977). Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica*, 427–438. [15]

Halac, M., E. Lipnowski, and D. Rappoport (2021). Rank uncertainty in organizations. *American Economic Review 111*(3), 757–786. [5]

Halac, M., E. Lipnowski, and D. Rappoport (2022). Addressing strategic uncertainty with incentives and information. *AEA Pap. Proc. 112*, 431–437. [5]

He, W. and J. Li (2022). Correlation-robust auction design. *Journal of Economic Theory 200*, 105403. [3, 37, 38]

Healy, P. J. and L. Mathevet (2012). Designing stable mechanisms for economic environments. *Theoretical economics 7*(3), 609–661. [5, 38, 39]

Hochstadt, H. (1989). *Integral equations*. Wiley Classics Library. John Wiley & Sons. [50]

Hu, N., J. Haghpanah, and R. Hartline (2021). Full surplus extraction from samples. *Journal of Economic Theory 193*. [27]

Jehiel, P. and L. Lamy (2018). A mechanism design approach to the tiebout hypothesis. *Journal of Political Economy 126*(2), 735–760. [15]

Jehiel, P., M. Meyer-ter Vehn, and B. Moldovanu (2012). Locally robust implementation and its limits. *Journal of Economic Theory 147*(6), 2439–2452. [3, 33, 34, 37]

Jehiel, P., M. Meyer-ter Vehn, B. Moldovanu, and W. R. Zame (2006). The limits of ex post implementation. *Econometrica 74*(3), 585–610. [9]

Jehiel, P. and B. Moldovanu (2001). Efficient design with interdependent valuations. *Econometrica 69*(5), 1237–1259. [9, 15]

Jewitt, I. (1988). Justifying the first-order approach to principal-agent probblems. *Econometrica*, 1177–1190. [15, 37, 39]

Laffont, J.-J. and E. Maskin (1980). A differential approach to dominant strategy mechanisms. *Econometrica*, 1507–1520. [15, 39]

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review 107*(11), 3257–3287. [38]

Lopomo, G., L. Rigotti, and C. Shannon (2021). Uncertainty in mechanism design. *arXiv:2108.12633*. [3, 18, 33, 34, 37]

Lopomo, G., L. Rigotti, and C. Shannon (2022). Uncertainty and robustness of surplus extraction. *Journal of Economic Theory 199*, 105088. [3, 27, 37]

Maskin, E. (1999). Nash equilibrium and welfare optimality. *The Review of Economic Studies 66*(1), 23–38. [5]

Mathevet, L. (2010). Supermodular mechanism design. *Theoretical Economics 5*(3), 403–443. [5, 38, 39]

Mathevet, L. and I. Taneva (2013). Finite supermodular design with interdependent valuations. *Games and Economic Behavior 82*, 327–349. [5, 38, 39]

McAfee, R. P. and P. J. Reny (1992). Correlated information and mecanism design. *Econometrica*, 395–421. [2, 5, 6, 7, 27]

Milgrom, P. R. (2004). *Putting auction theory to work*. Cambridge University Press. [15]

Milgrom, P. R. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica 70*(2), 583–601. [15, 35]

Müller, C. (2016). Robust virtual implementation under common strong belief in rationality. *Journal of Economic Theory 162*, 407–450. [36]

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research 6*(1), 58–73. [15, 17]

Neeman, Z. (2004). The relevance of private information in mechanism design. *Journal of Economic Theory 117*, 55–77. [7]

Ollár, M. and A. Penta (2017). Full implementation and belief restrictions. *American Economic Review 107*(8), 2243–2277. [3, 5, 9, 33, 36, 39]

Ollár, M. and A. Penta (2022). Efficient full implementation via transfers: Uniqueness and sensitivity in symmetric environments. Volume 112, pp. 438–443. American Economic Association. [3, 5, 36, 39]

Ollár, M. and A. Penta (2023). A network solution to robust implementation: The case of identical but unknown distributions. *Review of Economic Studies 90*(5), 2517–2554. [3, 5, 9, 33, 34, 36, 39]

Ollár, M. and A. Penta (2024a). Incentive compatibility with multi-dimensional types: the role of belief restricitons. Technical report. [9]

Ollár, M. and A. Penta (2024b). Robust implementation via transfers: the case of general smooth valuations. Technical report. [5, 36]

Palfrey, T. R. and S. Srivastava (1989). Implementation with incomplete information in exchange economies. *Econometrica 57*, 115–134. [5]

Penta, A. (2011). Backward induction reasoning in games with incomplete information. *University of Winconsin-Madison*. [36]

Penta, A. (2015). Robust dynamic implementation. *Journal of Economic Theory 160*, 280–316. [9, 36]

Rogerson, W. P. (1985). The first-order approach to principal-agent probblems. *Econometrica*, 1357–1367. [15, 37, 39]

Sandholm, W. H. (2002). Evolutionary implementation and congestion pricing. *The Review of Economic Studies 69*(3), 667–689. [38, 39]

Sandholm, W. H. (2005). Negative externalities and evolutionary implementation. *The Review of Economic Studies 72*(3), 885–915. [38, 39]

Sandholm, W. H. (2007). Pigouvian pricing and stochastic evolutionary implementation. *Journal of Economic Theory 132*(1), 367–382. [38, 39]

Segal, I. (2003). Optimal pricing mechanisms with unknown demand. *American Economic Review 93*(3), 509–529. [15]

Wilson, R. (1987). *Game-theoretic analyses of trading processes*. Advances in Economic Theory. in Bewley (ed.), Cambridge University Press. [2, 3]

Winter, E. (2004). Incentives and discrimination. *American Economic Review 94*(3), 764–773. [5]

Yamashita, T. (2015). Implementation in weakly undominated strategies: Optimality of second-price auction and posted-price mechanism. *The Review of Economic Studies 82*(3), 1223–1246. [38]

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 Θ_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 $t_i^*(m) = -v_i(d(m), m) + \int_{\underline{\theta_i}}^{m_i} \frac{\partial v_i}{\partial \theta_i}(d(s, m_{-i}), s, m_{-i})\, ds$. Since d is p.diff, restricted to its pieces, $\frac{\partial v_i}{\partial \theta_i}(d(\cdot), \cdot) : M \to \mathbb{R}$ is continuously differentiable over the same pieces as v_i is twice cont.diff. Therefore $\int^{m_i} \frac{\partial v_i}{\partial \theta_i}$ is p.diff over M, and thus t_i^* is p.diff over M.

Now, consider a piecewise differentiable $\mathcal{B}$-IC t_i, and we let $\beta_i := t_i - t_i^*$. Then, by Claim 1, β_i is p.diff over M. Next, since t_i is $\mathcal{B}$-IC, for all θ_i, $b \in B_{\theta_i}$, we have that, when the derivative exists, $\left[\partial_i \mathbb{E}^b\big(v_i(d(m_i, \theta_{-i}), \theta) + t_i(m_i, \theta_{-i})\big)\right]\big|_{m_i = \theta_i} = 0$. 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 θ_i and $b \in B_{\theta_i}$ for which both derivatives exist, we have $\left[\partial_i \mathbb{E}^b\big(t_i - t_i^*\big)(m_i)\right]\big|_{m_i = \theta_i} = 0$. Next, we use the following claim to extend this result to all differentiability points of $\mathbb{E}^b \beta_i$, beyond the joint differenttiability points of $\mathbb{E}^b t_i$ and $\mathbb{E}^b t_i^*$. $\square$

CLAIM 2: For a p.diff $f : M \to \mathbb{R}$ and $b \in \Delta(\Theta_{-i})$ 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: $S_1^b, \ldots, S_K^b$. Write $\mathbb{E}^b f(m_i) = \int_{\Theta_{-i}} f(m_i, \theta_{-i})\, db = \sum_{j=1}^K \int_{int\ S_j^b} f(m_i, \theta_{-i})\, db$. For each j, let $A_j(m_i) := \int_{int\ S_j^b} f(m_i, \theta_{-i})\, db$. Since f is p.diff over M, it is p.diff over each S_j^b and it has finitely many pieces of S_j^b: $S_{j,1}^b, \ldots, S_{j,l}^b, \ldots, S_{j,L_j}^b$. Rewrite A_j such that $A_j(m_i) = \sum_{l=1}^{L_j} \int_{int\ S_{j,l}^b} f(m_i, \theta_{-i})\, db$, and note that f is continuouse over $int\ S_{jl}^b$.

[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.

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). $\square$

Note that by Claim 2, if b has a p.diff cdf, then $\mathbb{E}^b v_i$ 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. $\blacksquare$

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 β_i can take.

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

Proof of Corollary 2. Let $\mathcal{B}^\diamond$ be a Bayesian environment with independent types, and note that by independence the belief does not change with the type, so let $b_i^\diamond \in \Delta(\Theta_{-i})$ denote agent i's beliefs, regardless of his type. First, recall that $\mathbb{E}^{b_i^\diamond} [\beta_i (\cdot, \theta_{-i})]$ 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 κ_i over M_i, which proves (1) of this corollary. By the characterization in Theorem 1, (2) and (3) follow. $\blacksquare$

[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.

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, β_i is also twice continuously differentiable and as the functions have compact domains, by the Leibniz rule, (1) obtains from Theorem 1. Further, under t_i, reporting θ_i is locally optimal and thus (2) obtains from the decomposition of the payoff function into U_i^* and β_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 θ_i, hence t_i is $\mathcal{B}$-IC. ■

Characterization of Belief-based Terms in Ex. 2. CLAIM: Consider the belief-restrictions $\mathcal{B}^\gamma$; for all $i \in \{1,2\}$ and for all θ_i, $B^\gamma_{\theta_i} = \left\{b \in \Delta(\Theta_j) : \mathbb{E}^b \theta_j = \gamma_i \theta_i\right\}$. 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 $\beta_i : M \to \mathbb{R}$ which is differentiable in m_i is a belief-based term if and only if for some real functions H_i on M and τ_i on M_{-i}, it takes the form $\beta_i(m) = \int_0^{m_i} \left(s - \frac{m_j}{\gamma_i}\right) H_i(s)\, ds + \tau_i(m_{-i})$.

Proof of the Claim. First, if β_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 θ_i, at the truthtelling profile for all beliefs in B_{θ_i} satisfies the expected value condition, thus it is a belief-based term. Second, in the other direction, if β_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 θ_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 α such that $\alpha x + (1-\alpha) y = \gamma_i \theta_i$. Note that such α 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 onM_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 θ_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 β_i takes the form as stated in Ex. 2. □

REMARK. Notice that this result is a consequence of linear algebraic principles. Consider a closed set X and a continuous function $f : X \to \mathbb{R}$ from the usual normed vector space of the continuous functions with the L_1 norm. The set of continuous functions in $\mathcal{M} := \{\mu \in \Delta(X) : \mathbb{E}^{\mu} f = 0\}$ spans the orthogonal complement of f and thus a function g from the same space satisfies $\mathbb{E}^{\mu} g = 0$ for all $\mu \in \mathcal{M}$ only if $g = \alpha f$ for some $\alpha \in \mathbb{R}$. Applying this in the setting of the above example, for fixed θ_i: X is Θ_{-i}, f is $\theta_j - \theta_i/\gamma_i$ and g is $\partial_i \beta_i$ and it must satisfy the expected value condition in Eq. 5. Thus for any fixed θ_i, $\partial_i \beta_i$ must be in the linear space of $\theta_j - \theta_i/\gamma_i$, that is for some $\alpha_{\theta_i} \in \mathbb{R}$, $\partial_i \beta_i = \alpha_{\theta_i}(\theta_j - \theta_i/\gamma_i)$. We generalize this characterization of belief-based terms by using this observation later in the Proofs of Prop. 3 and 4.

Proof of Theorem 3. The payoffs $U_i = v_i + t_i^* + \beta_i$, by using (3) and adding and subtracting $\int_{m_i}^{\theta_i} \frac{\partial v_i}{\partial \theta_i}(d(s, m_{-i})\, s, m_{-i})\, ds + \beta_i(\theta_i, m_{-i})$, can be rewritten, at the profile $m_{-i} = \theta_{-i}$, as

$$U_i(m_i, \theta_{-i}; \theta) = \int_{\underline{\theta_i}}^{\theta_i} \frac{\partial v_i}{\partial \theta_i}(d(s, \theta_{-i}), s, \theta_{-i})\ ds + \beta_i(\theta)$$
$$- \int_{m_i}^{\theta_i} \underbrace{\left(\frac{\partial v_i}{\partial \theta_i}(d(s, \theta_{-i}), s, \theta_{-i}) - \frac{\partial v_i}{\partial \theta_i}(d(m_i, \theta_{-i}), s, \theta_{-i}) \right)}_{=:\mathcal{SC}_i(m_i, s, \theta_{-i})} ds + \beta_i(m_i, \theta_{-i}) - \beta_i(\theta).$$

The first two terms do not depend on the report m_i, and the latter three terms give 0 if $m_i = \theta_i$. Thus $m_i = \theta_i$ is best response if and only if the expected gain from misreport, $-\mathbb{E}^b \int_{m_i}^{\theta_i} \mathcal{SC}_i(m_i, s, \theta_{-i})\, ds + \mathbb{E}^b \beta_i(m_i) - \mathbb{E}^b \beta_i(\theta_i)$, is nonpositive; which is the condition from the inequality of this theorem. ■

Proof of Proposition 3. Fix agent i. If $\mathcal{B}$ is maximal with respect to $(L_i, f_i)_{i \in I}$, then any belief-based term β_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.[17]

[17] One direction is clear. To prove the other direction, recall the following. Consider a closed set X and a continuous function $f : X \to \mathbb{R}$ from the usual normed vector space of continuous functions with the L_1 norm. The set of continuous functions in $\mathcal{M} := \{\mu \in \Delta(X) : \mathbb{E}^{\mu} f = 0\}$ spans the orthogonal complement of f and thus a function g from the same space satisfies $\mathbb{E}^{\mu} g = 0$ for all $\mu \in \mathcal{M}$ only if $g = \alpha f$ for some $\alpha \in \mathbb{R}$. Applying this here, for fixed θ_i: X is Θ_{-i}, f is $L_i(\theta_{-i}) - f_i(\theta_i)$ and g is $\partial_i \beta_i$ and it must satisfy the expected value condition in Eq. 5. Thus for any fixed θ_i, $\partial_i \beta_i$ must be in the linear space of $L_i(\theta_{-i}) - f_i(\theta_i)$.

Then, if t_i is $\mathcal{B}$-IC, by Theorem 1, it can be written as,

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

Next, we need to check when the SOC at the truthful profile holds.[18] To this end, we need to study when it is the case that for all $b_{\theta_i} \in B_{\theta_i}$,

$$\partial_{ii}^2 \mathbb{E}^{b_{\theta_i}} U_i^*\left(m_i, \theta_{-i}, \theta\right)\Big|_{m_i=\theta_i} + \partial_{ii}^2 \mathbb{E}^{b_{\theta_i}} \beta_i\left(m_i, \theta_{-i}\right)\Big|_{m_i=\theta_i} \le 0$$

$$-\mathbb{E}^{b_{\theta_i}}\left(\frac{\partial^2 v_i\left(d\left(\theta\right), \theta\right)}{\partial x \partial \theta_i}\frac{\partial d\left(\theta\right)}{\partial \theta_i}\right) \le f_i'\left(\theta_i\right) H_i\left(\theta_i\right)$$

Let us set

$$\overline{SCM}_i\left(\theta_i\right) := \sup_{b_{\theta_i}\in B_{\theta_i}} \mathbb{E}^{b_{\theta_i}}\left(-\frac{\partial^2 v_i\left(d\left(\theta\right), \theta\right)}{\partial x \partial \theta_i}\frac{\partial d\left(\theta\right)}{\partial \theta_i}\right).$$

With this notation, if $f_i' > 0$, then $\overline{SCM}_i/f_i'$ is a lower bound on H_i and if $f_i' < 0$, then $\overline{SCM}_i/f_i'$ 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:

$$\mathbb{E}^{b_{\theta_i}}\beta_i\left(\theta\right) = \mathbb{E}^{b_{\theta_i}}\int_{\underline{\theta_i}}^{\theta_i}\left(L_i\left(\theta_{-i}\right) - f_i\left(s\right)\right)H_i\left(s\right)\,ds$$

$$= \int_{\underline{\theta_i}}^{\theta_i}\left(\mathbb{E}^{b_{\theta_i}}L_i\left(\theta_{-i}\right) - f_i\left(s\right)\right)H_i\left(s\right)\,ds = \int_{\underline{\theta_i}}^{\theta_i}\left(f_i\left(\theta_i\right) - f_i\left(s\right)\right)H_i\left(s\right)\,ds.$$

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 $\mathbb{E}^{b_{\theta_i}}\beta_i\left(\theta\right) \ge \int_{\underline{\theta_i}}^{\theta_i}\left(f_i\left(\theta_i\right) - f_i\left(s\right)\right)\left[\overline{SCM}_i/f_i'\right]\left(s\right)\,ds.$ 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 $\mathcal{B}$-IC and extracts the surplus. By Theorem 1, t_i can be written as $t_i\left(m\right) = t_i^*\left(m\right) + \int_{\underline{\theta_i}}^{m_i}\left(L_i\left(m_{-i}\right) - f_i\left(s\right)\right)H_i\left(s\right)\,ds$ $\tau_i\left(m_{-i}\right)$. Moreover, for all θ_i and $b \in B_{\theta_i}$, $\mathbb{E}^b U_i^t\left(\theta;\theta\right) = 0$. Using the formula in 3, and

[18] The canonical externalities are $\partial_{ij}^2 U_i^*\left(m,\theta\right) = \left(\frac{\partial^2 v_i(\theta,d(m))}{\partial^2 x}\frac{\partial d}{\partial \theta_j} - \frac{\partial^2 v_i(m,d(m))}{\partial x \partial \theta_j} - \frac{\partial^2 v_i(m,d(m))}{\partial^2 x}\frac{\partial d}{\partial \theta_j}\right)\frac{\partial d}{\partial \theta_i} + \left(\frac{\partial v_i(\theta,d(m))}{\partial x} - \frac{\partial v_i(m,d(m))}{\partial x}\right)\frac{\partial^2 d}{\partial \theta_j \partial \theta_i}.$

the calculation for $\mathbb{E}^{b_{\theta_i}} \int_{\underline{\theta_i}}^{\theta_i} \left(L_i\left(\theta_{-i}\right) - f_i\left(s\right)\right) H_i\left(s\right)\ ds = \int_{\underline{\theta_i}}^{\theta_i} \left(f_i\left(\theta_i\right) - f_i\left(s\right)\right) H_i\left(s\right)\ ds$ as in the Proof of Prop. 3, these impy that

$$\mathbb{E}^b \left(\int_{\underline{\theta_i}}^{\theta_i} \frac{\partial v_i}{\partial \theta_i} \left(d\left(s, \theta_{-i}\right) s, \theta_{-i}\right)\ ds + \tau_i\left(\theta_{-i}\right) \right) = - \int_{\underline{\theta_i}}^{\theta_i} \left(f_i\left(\theta_i\right) - f_i\left(s\right)\right) H_i\left(s\right)\ ds.$$

The RHS of this expression depends on θ_i but not on b, therefore the LHS must be the same for all $b \in B_{\theta_i}$. Applying the argument of Footnote 17 to these functions, we have that the function $\int_{\underline{\theta_i}}^{\theta_i} \frac{\partial v_i}{\partial \theta_i} \left(d\left(s, \theta_{-i}\right) s, \theta_{-i}\right)\ ds + \tau_i\left(\theta_{-i}\right)$ must be L_i-linear. This function is differentiable in θ_i and thus its derivative $\frac{\partial v_i}{\partial \theta_i}\left(d\left(\theta\right), \theta\right)$ must be L_i-linear as well. In summary, unless $\frac{\partial v_i}{\partial \theta_i}\left(d\left(\theta\right), \theta\right)$ is L_i-linear, $\mathcal{B}$-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 $\mathcal{B}$ and $\mathcal{B}'$. Note that for each i, there is a function $L_i : M_{-i} \to \mathbb{R}$ such that $\frac{\partial v_i}{\partial \theta_i}\left(d\left(\theta\right), \theta\right)$ is not L_i-linear. For each i fix $\gamma_i \in (0,1)$, and let the belief-restrictions $\mathcal{B}$ be maximal with respect to the responsive moment condition $\left(L_i, \gamma_i \theta_i\right)_{i \in I}$.Prop. 1 implies that $\mathcal{B}$-IC transfers exist, thus $F\left(\mathcal{B}\right)$ is non-empty and $\infty > \tau\left(\mathcal{B}\right)$. Yet, as a consequence of Prop. 4, FSE is not possible, that is, $\tau\left(\mathcal{B}\right) > 0$. Next, let $\mathcal{B}'$ be s.t. $B'_{\theta_i} = \{p_{\theta_i}\}$ and s.t. (i) p_{θ_i} has a pdf that is continuouse and non-zero over the support $\times_{j \neq i} \left[\underline{\theta}_j, \underline{\theta}_j + \left(\theta_i - \underline{\theta}_i\right)\left(l_j / l_i\right)\right]$, where for each i, $l_i := \overline{\theta}_i - \underline{\theta}_i$, and (ii) for all θ_i, $\mathbb{E}^{p_{\theta_i}} L_i\left(\theta_{-i}\right) = \gamma_i \theta_i$. (Note that for each θ_i, matching the fixed first moment is possible.) For $\mathcal{B}'$ thus constructed, the construction in Ex. 3 shows that a t exists which ensured FSE and is $\mathcal{B}$-IC and hence $\mathcal{B}'$-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 θ_i and beliefs given such that $\theta_j | \theta_i = \gamma \nu_{\theta_i} + (1 - \gamma)\, \eta_{ij}$ where ν_{θ_i} is $U\left[0, \theta_i\right]$ and, independently of this, η_{ij} is $U\left[0, 1\right]$. Let us examine the solvability of $\int_0^1 \alpha_i\left(\theta_j\right) p\left(\theta_j | \theta_i\right) d\theta_j = f\left(\theta_i\right)$. (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 $1-\gamma > \gamma\theta_i$,

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

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

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

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

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

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:

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

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

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

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.[19]

Part 2: If $\gamma > 1/2$, then there are two cases to be considered: either $1-\gamma > \gamma\theta_i$ or $1-\gamma \le \gamma\theta_i$. 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

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

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

$$\begin{aligned} 0 + \int_\gamma^{1-\gamma+s} \alpha(\theta_j)\ d\theta_j &= (1-\gamma)\left(f\left(\frac{s}{\gamma}\right)s\right)' \\ \alpha(1-\gamma+s) &= (1-\gamma)\left(f''\left(\frac{s}{\gamma}\right)\frac{s}{\gamma} + 2f'\left(\frac{s}{\gamma}\right)\frac{1}{\gamma}\right), \end{aligned}$$

to, finally, change variables, use the notation L_γ 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): $1-\gamma \le \gamma\theta_i$. 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

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

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

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

[19] Note that $L_\gamma(s) = \left(f\left(\frac{s}{\gamma}\right)s\right)''$.

Finally, change variables, use the notation L_γ, 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. $\square$