Locally robust implementation and its limits*

Philippe Jehiel†, Moritz Meyer-ter-Vehn‡, Benny Moldovanu§

March 12, 2012

Abstract

We study a notion of locally robust implementation that captures the idea that the planner may know agents' beliefs well, but not perfectly. Locally robust implementation is a weaker concept than ex-post implementation, but we show that no regular allocation function is locally robust implementable in generic settings with quasi-linear utility, interdependent and bilinear values, and multi-dimensional payoff types.

1 Introduction

Bayesian mechanism design is frequently criticized for assuming too much knowledge about agents' beliefs. This knowledge gives the planner an implausible amount of power when designing the mechanism, and optimal mechanisms can be very sensitive to this knowledge, e.g., the well-known full surplus extraction mechanism of Crémer and McLean [6]. To address this issue, the robust mechanism design literature follows Harsanyi [10] by modeling an agent's belief as part of her private type, and requiring a robust mechanism to be incentive compatible for a range of agents' beliefs so as to reflect the planner's uncertainty about these beliefs (see Bergemann and Morris [1] and also Neeman [21] for an earlier investigation on mechanism design with a focus on payoff and belief types.).

Much of the robust mechanism design literature, e.g. Bergemann and Morris [2], takes the above criticism of the Bayesian paradigm to the opposite extreme, and assumes that the planner

*We would like to thank seminar audiences at the Princeton Mechanism Design Conference 2010, Toronto and UCLA for helpful comments. This work originated at the Trimester Program on Mechanism Design at the university of Bonn. We are grateful to the Hausdorff Research Institute for Mathematics for its hospitality and financial support. Keywords: robust implementation, interdepedent values, multi-dimensional types, perturbed beliefs. JEL: D82

†jehiel@pse.ens.fr, http://www.enpc.fr/ceras/jehiel/, PSE, 48 Bd Jourdan, 75014 Paris, France, and University College London

‡Corresponding author. mtv@econ.ucla.edu, http://www.econ.ucla.edu/mtv/ UCLA, Dept. of Economics, Bunche Hall 8283, Los Angeles, CA 90095, USA, Tel: +1-310-825-1011, Fax: +1-310-825-9528

§mold@uni-bonn.de, http://www.econ2.uni-bonn.de/moldovanu, University of Bonn, Lennéstr. 37, 53113 Bonn, Germany, Tel: + 49-228-73 63 95, Fax: + 49-228-73 7940

knows nothing at all about agents' beliefs. When the planner allows for all first-order beliefs of agents, any robustly implementable choice function is also dominant-strategy implementable when valuations are private, or ex-post implementable when valuations are interdependent, as shown by Ledyard [16] and by Bergemann and Morris [1], respectively.

Dominant-strategy and ex-post implementation are overly restrictive in important settings. In private value environments with unrestricted preference types and three or more social alternatives, Gibbard [8] and Satterthwaite [25] show that only dictatorial choice functions are implementable in dominant strategies. Restricting attention to quasi-linear utilities gives rise to more positive results when values are private, as shown by Vickrey [26], Clarke [5], Groves [9] and Roberts [23]. In interdependent value environments, positive results regarding ex-post implementation are obtained when signals are one-dimensional and value functions satisfy a single-crossing property (see Dasgupta and Maskin [7] and Jehiel and Moldovanu [12]). But, Jehiel, Meyer-ter-Vehn, Moldovanu and Zame [13], JMMZ henceforth, show that only trivial allocation functions are implementable when payoff types are multi-dimensional and the interdependent value functions are generic.[1] The strong negative results due to Gibbard, Satterthwaite and JMMZ suggest a weakening of the implementation concept.

In this paper we relax the requirement that a mechanism be incentive compatible for any first-order beliefs of the agents. More precisely, we only require the mechanism to be incentive compatible for beliefs that lie in a neighborhood of some benchmark beliefs (which may be derived from some common prior as usually assumed in the mechanism design literature). We call such a mechanism *locally robust*, and ask which social choice functions can be locally robustly implemented in this sense.

We show by example that some social choice functions can be locally robustly implemented while not being ex-post implementable. Thus, the notion of locally robust implementation does not reduce to ex-post implementation. Yet, the main result of this paper extends the impossibility result of JMMZ to locally robust implementation. More precisely, with quasi-linear utility and multi-dimensional payoff types, locally robust implementation implies a geometric condition that equates the marginal rates of information substitution of agents' value functions and the allocation function. This condition, in turn, implies a system of differential equations that needs to be satisfied by the value functions. But for almost all bilinear value functions, this system does not have a solution.[2]

The connection between our main present result and the impossibility result of JMMZ is in-

[1] The JMMZ genericity notion excludes several interesting settings. For example, Bikhchandani [4] shows that non-trivial ex-post implementation can be achieved in auction environments without consumption externalities. Jehiel, Meyer-ter-Vehn and Moldovanu [14] also display some possibility results in a non-generic framework with multi-dimensional signals.

[2] We restrict attention to the finite-dimensional space of bilinear value functions to allow for an elementary genericity notion, and for an elementary proof of our main result. This approach also allows us to avoid a technical assumption in JMMZ which is violated in our example in Section 3.

structive. As for many other implementation concepts, locally robust mechanisms need to satisfy a monotonicity condition and an integrability condition (commonly referred to as 'payoff equivalence'). Locally robust implementation is weaker than ex-post implementation because an allocation function that is monotone for a small set of beliefs need not be monotone ex-post. This is so because monotonicity is an inequality constraint: if the inequality is strict in expectation then it is still satisfied when some probability is shifted to realizations where monotonicity is violated ex-post. Locally robust implementation is generically not feasible because integrability for a small set of beliefs implies integrability ex-post. This is so because integrability on multi-dimensional payoff type spaces implies that equilibrium marginal utility is a conservative vector field, determined by the value function and the allocation function. Conservativeness imposes an equality constraint on the cross-partials of these functions. This equality must hold ex-post if it holds in expectation for an open set of beliefs.

The concept of *locally robust incentive compatibility* defined in this paper is very similar to *optimal incentive compatibility* defined in Lopomo, Rigotti and Shannon [17], LRS henceforth.[3] For payoff environments more general than the quasi-linear environment considered in this paper, LRS show that optimal incentive compatibility together with ex-post cyclical monotonicity implies ex-post incentive compatibility. But, ex-post cyclical monotonicity is a strong assumption which by itself implies ex-post implementability in quasi-linear environments, as shown by Rochet [24]. Conversely, locally robust implementability by itself does not imply ex-post implementability as shown by an example in Section 3 below. Therefore, our main result does not follow by combining the results of LRS and JMMZ. Locally robust implementation is also similar to *continuous implementation*, as defined in Oury and Tercieux [22] who relate partial implementation of a social choice function on the neighborhood of a type space to full implementation of this social choice function.

We proceed as follows. Section 2 introduces the model; Section 3 shows by example that locally robust implementation is more permissible than ex-post implementation; Section 4 shows that locally robust mechanisms satisfy monotonicity and integrability in expectation, and integrability ex-post; Section 5 introduces a regularity condition on allocation functions and proves the main impossibility result, Theorem 1.

2 The Model

The Payoff Environment: We consider the simplest setup in which our main result, Theorem 1, holds. Specifically, there are two social allocations $x \in \{0, 1\}$, and there are two agents $i \in \{1, 2\}$ with payoff types θ_i drawn from d_i-dimensional cubes $\Theta_i = [0, 1]^{d_i}$. Agents have quasi-linear

[3] For LRS the uncertainty about beliefs is in the mind of the ambiguity-averse agent, while for us the uncertainty is in the mind of the planner. Both interpretations give rise to the same model.

Bernoulli utility functions of the form $u_i = xv_i(\theta_i, \theta_{-i}) - p_i$, where p_i is a monetary payment by agent i and v_i is i's smooth interdependent value function. For the proof of genericity in Theorem 1 we will restrict attention to the finite-dimensional space of bilinear value functions v_i.

The Type Space: *Baseline beliefs* are given by continuous functions $\pi_i^* : \Theta_i \to \Delta(\Theta_{-i})$, where Θ_i is equipped with the sup-norm and $\Delta(\Theta_{-i})$ with the metric of absolute variation; for $\varepsilon > 0$ we let B_ε be the open ε-balls in these metric spaces.

Even though not required for our main result, we observe that the baseline belief π_i^* could be derived from a common prior distribution π^* over Θ where $\pi_i^*(\theta_i)$ would be the marginal of π^* over θ_{-i} conditional on θ_i. This common distribution π^* could allow for correlation between θ_i and θ_{-i} as in the work of Crémer and McLean [6].[4]

To model the planner's local uncertainty about agents' beliefs $\pi_i^*(\theta_i)$ we assume that there is $\varepsilon > 0$ such that agent i's type space $T_i \subset \Theta_i \times \Delta(\Theta_{-i})$ includes all ε-perturbed beliefs, that is $\theta_i \times B_\varepsilon(\pi_i^*(\theta_i)) \subset T_i$ for all θ_i. We interpret $\pi_i \in \Delta(\Theta_{-i})$ as a belief over T_{-i} with marginal π_i over Θ_{-i} such that $\pi_i\{(\theta_{-i}, \pi_{-i}^*(\theta_{-i}))|\ \theta_{-i} \in \Theta_{-i}\} = 1$. This means that the type space T_i differs from a standard Bayesian type space $\{(\theta_i, \pi_i^*(\theta_i))\,|\theta_i \in \Theta_i\}$ only to the degree that agent i could have different beliefs about $-i$'s payoff types, but i believes with probability one that $-i$'s beliefs are specified by π_{-i}^*.

We view T_i as a small type space because every neighborhood of $\{(\theta_i, \pi_i^*(\theta_i))\,|\theta_i \in \Theta_i\}$ in the universal type space with respect to the product, or to the uniform-weak topology includes such a neighborhood T_i. Importantly, the definition ensures that T_i is large enough to ensure that beliefs are locally independent in the following sense: For every θ_i there exists $\varepsilon > 0$, an ε-ball of payoff types $B_\varepsilon(\theta_i)$, and an ε-ball of belief types $B_\varepsilon(\pi_i^*(\theta_i))$, such that:[5,6]

$$B_\varepsilon(\theta_i) \times B_\varepsilon(\pi_i^*(\theta_i)) \subset T_i. \tag{1}$$

Implementation: The planner wants to implement a (possibly stochastic) allocation $q : \Theta \to [0,1]$ as a function of payoff types θ. An allocation function q is *locally robust implementable* if there exist type spaces T_i as above and a (possibly belief-dependent) payment function $p : T \to \mathbb{R}^2$, such that the direct revelation mechanism (q, p) is incentive compatible on T, i.e.

$$\mathbb{E}_{\pi_i}[v_i(\theta)\,q(\theta) - p_i(t)] \geq \mathbb{E}_{\pi_i}\left[v_i(\theta)\,q(\theta') - p_i(t')\right] \tag{IC}$$

for all $t_i = (\theta_i, \pi_i), t_i' = (\theta_i', \pi_i')$, where we set $\theta = (\theta_i, \theta_{-i}), \theta' = (\theta_i', \theta_{-i}), t = (\theta_i, \pi_i, \theta_{-i}, \pi_{-i}), t' = (\theta_i', \pi_i', \theta_{-i}, \pi_{-i})$.

[4] We could also allow the baseline belief to bear on payoff-irrelevant aspects of the type (such as, for agent i, signals over agent $-i$'s realization of θ_{-i}). Yet, the same result as Theorem 1 would hold for this more general setting.

[5] This is an elementary version of 'overlapping beliefs', as defined by LRS.

[6] This argument relies on the continuity of the belief functions $\pi_i^*(\cdot)$.

Locally robust implementability is not directly comparable to *optimal incentive compatibility* introduced in LRS because it pertains to allocation functions $q : \Theta \to [0,1]$ that need to be augmented by payment functions $p : T \to \mathbb{R}^2$ which may additionally depend on agents' beliefs. However, when we fix a mechanism with belief-independent payments $(q,p) : \Theta \to [0,1] \times R^2$ we can address LRS' question, that is whether locally robust incentive compatibility of (q,p) implies ex-post incentive compatibility.

Locally robust implementation is a weak implementation concept since: (1) payments are allowed to depend on beliefs; (2) condition (IC) only requires partial implementation; (3) the type space T is small. This implies that our negative result, Theorem 1, is strong. In contrast, any positive result for locally robust implementation may be subjected to the critique that it is due to the above three factors. Therefore, we argue at the end of Section 3 that the positive result in that section is not due to these factors, but that it obtains under more demanding notions of locally robust implementation.

3 Locally Robust vs. Ex-Post Implementation

In this Section we illustrate that a locally robust implementable allocation function q need not be ex-post implementable. While this fact may not be surprising, it is not obvious either, as highlighted by LRS.

Assume that type spaces are one-dimensional $\Theta_i = \Theta_{-i} = [0,1]$, and that value functions are given by $v_i(\theta) = \theta_i(3\theta_{-i} - 1)$ and $v_{-i}(\theta) = \theta_{-i}$. Thus agent i's value is increasing in own type when $\theta_{-i} = 1$ (as $\partial_i v_i(\cdot, \theta_{-i}) \equiv 2$), and is decreasing in own type when $\theta'_{-i} = 0$ (as $\partial_i v_i\left(\cdot, \theta'_{-i}\right) \equiv -1$). When i's belief π_i assigns sufficient weight to high values of θ_{-i}, for example if π_i is uniform on $\Theta_{-i} = [0,1]$, then i's expected value is increasing in own type since

$$\mathbb{E}_{\pi_i}[\partial_i v_i(\cdot, \theta_{-i})] = 3\mathbb{E}_{\pi_i}[\theta_{-i}] - 1 > 0.$$

We consider two allocation functions q: the first to contrast locally robust implementation and ex-post implementation in the simplest possible manner, and the second to satisfy the regularity condition of Section 5 below.

Dictatorial Example: Consider a dictatorial allocation function that only takes i's payoff type into account. Specifically, we define q by a cutoff $\theta_i^* \in (0,1)$ with the property that

$$q(\theta_i, \theta_{-i}) = \begin{cases} 1 \text{ if } \theta_i \geq \theta_i^*, \\ 0 \text{ else.} \end{cases}$$

This allocation function is not ex-post implementable because for $\theta'_{-i} = 0$ it chooses allocation 0 for payoff types $\theta_i < \theta_i^*$ who have a high value for allocation 1, and it chooses allocation 1 for payoff types $\theta_i \geq \theta_i^*$ who have a low value for allocation 1. This ex-post violation of monotonicity is not compatible with agent i's ex-post incentive constraint.

Nevertheless, q is locally robust implementable. To see that, consider uniform beliefs $\pi_i^*(\theta_i) \in \Delta(\Theta_{-i})$ for all θ_i and the payment rule

$$p_i(\theta_i, \theta_{-i}) = \begin{cases} v_i(\theta_i^*, \theta_{-i}) & \text{if } \theta_i \geq \theta_i^*, \\ 0 & \text{else.} \end{cases}$$

Agent i's type (θ_i, π_i) is then effectively choosing between the outcome $(q, p_i) = (1, v_i(\theta_i^*, \theta_{-i}))$ with an expected payoff of

$$\mathbb{E}_{\pi_i}[v_i(\theta_i, \theta_{-i}) - p_i] = \mathbb{E}_{\pi_i}[v_i(\theta_i, \theta_{-i}) - v_i(\theta_i^*, \theta_{-i})]$$

and outcome $(q, p_i) = (0, 0)$ with a payoff of 0. For small $\varepsilon > 0$ and any belief $\pi_i \in B_\varepsilon(\pi_i^*)$ we have $\mathbb{E}_{\pi_i}[\partial_i v_i(\cdot, \theta_{-i})] > 0$, so that the agent indeed chooses $q = 1$ when $\theta_i \geq \theta_i^*$ and $q = 0$ when $\theta_i < \theta_i^*$.

Regular Example: Consider the stochastic allocation function $q(\theta) = (\theta_i + \theta_{-i})/2$. For the same reason as above this allocation function is not ex-post implementable, but it is locally robust implementable. To show this we first assume that agent $-i$ reports truthfully $\hat{\theta}_{-i} = \theta_{-i}$, and analyze i's IC constraint. Consider uniform beliefs $\pi_i^*(\theta_i)$ and the (belief-independent) payment rule

$$p_i(\hat{\theta}_i, \theta_{-i}) = q(\hat{\theta}_i, \theta_{-i}) v_i(\hat{\theta}_i, \theta_{-i}) - \int_0^{\hat{\theta}_i} q(\tilde{\theta}_i, \theta_{-i}) \partial_i v_i(\tilde{\theta}_i, \theta_{-i}) d\tilde{\theta}_i.$$

Then utility of type θ_i reporting type $\hat{\theta}_i$ is given by

$$u_i(\theta_i, \hat{\theta}_i; \theta_{-i}) = q(\hat{\theta}_i, \theta_{-i}) v_i(\theta_i, \theta_{-i}) - q(\hat{\theta}_i, \theta_{-i}) v_i(\hat{\theta}_i, \theta_{-i}) + \int_0^{\hat{\theta}_i} q(\tilde{\theta}_i, \theta_{-i}) \partial_i v_i(\tilde{\theta}_i, \theta_{-i}) d\tilde{\theta}_i$$

and marginal utility in the report $\hat{\theta}_i$ equals

$$\partial_{\hat{\theta}_i} u_i(\theta_i, \hat{\theta}_i; \theta_{-i}) = \underbrace{\partial_{\hat{\theta}_i} q(\hat{\theta}_i, \theta_{-i})}_{=1/2} \underbrace{(v_i(\theta_i, \theta_{-i}) - v_i(\hat{\theta}_i, \theta_{-i}))}_{=(\theta_i - \hat{\theta}_i)(3\theta_{-i} - 1)}.$$

Ex-post with $\theta_{-i} = 0$ we have $3\theta_{-i} - 1 < 0$, so every type θ_i optimally reports either $\hat{\theta}_i = 0$ or $\hat{\theta}_i = 1$.

In contrast, ex-ante we have

$$\partial_{\hat{\theta}_i} \mathbb{E}_{\pi_i} \left[u_i(\theta_i, \hat{\theta}_i; \theta_{-i}) \right] = \frac{1}{2} (\theta_i - \hat{\theta}_i) \mathbb{E}_{\pi_i} [3\theta_{-i} - 1]$$

and $\mathbb{E}_{\pi_i}[3\theta_{-i} - 1] > 0$ for all π_i in the neighborhood of π_i^*, so every type θ_i optimally reports

$\hat{\theta}_i = \theta_i$.

So far we have assumed that agent $-i$ reports truthfully $\hat{\theta}_{-i} = \theta_{-i}$. This is justified, because an analogous construction of payments p_{-i} makes truthful reporting $\hat{\theta}_{-i} = \theta_{-i}$ a strictly dominant strategy.

This positive result for locally robust implementation and the contrast to ex-post implementation is due to the core idea of local robustness, i.e. agents' beliefs are known to be close to some baseline. It is not due to the weaknesses of the solution concept discussed in the previous section since: (1) the mechanisms (q, p) have payments defined as a function of payoff types alone; (2) every rationalizable strategy of type (θ_i, π_i) leads to outcome $q(\theta_i)$, so (q, p) fully implements q; and (3) incentive compatibility is maintained on any larger type space with the same first-order beliefs because higher-order beliefs do not matter in the mechanisms (q, p).

4 Monotonicity and Integrability

As a first step towards the main result, we follow Jehiel, Moldovanu, Stacchetti [11], and show that implementable allocation functions must satisfy locally robust versions of monotonicity and integrability. In deriving these necessary conditions we only exploit agents' ability to misreport payoff types for any given belief type, but ignore their ability to misreport belief types.[7]

Lemma 1 *If the direct mechanism (q, p) is incentive compatible on T, then it satisfies:*

(a) ***Monotonicity:*** *For all θ_i, θ_i' and π_i such that $(\theta_i, \pi_i), (\theta_i', \pi_i) \in T_i$ we have*

$$\mathbb{E}_{\pi_i}\left[\left(v_i(\theta) - v_i(\theta')\right)\left(q(\theta) - q(\theta')\right)\right] \geq 0 \tag{2}$$

where $\theta = (\theta_i, \theta_{-i})$ and $\theta' = (\theta_i', \theta_{-i})$.

(b) ***Integrability:*** *Let θ_i and $\varepsilon > 0$ be such that condition (1) holds. Let*

$$U_{i,\pi_i}(\theta_i) = \mathbb{E}_{\pi_i}[q(\theta_i, \theta_{-i})v_i(\theta_i, \theta_{-i}) - p_i(\theta_i, \pi_i, \theta_{-i}, \pi_{-i})]$$

[7] Ignoring such misreports of beliefs does not significantly weaken the IC constraints, because one can elicit beliefs by a continuous version of the log-scoring rule (see for example Johnson et al. [15]). The discrete version of this rule punishes agent i with the payment rule $p_i(\hat{\pi}_i, t_{-i}) = -\log(\hat{\pi}_i(t_{-i}))$ when i reports belief $\hat{\pi}_i$ and others report type t_{-i} . If i's true belief is π_i and others truthfully report t_{-i}, the net benefit from misreporting her belief as $\hat{\pi}_i$ is non-positive:

$$\mathbb{E}_{\pi_i}\left[\log(\hat{\pi}_i(t_{-i})) - \log(\pi_i(t_{-i}))\right] = \mathbb{E}_{\pi_i}\left[\log\left(\frac{\hat{\pi}_i(t_{-i})}{\pi_i(t_{-i})}\right)\right] \leq \log \mathbb{E}_{\pi_i}\left[\frac{\hat{\pi}_i(t_{-i})}{\pi_i(t_{-i})}\right] = \log 1 = 0$$

where the inequality follows from the concavity of the log function and Jensen's inequality.

Building on this insight, Bergemann, Morris and Takahashi [3] show that an allocation function can be implemented on a finite type space T if and only if it satisfies 'π_i-cyclical montonocity' on $\Theta_i(\pi_i) := \{\theta_i : (\theta_i, \pi_i) \in T_i\}$ for every π_i.

be agent i's expected equilibrium utility with payoff type θ_i under (q,p) and belief π_i. Then for all $(\theta_i',\pi_i)\in B_\varepsilon(\theta_i)\times B_\varepsilon(\pi_i^*(\theta_i))$ and all differentiable paths $s:[0,1]\to B_\varepsilon(\theta_i)$ with $s(0)=\theta_i$ and $s(1)=\theta_i'$, we have

$$U_{i,\pi_i}\left(\theta_i'\right)-U_{i,\pi_i}\left(\theta_i\right)=\int_{\theta_i}^{\theta_i'}\mathbb{E}_{\pi_i}\left[q\left(s,\theta_{-i}\right)\nabla_i v_i\left(s,\theta_{-i}\right)\right]\cdot ds \tag{3}$$

where $\nabla_i v_i$ is the d_i-dimensional vector of partial derivatives of v_i with respect to i's own payoff type. Thus the vector field $\mathbb{E}_{\pi_i}\left[q\left(\cdot,\theta_{-i}\right)\nabla_i v_i\left(\cdot,\theta_{-i}\right)\right]:\Theta_i\to\mathbb{R}^{d_i}$ is conservative on $B_\varepsilon(\theta_i)$.

Proof. To show monotonicity, consider as usual the IC constraints of types (θ_i,π_i) and (θ_i',π_i) not to misreport each other's type:

$$\begin{aligned}\mathbb{E}_{\pi_i}\left[v_i(\theta)\,q(\theta)-p_i(t)\right] &\geq \mathbb{E}_{\pi_i}\left[v_i(\theta)\,q\left(\theta'\right)-p_i\left(t'\right)\right]\\ \mathbb{E}_{\pi_i}\left[v_i\left(\theta'\right)q\left(\theta'\right)-p_i\left(t'\right)\right] &\geq \mathbb{E}_{\pi_i}\left[v_i\left(\theta'\right)q(\theta)-p_i(t)\right]\end{aligned}$$

Adding up the above two inequalities yields (2).

Integrability (or payoff equivalence) basically follows from the envelope theorem. More precisely, fix agent i's belief, and let

$$\mathbb{E}_{\pi_i}\left[u_i(\theta_i,\hat{\theta}_i;\theta_{-i})\right]=\mathbb{E}_{\pi_i}\left[q(\hat{\theta}_i,\theta_{-i})v_i(\theta_i,\theta_{-i})-p_i(\hat{\theta}_i,\pi_i,\theta_{-i},\pi_{-i})\right]$$

be the expected utility of type θ_i when reporting $\hat{\theta}_i$. Let $\theta_i^*(\theta_i)\in\arg\max_{\hat{\theta}_i}\mathbb{E}_{\pi_i}\left[u_i(\theta_i,\hat{\theta}_i;\theta_{-i})\right]$ be any selection from the arg max-correspondence. Then the multi-dimensional version of Corollary 1 in Milgrom and Segal [20] states that

$$U_{i,\pi_i}\left(\theta_i'\right)-U_{i,\pi_i}\left(\theta_i\right)=\int_{\theta_i}^{\theta_i'}\nabla_{\theta_i}\mathbb{E}_{\pi_i}\left[u_i(s,\theta_i^*(s);\theta_{-i})\right]\cdot ds.$$

To conclude the argument, we apply the theorem of dominated convergence to change the order of differentiation and integration, i.e. to pull the gradient $\nabla_{\theta_i}=\nabla_i$ into the expectation. ■

At first, one might be surprised by the fact that Lemma 1.b holds even though no assumption about the independence of the baseline belief across agents has been made. But, note that integrability holds only locally, where the belief of agent i can be held constant (due to our consideration of a neighborhood of the baseline belief). When the belief is constant, the situation is similar to the one arising with independent distributions of types.

Coming back to the examples of Section 3, the contrast between locally robust implementation and ex-post implementation is due to the fact that monotonicity can be satisfied for all close-by beliefs $\pi_i \in B_\varepsilon(\pi_i^*)$, but violated for other far-away beliefs $\pi_i' \in \Delta(\Theta_{-i})$. This is indeed the case in the examples in Section 3 where $\mathbb{E}_{\pi_i}[\partial_i v_i(\cdot, \theta_{-i})] > 0$ for beliefs π_i close-to uniform, but $\mathbb{E}_{\pi_i'}[\partial_i v_i(\cdot, \theta_{-i})] = -1$ for belief π_i' that puts probability one on type $\theta_{-i} = 0$.

Loosely speaking, monotonicity is a locally robust property in the following sense: When inequality (2) is strict for some belief π_i^*, then it is still satisfied when some probability is shifted to ex-post realizations θ_{-i} for which the inequality is violated.

The situation is different for integrability since the requirement that the vector field $\mathbb{E}_{\pi_i}[q(\cdot, \theta_{-i}) \nabla_i v_i(\cdot, \theta_{-i})]$ be conservative translates into an equality constraint (on cross derivatives), which needs to be satisfied also ex-post.

Lemma 2 *If (q, p) is incentive compatible, then for every $\theta_{-i} \in \Theta_{-i}$ the vector field $q(\cdot, \theta_{-i}) \nabla_i v_i(\cdot, \theta_{-i}) : \Theta_i \to \mathbb{R}^{d_i}$ is conservative on Θ_i. That is,*

$$\int_{\theta_i}^{\theta_i} q(s, \theta_{-i}) \nabla_i v_i(s, \theta_{-i}) \cdot ds = 0$$

for all differentiable paths $s : [0,1] \to \Theta_i$ with $s(0) = s(1) = \theta_i$.

Proof. By definition of type space T_i, there exists for every payoff type θ_i some $\varepsilon > 0$ such that $B_\varepsilon(\theta_i) \times B_\varepsilon(\pi_i^*(\theta_i)) \subseteq T_i$. Consider $s : [0,1] \to B_\varepsilon(\theta_i)$ with $s(0) = s(1) = \theta_i$. By Lemma 1, the vector field $\mathbb{E}_{\pi_i}[q(\cdot, \theta_{-i}) \nabla_i v_i(\cdot, \theta_{-i})]$ is conservative on $B_\varepsilon(\theta_i)$ for all $\pi_i \in B_\varepsilon(\pi_i^*(\theta_i))$, and in particular for $\pi_i' = (1-\varepsilon)\pi_i^*(\theta_i) + \varepsilon \mathbb{I}_{\theta_{-i}}$ for any θ_{-i}. Thus,

$$\begin{aligned} 0 &= \int_{\theta_i}^{\theta_i} \Big(\mathbb{E}_{\pi_i'}[q(\cdot, \theta_{-i}) \nabla_i v_i(\cdot, \theta_{-i})] - (1-\varepsilon)\, \mathbb{E}_{\pi_i^*(\theta_i)}[q(\cdot, \theta_{-i}) \nabla_i v_i(\cdot, \theta_{-i})] \Big) \cdot ds \\ &= \varepsilon \int_{\theta_i}^{\theta_i} q(s, \theta_{-i}) \nabla_i v_i(s, \theta_{-i}) \cdot ds \end{aligned}$$

so $q(\cdot, \theta_{-i}) \nabla_i v_i(\cdot, \theta_{-i})$ is conservative on $B_\varepsilon(\theta_i)$.[8] This argument is valid for all θ_i and θ_{-i}, completing the proof. ■

5 Generic Impossibility of Locally Robust Implementation

We now derive the main result of the paper: generically, no regular allocation function is locally robust implementable. We proceed by deriving from Lemma 2 some geometric conditions on the agents' value functions, which do not admit a solution for bilinear value functions with generic parameters.

[8] This argument is an elementary version of the proof of Theorem 1 in LRS.

5.1 The Regularity Assumption

The proof of our main result, Theorem 1, relies on geometric arguments on the boundary $I \subset \Theta$ that separates the regions where different allocations are chosen. In order to facilitate these arguments we focus on *regular* allocation functions that are smooth and non-satiated in the following sense.

1. A stochastic allocation function $q : \Theta \to [0, 1]$ is *regular* if it is smooth with $\nabla_i q(\theta) \neq 0$ for all i and θ.

2. A deterministic allocation function $q : \Theta \to \{0, 1\}$ is *regular* if it maximizes a smooth, non-satiated objective function, i.e. there exists $\psi : \Theta \to \mathbb{R}$ such that $q(\theta) \in \arg\max_{x \in \{0,1\}} x\psi(\theta)$, where ψ satisfies $\nabla_i \psi(\theta) \neq 0$ for all i and θ, and there exists an interior $\theta \in \Theta$ with $\psi(\theta) = 0$.[9,10]

For deterministic regular q we fix $I = \psi^{-1}(0)$. For stochastic regular q we choose any interior θ and fix $I = q^{-1}(q(\theta))$. In either case I is a $d_i + d_{-i} - 1$-dimensional submanifold of Θ. For any interior $\theta^* = (\theta_i^*, \theta_{-i}^*) \in I$ let $I_i(\theta^*) \subseteq \Theta_i$ be the path-connected component of $\{\theta_i \in \Theta_i : (\theta_i, \theta_{-i}^*) \in I\}$ that includes θ_i^*.

Lemma 3 *Assume that $d_i \geq 2$, that q is regular, and that $q\left(\cdot, \theta_{-i}^*\right) \nabla_i v_i \left(\cdot, \theta_{-i}^*\right) : \Theta_i \to \mathbb{R}^n$ is a conservative vector field for some interior $\theta^* \in I$. Then $v_i\left(\cdot, \theta_{-i}^*\right)$ is constant on $I_i(\theta^*)$.*

Proof. The idea of the proof is to construct a 'taxation mechanism' with allocation function $q(\cdot, \theta_{-i}^*)$, that is generally not incentive compatible but satisfies a first-order condition that implies the Lemma.

To simplify notation in this proof, we drop the argument θ_{-i}^* from the functions q and v_i. Fix $\theta_i^0 \in \Theta_i$. As $q\nabla v_i$ is conservative, the integral

$$\mathcal{U}_i(\theta_i) = \int_{\theta_i^0}^{\theta_i} q(s) \nabla v_i(s) \cdot ds$$

is the same for any differentiable path $s : [0, 1] \to \Theta_i$ with $s(0) = \theta_i^0$ and $s(1) = \theta_i$. We interpret $\mathcal{U}_i$ as i's 'equilibrium utility' of a 'taxation mechanism' with allocation function q and 'payments'

$$\mathcal{P}_i(\theta_i) = q(\theta_i) v_i(\theta_i) - \mathcal{U}_i(\theta_i).$$

[9] Our results extend immediately to piece-wise constant allocation functions $q = \phi \circ \psi : \Theta \to \{q_1, \cdots, q_n\}$ with full range, where $0 \leq q_1 < \cdots < q_n \leq 1$, $\phi : \mathbb{R} \to \{q_1, \cdots, q_n\}$ is increasing, and $\psi : \Theta \to \mathbb{R}$ satisfies $\nabla_i \psi(\theta) \neq 0$ for all i and θ.

[10] The dictatorial allocation function illustrating the difference between locally robust implementation and ex-post implementation in Section 3 violates regularity. We have been able to define analogous examples with regular, deterministic allocation functions q. However, these examples require value functions that are non-linear in own type and are thus omitted from this paper.

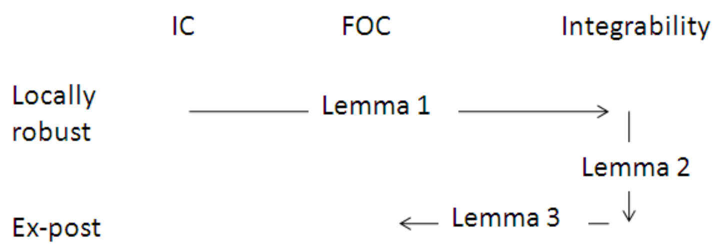

Figure 1: This figure illustrates the logic of our approach: Lemma 1 shows that the vector field $\mathbb{E}\left[q\nabla_i v_i\right]$ must be integrable when (q,p) is incentive compatible. Lemma 2 shows that interim integrability implies ex-post integrability. Lemma 3 reconverts ex-post integrability into a first-order condition.

First, consider a regular stochastic allocation function q. Then, by

$$\nabla\mathcal{P}_i = v_i\nabla q + q\nabla v_i - \nabla\mathcal{U}_i = v_i\nabla q \tag{4}$$

the 'payment' $\mathcal{P}_i$ is constant on $I_i\left(\theta^*\right)$, and we can write $\mathcal{P}_i(\theta_i) = \hat{\mathcal{P}}_i(q(\theta_i))$ for all θ_i in a neighborhood of $I_i\left(\theta^*\right)$. By equation (4) and $\nabla q \neq 0$ we know that $\hat{\mathcal{P}}_i$ is differentiable with $\hat{\mathcal{P}}_i'(q(\theta_i)) = v_i(\theta_i)$. We can interpret this equation as the first-order condition of a 'taxation mechanism' that lets agent i choose allocation q to maximize $v_i\left(\theta_i\right)q - \hat{\mathcal{P}}_i(q)$.[11] As $\hat{\mathcal{P}}_i'(q(\cdot))$ is constant on $I_i\left(\theta^*\right)$, also $v_i(\cdot)$ is constant on $I_i\left(\theta^*\right)$.

Second, consider a regular deterministic allocation function q. By construction, $\mathcal{P}_i$ is constant on path-connected components of $q^{-1}\left(0\right)$ and $q^{-1}\left(1\right)$. As $\mathcal{U}_i\left(\cdot\right)$ is continuous and $q\left(\cdot\right)v_i\left(\cdot\right)$ has a discontinuity of $v_i\left(\theta_i\right)$ at any $\theta_i \in I_i\left(\theta^*\right)$ we can again conclude that $v_i\left(\cdot\right)$ is constant on $I_i\left(\theta^*\right)$. ■

Lemma 3 allows us to write i's value on the boundary as $\rho_i\left(\theta^*_{-i}\right) = v_i\left(\theta_i, \theta^*_{-i}\right)$ for any $\theta_i \in I_i\left(\theta^*\right)$. Next we want to extend ρ_i to a neighborhood of θ^*_{-i}. In doing so we have to be careful in choosing the correct path-connected component of $\{\theta_i \in \Theta_i : (\theta_i, \theta_{-i}) \in I\}$. To this end we fix open neighborhoods $N_i\left(\theta^*_i\right)$ of θ^*_i and $N_{-i}\left(\theta^*_{-i}\right)$ of θ^*_{-i} that are small enough so that for every $\theta_{-i} \in N_{-i}\left(\theta^*_{-i}\right)$ the set

$$I_i^{loc}\left(\theta^*; \theta_{-i}\right) := \{\theta_i \in N_i\left(\theta^*_i\right) : (\theta_i, \theta_{-i}) \in I\}$$

is a path-connected $d_i - 1$-dimensional submanifold of Θ_i, and similarly for every $\theta_i \in N_i\left(\theta^*_i\right)$ the set

$$I_{-i}^{loc}\left(\theta^*; \theta_i\right) := \{\theta_{-i} \in N_{-i}\left(\theta^*_{-i}\right) : (\theta_i, \theta_{-i}) \in I\}$$

[11] Note that this 'taxation mechanism' is generally not incentive compatible (for fixed θ^*_{-i}) as emphasized by the 'regular example' in Section 3 and indicated in Figure 1.

is a path-connected $d_{-i}-1$-dimensional submanifold of Θ_{-i}.

For $\theta_{-i}=\theta^*_{-i}$ the set $I_i^{loc}\left(\theta^*;\theta^*_{-i}\right)$ includes θ^*_i and is thus non-empty. So for fixed neighborhoods $N_i\left(\theta^*_i\right), N_{-i}\left(\theta^*_{-i}\right)$ with the above property, we can choose smaller neighborhoods $M_i\left(\theta^*_i\right)\subseteq N_i\left(\theta^*_i\right)$ and $M_{-i}\left(\theta^*_{-i}\right)\subseteq N_{-i}\left(\theta^*_{-i}\right)$ such that for any $\theta_{-i}\in M_{-i}\left(\theta^*_{-i}\right)$ the set $I_i^{loc}\left(\theta^*;\theta_{-i}\right)$ is non-empty, and for any $\theta_i\in M_i\left(\theta^*_i\right)$ the set $I_i^{loc}\left(\theta^*;\theta_i\right)$ is non-empty.

Thus, for every internal $\theta^*\in I$, we can define 'ex-post transfers' $\rho_i : M_{-i}\left(\theta^*_{-i}\right)\to\mathbb{R}$ and $\rho_{-i}:M_i\left(\theta^*_i\right)\to\mathbb{R}$ by

$$\begin{aligned}\rho_i(\theta_{-i}) &= v_i(\theta_i,\theta_{-i}) \text{ for any } \theta_i\in I_i^{loc}\left(\theta^*;\theta_{-i}\right),\\ \rho_{-i}(\theta_i) &= v_{-i}(\theta_i,\theta_{-i}) \text{ for any } \theta_{-i}\in I_{-i}^{loc}\left(\theta^*;\theta_i\right).\end{aligned}$$

The implicit function theorem implies then that ρ_i is differentiable at θ^*_{-i},[12] and ρ_{-i} is differentiable at θ^*_i.

5.2 The Main Result

With the above preparations in place, we can now show that locally robust implementation imposes similar conditions on value functions as ex-post implementation.

Lemma 4 *Assume $d_i\geq 2$. If a regular allocation function q is locally robust implementable, then there exists an interior $\theta^*\in I$ such that the vectors*

$$\nabla_i v_i(\theta^*_i,\theta_{-i}) \text{ and } \nabla_i(v_{-i}(\theta^*_i,\theta_{-i})-\rho_{-i}(\theta^*_i)) \text{ are parallel for all } \theta_{-i}\in M_{-i}\left(\theta^*_{-i}\right) \text{ with } (\theta^*_i,\theta_{-i})\in I. \tag{5}$$

Proof. For any such θ_{-i} we argue that both these vectors are perpendicular on $I_i^{loc}\left(\theta^*;\theta_{-i}\right)$. For $\nabla_i v_i(\theta^*_i,\theta_{-i})$ this follows from Lemma 3. For $\nabla_i(v_{-i}(\theta^*_i,\theta_{-i})-\rho_{-i}(\theta^*_i))$ it follows by the construction of ρ_{-i}, because $v_{-i}(\cdot,\theta_{-i})-\rho_{-i}(\cdot)$ vanishes on $M_i\left(\theta^*_i\right)$. ■

Lemma 4 is a close analogue to Proposition 3.3 in JMMZ, but there are two differences. First, Proposition 3.3 in JMMZ shows that the two vectors are not only parallel but also point in the same direction. This corresponds to the fact that an ex-post implementable allocation function must be ex-post monotone, while a locally robust implementable allocation function need not be

[12] More specifically, the gradient of ρ_i is given by

$$\begin{aligned}\nabla\rho_i\left(\theta^*_{-i}\right) &= \nabla_{-i}v_i\left(\theta^*_i,\theta^*_{-i}\right)+\frac{\partial_x v_i\left(\theta^*_i,\theta^*_{-i}\right)}{\partial_x\psi\left(\theta^*_i,\theta^*_{-i}\right)}\nabla_{-i}\psi\left(\theta^*_i,\theta^*_{-i}\right) && \text{(deterministic regular)}\\ \nabla\rho_i\left(\theta^*_{-i}\right) &= \nabla_{-i}v_i\left(\theta^*_i,\theta^*_{-i}\right)+\frac{\partial_x v_i\left(\theta^*_i,\theta^*_{-i}\right)}{\partial_x q\left(\theta^*_i,\theta^*_{-i}\right)}\nabla_{-i}q\left(\theta^*_i,\theta^*_{-i}\right) && \text{(stochastic regular)}\end{aligned}$$

where x is any direction in Θ_i for which $\partial_x\psi\left(\theta_i,\theta_{-i}\right)\neq 0$ (resp. $\partial_x q\left(\theta^*_i,\theta^*_{-i}\right)\neq 0$).

ex-post monotone. Second, by focusing on regular allocation functions, we simplify the analysis in comparison to JMMZ; among other things, this rules out case (ii) of Proposition 3.3 in JMMZ.

For any regular allocation function q to be locally robust implementable, Lemma 4 requires the existence of $\theta_i^* \in \Theta_i$ and $\nabla_i \rho_{-i}(\theta_i^*) \in \mathbb{R}^{d_i}$ such that (5) is satisfied. Condition (5) imposes more equations on the value functions v_i and v_{-i} than can be satisfied by the free parameters θ_i^* and $\nabla_i \rho_{-i}(\theta_i^*)$. Dealing with ex-post implementation, JMMZ took a 'highbrow' approach to show a similar result, proving genericity in an infinite-dimensional functional space of sufficiently smooth value functions. Here, we provide an alternative, elementary approach for the finite-dimensional space of bilinear value functions. This approach has the additional advantage that it does not require $\nabla_i v_i \neq 0$ everywhere, which was assumed in JMMZ. We conjecture that the proof of Theorem 1 generalizes, for any $n \in \mathbb{N}$, to the finite-dimensional space of polynomial value functions with degree below n.

Bilinear value functions can uniquely be represented as

$$v_i(\theta) = \theta_i^T (f_i + F_i \theta_{-i}) + g_i + G_i^T \theta_{-i} \tag{6a}$$

$$v_{-i}(\theta) = \theta_i^T (f_{-i} + F_{-i} \theta_{-i}) + g_{-i} + G_{-i}^T \theta_{-i} \tag{6b}$$

with vectors $f_i, f_{-i} \in \mathbb{R}^{d_i}, G_i, G_{-i} \in \mathbb{R}^{d_{-i}}$, constants, $g_i, g_{-i} \in \mathbb{R}$, and matrices $F_i, F_{-i} \in M(d_i \times d_{-i}|\mathbb{R})$, where θ_i^T is the transposed row vector corresponding to column vector θ_i. For the following arguments we introduce some geometric terminology. For vectors $x, y \in \mathbb{R}^n$ we write $x \parallel y$ if x and y are parallel, i.e. there exists $\lambda \in \mathbb{R}$ with $y = \lambda x$ or $x = 0$, and $x \perp y$ if x and y are orthogonal, i.e. $x \cdot y = 0$. Finally, we say that bilinear value functions (6) are *generic* if neither

$$F_i \delta_{-i} \parallel F_{-i} \delta_{-i} \text{ for all } \delta_{-i} \in \mathbb{R}^{d_{-i}} \tag{G1}$$

nor

$$F_{-i}^T \theta_i \parallel G_{-i} \text{ for all } \theta_i \in \Theta_i. \tag{G2}$$

The set of non-generic value functions has Lebesgue-measure zero: Condition (G1) requires the $d_i d_{-i}$-dimensional matrices F_i and F_{-i} to be scalar multiples of each other, and therefore describes a submanifold of co-dimension $d_i d_{-i} - 1$ in $\mathbb{R}^{2d_i d_{-i}} \ni (F_i, F_{-i})$. Condition (G2) requires each of d_i columns of F_{-i}^T to be a scalar multiple of the d_{-i}-dimensional vector G_{-i}, and therefore describes a manifold of co-dimension $d_i(d_{-i} - 1)$ in $\mathbb{R}^{d_i d_{-i}} \times \mathbb{R}^{d_{-i}} \ni (F_{-i}, G_{-i})$.

Theorem 1 *Assume $d_i \geq 2$. For generic bilinear value functions, no regular allocation function is locally robust implementable.*

Proof. Let q be regular and fix $\theta^* = (\theta_i^*, \theta_{-i}^*)$ in the interior of I. With bilinear value functions the geometric condition (5) simplifies to

$$f_i + F_i\theta_{-i} \parallel f_{-i} + F_{-i}\theta_{-i} - \nabla_i\rho_{-i}(\theta_i^*) \text{ for all } \theta_{-i} \in I_{-i}(\theta^*) . \tag{7}$$

We can assume that $I_{-i}(\theta^*)$ is a hyperplane in Θ_{-i}: If $\nabla_{-i}v_{-i}(\theta_i^*, \cdot) \neq 0$ then the level set $\{\theta_{-i}|v_{-i}(\theta_i^*, \theta_{-i}) = v_{-i}(\theta_i^*, \theta_{-i}^*)\}$ is a hyperplane in Θ_{-i} and coincides with $I_{-i}(\theta^*)$ by Lemma 3. Otherwise, if $\nabla_{-i}v_{-i}(\theta_i^*, \cdot) = 0$**,** assumption (G2) implies that the direction of $\nabla_{-i}v_{-i}(\theta_i, \cdot) = F_{-i}^T\theta_i + G_{-i}^T$ varies in θ_i, so there exists $\theta_i \in M_i(\theta_i^*)$ with $\nabla_{-i}v_{-i}(\theta_i, \cdot) \neq 0$. By our regularity assumption $I_{-i}(\theta_i, \theta_{-i}^*)$ is non-empty and condition (7) holds when we replace θ_i^* by θ_i. Subject to this replacement, we can assume that $\nabla_{-i}v_{-i}(\theta^*) \neq 0$ and that $I_{-i}(\theta^*)$ is indeed a hyperplane.

As F_i, F_{-i} are linear functions of θ_{-i}, we rewrite (7) as

$$f_i + F_i\theta_{-i}^* + \alpha F_i(\theta_{-i}^* - \theta_{-i}) \parallel f_{-i} + F_{-i}\theta_{-i}^* - \nabla_i\rho_{-i}(\theta_i^*) + \alpha F_{-i}(\theta_{-i}^* - \theta_{-i}) \text{ for all } \theta_{-i} \in I_{-i}(\theta^*) \text{ and } \alpha \in [0,1] .$$

We then apply Lemma 5 (see Appendix) to $y = F_i(\theta_{-i}^* - \theta_{-i})$ and $y' = F_{-i}(\theta_{-i}^* - \theta_{-i})$ to obtain

$$F_i(\theta_{-i}^* - \theta_{-i}) \parallel F_{-i}(\theta_{-i}^* - \theta_{-i}) \text{ for all } \theta_{-i} \in I_{-i}(\theta^*) .$$

Alternatively

$$F_i\delta_{-i} \parallel F_{-i}\delta_{-i} \text{ for all } \delta_{-i} \perp \nabla_{-i}v_{-i}(\theta_i^*, \cdot) \tag{8}$$

as the gradient $\nabla_{-i}v_{-i}(\theta_i^*, \cdot)$ is the normal vector on the hyperplane $I_{-i}(\theta^*)$.

By regularity, (8) holds for all θ_i in a neighborhood of θ_i^* and by condition (G2) the direction of $\nabla_{-i}v_{-i}(\theta_i, \cdot)$ varies linearly in θ_i. So by Lemma 6 (see Appendix) there exists a single parameter $\lambda \in \mathbb{R}$ such that $\lambda F_i\delta_{-i} = F_{-i}\delta_{-i}$ for all $\delta_{-i} \in \mathbb{R}^{d_{-i}}$, or $F_i \equiv 0$. This contradicts (G1), finishing the proof. ■

Discussion of Regularity: We rely on the regularity assumption in two ways. First we use it whenever we assume that the sets $I_i(\theta^*)$ are 'well-behaved' as in the proof of Lemma 3. This use of regularity is an innocuous way to keep the analysis clean. Second, we use the assumption $\nabla_i\psi \neq 0$ or $\nabla_i q \neq 0$ when we argue that $\rho_i(\cdot)$ is differentiable, or even well-defined. This use of regularity is more substantial because it rules out dictatorial choice functions $q = q(\theta_i)$ where a small change of θ_i can tip the allocation from $q = 0$ to $q = 1$ for all θ_{-i}. For such dictatorial function q the boundary $I_{-i}(\theta^*) \subset \Theta_{-i}$ does not exist for any θ^*.[13]

We complement Theorem 1 by showing that, generically, dictatorial allocation functions are not locally robust implementable either. Indeed, consider any non-constant dictatorial allocation

[13] The same issue arises in JMMZ. There we treat 'irregular' allocation functions in part (ii) of Proposition 3.3., and parts (iii) and (iv) of Proposition 4.3.

function $q : \Theta_i \to [0,1]$. By monotonicity (2), the allocation $q(\cdot)$ must be increasing in the same direction as i's expected value $\mathbb{E}_{\pi_i^*}[v_i(\cdot,\theta_{-i})]$, so generically there exists a hyperplane $I \subset \Theta_i$ that separates lower values of $q(\theta_i)$ from higher values of $q(\theta_i)$. Lemma 3 implies that $v_i(\cdot,\theta_{-i})$ is constant on I for all θ_{-i}, so $\nabla_i v_i(\cdot,\theta_{-i}) = F_i\theta_{-i}$ does not depend on θ_{-i}. This is clearly a non-generic condition.

6 Conclusion

In this paper we have studied a notion of locally robust implementation that takes an intermediate position between Bayesian implementation and robust implementation. Specifically, the agent's type space is some neighborhood of a Bayesian type space, modeling slight uncertainty of the planner about agents' beliefs. While such a type space may seem much closer to a classical Bayesian type space than to, say, the universal type space, we show that for rich environments with multi-dimensional payoff types, locally robust implementation is still an overly demanding concept. Theorem 1 shows that, for generic bilinear values, no regular allocation function is locally robust implementable. This result parallels and reinforces the negative result on ex-post implementation in JMMZ.

One way to interpret this negative result is that in many payoff environments even local robustness is too demanding when applied to social choice functions. One should be then ready to allow for the implementation of social choice correspondences in which the outcome may depend (at least slightly) on agents' beliefs. This calls for a redirection of the robust mechanism design agenda towards the implementation of social choice correspondences - a direction actually present in Bergemann and Morris [1], but less so in the subsequent literature. In particular, following the spirit of the local perturbations considered in this paper, it would make sense to uncover the kind of local perturbations of beliefs and the baseline social choice functions for which a nearby outcome can be ensured. Some insights along these lines are developed by Meyer-ter-Vehn and Morris [19] who show that, for open sets of value functions and for arbitrary belief spaces, the planner is able to achieve belief-dependent, but close-to-optimal outcomes (see also Madarasz and Prat [18] in a multi-product monopoly setup for a related investigation).

7 Appendix

Lemma 5 *Let x, x', y, y' be vectors in $\mathbb{R}^n$. If $(x+\alpha y) \parallel (x'+\alpha y')$ for three or more values $\alpha \in \mathbb{R}$, then $y \parallel y'$.*

Proof. First note that $x \parallel x'$ iff $x_j x'_k \neq x'_j x_k$ for all coordinates $j,k \in \{1,\cdots,n\}$. Now, if $(x_j+\alpha y_j)(x'_k+\alpha y'_k) = (x_k+\alpha y_k)(x'_j+\alpha y'_j)$ for three or more values of α, then the coefficients on the α^2-terms must coincide, so $y_j y'_k = y_k y'_j$. ■

Lemma 6 *Let x, x', y, y' be vectors in $\mathbb{R}^n$. If $\alpha x + (1-\alpha) y \parallel \alpha x' + (1-\alpha) y'$ for all convex combinations of x, y and x', y', then there exists a single parameter $\lambda \in \mathbb{R}$ such that for all linear combinations of x, y and x', y' we have $\lambda (\alpha x + \beta y) = \alpha x' + \beta y'$ or $\alpha x + \beta y = 0$.*

Proof. Assume that x and y (and x' and y') are linearly independent (otherwise the proof is obvious), so $\lambda_1 x = x'$ and $\lambda_0 y = y'$. Then

$$\alpha x' + (1-\alpha) y' = \alpha \lambda_1 x + (1-\alpha) \lambda_0 y = \lambda_0 (\alpha x + (1-\alpha) y) + \alpha (\lambda_1 - \lambda_0) x$$

can only be parallel to $\alpha x + (1-\alpha) y$ if $\lambda_0 = \lambda_1$. ■

References

[1] Bergemann, D. and S. Morris (2005), "Robust Mechanism Design", *Econometrica* **73**, 1771-1813.

[2] Bergemann, D. and S. Morris (2009), "Robust Implementation in Direct Mechanisms", *Review of Economic Studies* **76**, 1175-1204.

[3] Bergemann, D., S. Morris and S. Takahashi (2012), "Efficient Auctions and Interdependent Types", *mimeo*

[4] Bikhchandani, S. (2006): "Ex-Post Implementation in Environments with Private Goods," *Theoretical Economics* **1**, 369-393

[5] Clarke E. (1971), "Multipart Pricing of Public Goods," *Public Choice* **8**, 19-33.

[6] Crémer, J. and R. McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions", *Econometrica* **56**, 1247-1257

[7] Dasgupta, P. and E. Maskin (2000): "Efficient Auctions", *Quarterly Journal of Economics* **115**, 341-388.

[8] Gibbard, A. (1973), "Manipulation of Voting Schemes", *Econometrica* **41**, 587-601.

[9] Groves T. (1973), "Incentives in Teams," *Econometrica* **41**, 617-631.

[10] Harsanyi, J. (1967/1968), "Games with Incomplete Information Played by Bayesian Agents," *Management Science*, **14**, 159–182, 320–334, 486–502.

[11] Jehiel, P., B. Moldovanu and E. Stacchetti (1999), "Multidimensional Mechanism Design for Auctions with Externalities", *Journal of Economic Theory* **85**, 258-293

[12] Jehiel, P. and B. Moldovanu (2001) "Efficient Design with Interdependent Valuations", *Econometrica* **69**, 1237-1259.

[13] Jehiel, P., M. Meyer-ter-Vehn, B. Moldovanu and W.R. Zame (2006), "The Limits of Ex-Post Implementation", *Econometrica* **74**, 585-610

[14] Jehiel, P., Meyer-ter-Vehn, M. and Moldovanu, B. (2008): "Ex-post Implementation and Preference Aggregation via Potentials", *Economic Theory* **37**(3), 469 - 490.

[15] Johnson, S., J. W. Pratt and R. Zeckhauser (1990): "Efficiency Despite Mutually Payoff-Relevant Private Information: The Finite Case," *Econometrica* **58**, 873-900.

[16] Ledyard, J. (1978): "Incentive Compatibility and Incomplete Information," *Journal of Economic Theory* **18**, 171-189.

[17] Lopomo, G., L. Rigotti and C. Shannon (2009), "Uncertainty in Mechanism Design", *mimeo*

[18] Madarasz, K. and A. Prat (2010), "Screening with an Approximate Type Space", *mimeo.*

[19] Meyer-ter-Vehn, M. and S. Morris (2011), "The Robustness of Robust Implementation", *Journal of Economic Theory* **146**, 2093–2104

[20] Milgrom, P. and I. Segal (2002), "Envelope Theorems for Arbitrary Choice Sets", *Econometrica* **70**, 583-601

[21] Neeman, Z. (2004), "The Relevance of Private Information in Mechanism Design", *Journal of Economic Theory* **117**, 55-77

[22] Oury, M. and O. Tercieux (2012), "Continuous Implementation", *Econometrica* forthcoming

[23] Roberts, K. (1979): "The Characterization of Implementable Choice Rules" in *Aggregation and Revelation of Preferences*, ed. Laffont J.J., North Holland.: 895-921

[24] Rochet, J. C. (1987), "A necessary and sufficient condition for rationalizability in a quasi-linear context,", *Journal of Mathematical Economics* **16**, 191-200.

[25] Satterthwaite, M. (1975), "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting procedures and Social Welfare Functions", *Journal of Economic Theory* **10**, 187-217.

[26] Vickrey, W. (1961), "Counterspeculation, Auctions and Competitive Sealed Tenders," *Journal of Finance* **16**, 8-37.