Locally Robust Implementation and its Limits

Philippe Jehiel, Moritz Meyer-ter-Vehn, Benny Moldovanu*

November 8, 2010

Abstract

We introduce 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 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 (1988). To address this issue, the robust mechanism design literature models an agent's belief as part of her private type, and requires a robust mechanism to be incentive compatible for a range of agents' beliefs so as to reflect the designer's uncertainty about these beliefs (see Bergemann and Morris (2005)[1]).

Much of the robust mechanism design literature, e.g. Bergemann and Morris (2009), takes the above criticism of the Bayesian paradigm to the opposite extreme, and assumes that the designer knows nothing at all about agents' beliefs. When the designer 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 (1978) and by Bergemann and Morris (2005), respectively.

Dominant-strategy and ex-post implementation are overly restrictive in important settings. In private value environments with unrestricted preference types and three of more social alternatives, Gibbard (1973) and Satterthwaite (1975) show that only dictatorial choice functions are

*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.

[1] See also Neeman (2004) for an earlier investigation on mechanism design with a focus on payoff and belief types.

implementable in dominant strategies. Restricting attention to quasi-linear utilities gives rise to more positive results when values are private, as shown by Vickrey (1961), Clarke (1971), Groves (1973) and Roberts (1979). 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 (2000) and Jehiel and Moldovanu (2001)).[2] But, Jehiel, Meyer-ter-Vehn, Moldovanu and Zame (2006), JMMZ henceforth, show that only trivial allocation functions are implementable when payoff types are multi-dimensional and the interdependent value functions are generic. 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 the 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, generically, the system does not have a solution.

The connection between our main present result and the impossibility result of JMMZ is instructive. As for many other implementation concepts, locally robust mechanisms need to satisfy a monotonicity condition and an integrability condition (commonly known 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 allocation function. Conservativeness imposes an equality constraint on the cross-partials of the value functions and the allocation function. This equality must hold ex-post if it holds in

[2] Bickhchandani (2006) and Jehiel, Meyer-ter-Vehn and Moldovanu (2008) display some possibility results in non-generic framework with multi-dimensional signals.

expectation for an open set of beliefs.

The concept of *locally robust incentive compatibility* defined in this paper is very similar to the *optimal incentive compatibility* defined in Lopomo, Rigotti and Shannon (2009) in order to study uncertainty averse agents. For payoff environments more general than the quasi-linear environment considered in this paper, Lopomo et al. 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 (1987). 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 Lopomo et al. and JMMZ. Locally robust implementation is also similar to the *continuous implementation*, as defined in Oury and Tercieux (2009) 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. It would *a fortiori* hold in more complex environments (i.e., involving more than two alternatives, or more than two agents). Specifically, there are two alternatives $x \in \{0, 1\}$, 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 Bernoulli utility functions of the form $u_i = xv_i(\theta_i, \theta_{-i}) - p_i$, where p_i is a monetary payment by the agent. We assume that $v_i(\theta_i, \theta_{-i})$ is continuously differentiable, and that the d_i-dimensional gradient of v_i with respect to θ_i is strictly positive in every dimension, i.e. $\nabla_i v_i(\theta_i, \theta_{-i}) \gg 0$.

The Interim Type Space: The *baseline beliefs* are given by arbitrary functions $\pi_i^* : \Theta_i \to \Delta(\Theta_{-i})$. 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. Moreover, this common distribution π^* could allow for correlation between θ_i and θ_{-i} as in the work of Crémer and McLean (1988).[3]

[3] 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, and allowing for this would only make the notation more cumbersome.

Agent i's type space $T_i \subset \Theta_i \times \Delta(\Theta_{-i})$ is a neighborhood of the graph $\{(\theta_i, \pi_i^*(\theta_i)) | \theta_i \in \Theta_i\}$, where Θ_i is endowed with the standard Euclidian topology, $\Delta(\Theta_{-i})$ with the weak topology, and $\Theta_i \times \Delta(\Theta_{-i})$ with the product topology.

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 think of 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 for every θ_i there exists $\varepsilon > 0$, an ε-ball of payoff types $B_\varepsilon(\theta_i) = \{\theta_i' \in \Theta_i | \left\|\theta_i - \theta_i'\right\|_\infty < \varepsilon\}$, and an ε-ball of belief types $B_\varepsilon(\pi_i^*(\theta_i)) = \{(1-\varepsilon)\pi_i^*(\theta_i) + \varepsilon\pi_i : \pi_i \in \Delta(\Theta_{-i})\}$, such that:[4]

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

Implementation: The planner wants to implement a deterministic allocation $q : \Theta \to \{0,1\}$ as a function of payoff types θ.[5] An allocation function q is *locally robust implementable* if there exists 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. if

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

for all $\theta = (\theta_i, \theta_{-i}), \theta' = \left(\theta_i', \theta_{-i}\right), t = (\theta_i, \pi_i, \theta_{-i}, \pi_{-i}), t' = (\theta_i', \pi_i', \theta_{-i}, \pi_{-i})$.

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.

[4] To see that any weak neigborhood of $\pi_i^*(\theta_i) \in \Delta(\Theta_{-i})$ includes a ball $B_\varepsilon(\pi_i^*(\theta_i))$ for some $\varepsilon > 0$, consider a sequence of measures $(\pi_{i,n})_{n\in\mathbb{N}}$ with $\pi_{i,n} \in B_{\varepsilon_n}(\pi_i^*(\theta_i))$ where $\varepsilon_n \to 0$. We need to show that $\pi_{i,n}$ converges to π_i^* in the weak topology, i.e. to show that $\limsup \pi_{i,n}(C) \leq \pi_i^*(C)$ for all closed sets $C \subseteq \Theta_{-i}$. By definition we have $\pi_{i,n} = (1-\varepsilon_n)\pi_i^* + \varepsilon_n \pi_{i,n}'$ for some $\pi_{i,n}' \in \Delta(\Theta_{-i})$. Thus, $\pi_{i,n}(C) - \pi_i^*(C) = \varepsilon_n(\pi_{i,n}'(C) - \pi_i^*(C)) \leq \varepsilon_n$ for any closed set $C \subseteq \Theta_{-i}$, and so $\limsup(\pi_{i,n}(C) - \pi_i^*(C)) \leq \limsup \varepsilon_n = 0$.

[5] The arguments presented here generalize in a straightforward manner to stochastic allocations $q \in [0,1]$.

3 Locally Robust vs. Ex-Post Implementation

We point out here by example 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 the work of Lopomo, Rigotti and Shannon (2009).

Fix one agent i, and define payoff type spaces $\Theta_i = [0,1]$ and $\Theta_{-i} = \{-1, 1\}$ and baseline beliefs $\pi_i^*(\theta_i)$ by $\Pr(\theta_{-i} = 1) = 0.8$ for all θ_i.[6] Agent i's type space allows for different beliefs over θ_{-i}, and is given by $T_i = \Theta_i \times [0.7, 0.9]$. Thus agent i's value is increasing in own type for $\theta_{-i} = 1$ (as $\nabla_i v_i(\cdot, \theta_{-i}) \equiv 1$), and is decreasing in own type for $\theta'_{-i} = -1$ (as $\nabla_i v_i\left(\cdot, \theta'_{-i}\right) \equiv -1$). Note that i's value is increasing in expectation in own type for any belief $\pi_i \in [0.7, 0.9]$ since

$$\mathbb{E}_{\pi_i}[\nabla_i v_i(\cdot, \theta_{-i})] = 2\pi_i - 1 > 0.$$

Consider a dictatorial allocation function that only takes i's payoff type into account, i.e. q is defined by a cutoff $\theta_i^* \in (0,1)$ such 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 $\theta'_{-i} = -1$ 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 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 every belief $\pi_i \in [0.7, 0.9]$ we have $\mathbb{E}_{\pi_i}[\nabla_i v_i(\cdot, \theta_{-i})] > 0$, so that the agent indeed chooses $q = 1$ when $\theta_i > \theta_i^*$ and $q = 0$ when $\theta_i < \theta_i^*$.

This positive result for locally robust implementation and the contrast to ex-post implementation is due to the core idea of local robustness, that agents' beliefs are known to be close to some baseline. It is not due to artificial weaknesses in the solution concept since: (1) the mechanism

[6] While these type spaces do not fit all the technical assumptions of Section 2, this merely aids in keeping the example as simple as possible, and does not drive the substantial results.

(q,p) has payments defined as a function of payoff types alone; (2) every undominated strategy of type (θ_i,π_i) with $\theta_i \neq \theta_i^*$ will lead to outcome $q(\theta_i)$, so that (q,p) fully implements q for almost all payoff types; and (3) incentive compatibility is maintained on any larger type space with the same first-order beliefs because i's higher-order beliefs do not matter in mechanism (q,p).

4 Monotonicity and Integrability

As a first step towards the main result, we follow Jehiel, Moldovanu, Stacchetti (1999), and show that implementable allocation functions must satisfy locally robust versions of monotonicity and integrability. In deriving these necessary conditions we only exploit agent i's ability to misreport his payoff type for any given belief type, but ignore her ability to misreport her belief type.[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*

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

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*

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

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

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

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 $B_\varepsilon\left(\theta_i\right)$.

[7] Ignoring such misreports of beliefs does not significantly weaken the IC constraints, because it is relatively easy to elicit beliefs by a continuous version of the log-scoring rule (see for example Johnson et al. (1990)). The discrete version of this rule punishes agent i with the payment rule $p_i(\pi_i,t_{-i})=-\log(\pi_i(t_{-i}))$ when i reports belief π_i and others report type t_{-i} . If i's true belief is π_i and others truthfully report t_{-i}, the benefit from misreporting her belief as π_i' is negative

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

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

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:

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

Adding up the above two inequalities yields (2).

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

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

be the expected utility of type θ_i when reporting $\widehat{\theta}_i$. Let $\theta_i^*(\theta_i) \in \arg\max_{\widehat{\theta}_i} f_{\pi_i}(\widehat{\theta}_i, \theta_i)$ be any selection from the arg max-correspondence. Then the multi-dimensional version of Corollary 1 in Milgrom and Segal (2002) states that

$$U_{i,\pi_i}\left(\theta_i'\right) - U_{i,\pi_i}\left(\theta_i\right) = \int_{\theta_i}^{\theta_i'} \nabla_i f_{\pi_i}(\theta_i^*(s), s) \cdot ds.$$

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 ∇_i into the expectation in f_{π_i}. ■

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 example of Section 3, the basic reason behind the positive result on locally robust implementation and the contrast to ex-post implementation is that monotonicity can be satisfied for all close-by beliefs $\pi_i \in B_\varepsilon(\pi_i^*) \subseteq \Delta(\Theta_{-i})$, while at the same time be violated for other far-away beliefs $\pi_i' \in \Delta(\Theta_{-i})$. This is indeed the case in the example in Section 3 where $\mathbb{E}_{\pi_i}[\nabla_i v_i\left(\cdot, \theta_{-i}\right)] > 0$ for all beliefs $\pi_i \in [0.7, 0.9]$ but $\mathbb{E}_{\pi_i'}[\nabla_i v_i\left(\cdot, \theta_{-i}\right)] = -1$ for belief $\pi_i' = 0$ that puts probability one on type $\theta_{-i} = -1$. The reason is that when inequality (2) is strict for some belief π_i^*, then it can still be 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 $\mathbb{E}_{\pi_i}[q\left(\cdot, \theta_{-i}\right)$ $\nabla_i v_i\left(\cdot, \theta_{-i}\right)]$ be conservative translates into an equality constraint (on cross derivatives), which needs also to be satisfied ex-post.

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

$$\int_{\theta_i}^{\theta_i'} q\left(s,\theta_{-i}\right)\nabla_i v_i\left(s,\theta_{-i}\right)\cdot ds$$

has the same value for all differentiable paths $s : [0,1] \to \Theta_i$ with $s(0) = \theta_i$ and $s(1) = \theta_i'$.

Proof. By definition of type space T_i, there exists for every payoff type θ_i some $\varepsilon > 0$ such that $B_\varepsilon\left(\theta_i\right) \times B_\varepsilon\left(\pi_i^*(\theta_i)\right) \subseteq T_i$. If the vector field $q\left(\cdot,\theta_{-i}'\right)\nabla_i v_i\left(\cdot,\theta_{-i}'\right)$ is not conservative for some θ_{-i}', then the vector field $\mathbb{E}_{\pi_i'}\left[q\left(\cdot,\theta_{-i}\right)\nabla_i v_i\left(\cdot,\theta_{-i}\right)\right]$ for prior $\pi_i' = (1-\varepsilon)\pi_i + \varepsilon\mathbb{I}_{\theta_{-i}'}$ would not be conservative either, where $\mathbb{I}_{\theta_{-i}'} \in \Delta(\Theta_{-i})$ is the belief that puts probability one on θ_{-i}'.[8] But, by Lemma 1, the vector field $\mathbb{E}_{\pi_i}\left[q\left(\cdot,\theta_{-i}\right)\nabla_i v_i\left(\cdot,\theta_{-i}\right)\right]$ is conservative on $B_\varepsilon\left(\theta_i\right)$ for all $\pi_i \in B_\varepsilon\left(\pi_i^*(\theta_i)\right)$, yielding a contradiction. ■

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. These conditions jointly imply a differential equation on value functions that has generically no solution.

5.1 The Regularity Assumption

The proof of Theorem 1 relies on geometric arguments on the boundary $I \subset \Theta$ that separates the areas $q^{-1}(0)$ and $q^{-1}(1)$ in the payoff type space where different allocations are chosen. In order to facilitate these arguments we focus on *regular* allocation functions:

Definition 1 *An allocation function $q : \Theta \to \{0,1\}$ is* regular *if both allocations 0 and 1 are chosen for interior types θ, and if q factors through a smooth and responsive score function ψ, i.e. there exists $\psi : \Theta \to \mathbb{R}$ continuously differentiable with $\nabla_i\psi, \nabla_{-i}\psi \gg 0$ such that $q(\theta) = 0$ if $\psi(\theta) < 0$ and $q(\theta) = 1$ if $\psi(\theta) > 0$.*

For a regular allocation function q , the sets $q^{-1}(0)$ and $q^{-1}(1)$ are separated by the $d_i + d_{-i} - 1$-dimensional manifold $I = \psi^{-1}(0)$. Moreover, the projection $pr_i(I) = \{\theta_i \in \Theta_i | \exists\theta_{-i} : (\theta_i,\theta_{-i}) \in I\}$ has a non-empty interior, and for every θ_i^* in the interior of $pr_i(I)$ the "slice" $I\left(\theta_i^*\right) = \{\theta_{-i} \in \Theta_{-i} : (\theta_i^*,\theta_{-i}) \in I\}$ is a $d_{-i} - 1$-dimensional manifold in Θ_{-i}.

Lemma 3 *Assume that q is regular and that $q\left(\cdot,\theta_{-i}^*\right)\nabla_i v_i\left(\cdot,\theta_{-i}^*\right)$ is a conservative vector field. Then i's value function $v_i\left(\cdot,\theta_{-i}^*\right)$ must be constant on $I\left(\theta_{-i}^*\right)$.*

[8] This argument is an elementary version of the proof of Theorem 1 in Lopomo, Rigotti and Shannon (2009).

Proof. For any regular q such that $q\left(\cdot,\theta_{-i}^{*}\right)\nabla_{i}v_{i}\left(\cdot,\theta_{-i}^{*}\right)$ is conservative, a potential function of $q\left(\cdot,\theta_{-i}^{*}\right)\nabla_{i}v_{i}\left(\cdot,\theta_{-i}^{*}\right)$ must be constant on the interior of $\{\theta_i : q(\theta_i,\theta_{-i}^{*})=0\}$, and equal to $v_i\left(\cdot,\theta_{-i}^{*}\right)+const.$ on the interior of $\{\theta_i : q(\theta_i,\theta_{-i}^{*})=1\}$. As potential functions are continuous, $v_i\left(\cdot,\theta_{-i}^{*}\right)$ must be constant on the boundary $I\left(\theta_{-i}^{*}\right)$ between $\{\theta_i|q(\theta_i,\theta_{-i}^{*})=0\}$ and $\{\theta_i|q(\theta_i,\theta_{-i}^{*})=1\}$. ■

Lemma 3 is closely related to the taxation principle for ex-post implementation, see e.g. Lemma 2.1 in JMMZ. That principle states that for fixed payoff types of others θ_{-i}^{*}, an ex-post incentive compatible mechanism chooses an allocation $q(\cdot,\theta_{-i}^{*})$ to maximize i's value minus some tax, $q(v_i(\theta_i,\theta_{-i}^{*})-\rho_i(\theta_{-i}^{*}))$. For continuous value functions this implies that $v_i(\cdot,\theta_{-i}^{*})$ is constant and equal to $\rho_i(\theta_{-i}^{*})$ on the boundary $I\left(\theta_{-i}^{*}\right)$.[9] Following this analogy, we define for any regular, locally robust implementable allocation function q agent i's *virtual ex-post transfer:*

$$\rho_i(\theta_{-i}^{*})=v_i(\theta_i,\theta_{-i}^{*})\text{ for any }\theta_i\in I\left(\theta_{-i}^{*}\right).$$

The implicit function theorem implies then that $\rho_i : pr_{-i}(I)\rightarrow\mathbb{R}$ is differentiable on the interior of $pr_{-i}(I)$.[10]

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. The following lemma is the analogue to Proposition 3.3 case (i) in JMMZ.

Lemma 4 *If a regular allocation function q is locally robust implementable, then there exists θ_i^{*} in the interior of $pr_i(I)$ such that the vectors*

$$\nabla_i v_i(\theta_i^{*},\theta_{-i})\text{ and }\nabla_i(v_{-i}(\theta_i^{*},\theta_{-i})-\rho_{-i}(\theta_i^{*}))\text{ are parallel for all }\theta_{-i}\in I\left(\theta_i^{*}\right). \tag{4}$$

Proof. For any $\theta_{-i}\in I\left(\theta_i^{*}\right)$ we argue that both these vectors are perpendicular on $I(\theta_{-i})$. 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 $I(\theta_{-i})$. ■

[9] The differential way of stating that $v_i(\cdot,\theta_i^{*})$ and $q(\cdot,\theta_i^{*})$ have the same level sets, is that the allocation function q must respect i's incentives by trading off changes in different dimensions of i's payoff type with the same marginal rate of information substitution as agent i's value function.

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

$$\nabla_{-i}\rho_i\left(\theta_{-i}^{*}\right)=\nabla_{-i}v_i\left(\theta_i,\theta_{-i}^{*}\right)+\frac{\partial_x v_i\left(\theta_i,\theta_{-i}^{*}\right)}{\partial_x\psi\left(\theta_i,\theta_{-i}^{*}\right)}\nabla_{-i}\psi\left(\theta_i,\theta_{-i}^{*}\right)$$

where θ_i is any element in the interior of $I\left(\theta_{-i}^{*}\right)$, and x is any direction in Θ_i for which $\partial_x\psi\left(\theta_i,\theta_{-i}^{*}\right)\neq 0$.

Thus for any regular allocation function q to be locally robust implementable, there must exist $\theta_i^* \in \Theta_i$ and $\nabla_i \rho_{-i}(\theta_i^*) \in \mathbb{R}^{d_i}$ such that (4) is satisfied.

It is instructive to note that there are two differences between the above result and Proposition 3.3 in JMMZ. First, Proposition 3.3 in JMMZ shows that the two vectors are are not only parallel but also point in the same direction (co-directional). This is because an ex-post implementable allocation function must be ex-post monotone, while a locally robust implementable allocation function need not be ex-post monotone (see Section 3). 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.

When the dimension of $I(\theta_i^*)$ is greater than zero, i.e. when $d_{-i} \geq 2$, condition (4) imposes a continuum of independent equations on the partial derivatives of v_i and v_{-i}. Generically, these cannot be satisfied by the choice of only $2d_i$ free parameters θ_i^* and $\nabla_i \rho_{-i}(\theta_i^*)$.

Formally, let $r > (2d_1+1)/(d_1-1)$, $d = d_1 + d_2$ and $m = dr + 2d_1 + 1 - 2d_1 r$. Let $C^m(S, \mathbb{R}^2)$ be the Banach space of maps $\Theta \to \mathbb{R}^2$ that admit an m-times continuously differentiable extension to an open neighborhood of Θ, equipped with the topology of uniform convergence of maps and m derivatives. Let $\mathcal{H} \subset C^m(\Theta, \mathbb{R}^2)$ be the open subset consisting of those pairs of relative valuation functions $(v_1, v_2) \in C^m(\Theta, \mathbb{R}^2)$ for which $\nabla_i v_i \gg 0$ everywhere.

Theorem 1 *Assume that the individual payoff type spaces have dimensions $d_i \geq 2$, $i = 1, 2$. Then there is a residual and finitely prevalent subset $\mathcal{G} \subset \mathcal{H}$ such that for every $(v_1, v_2) \in \mathcal{G}$, no regular choice function is locally robust implementable.*

Proof. See the proof of Theorem 4.2 in JMMZ. ■

Remark 2 *We rely on the regularity assumption, Definition 1, in two ways. First we use it whenever we assume that $q^{-1}(0)$, $q^{-1}(1)$ or I are 'well-behaved' as in the proof of Lemma 3. This use of regularity is an innocuous way to keep the analysis simple. Second, we use the assumption $\nabla_i \psi, \nabla_{-i} \psi \gg 0$ 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}, so that the boundary $I(\theta_i) \subset \Theta_{-i}$ does not exist for any θ_i. 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. We can also show here that, generically, dictatorial allocation functions are not locally robust implementable either. Indeed, consider any dictatorial allocation function $q : \Theta_i \to \{0, 1\}$ that chooses both allocations for interior types θ_i. In our above terminology, this implies that the boundary $I(\theta_{-i})$ that separates $\{\theta_i | q(\theta_i, \theta_{-i}) = 0\}$ from $\{\theta_i | q(\theta_i, \theta_{-i}) = 1\}$ is independent of θ_{-i}, i.e. $I(\theta_{-i}) = I_i$. Lemma 3 implies that $v_i(\cdot, \theta_{-i})$ is constant on I_i for all θ_{-i}. For a fixed set I_i, this is obviously a* *strong assumption and Theorem 4.2 in JMMZ implies that this condition is generically not satisfied for any set I_i.*

6 Conclusion

In this note we have introduced 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 designer 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, generically, 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 (2005), but less so in the subsequent literature. In particular, following the spirit of the local perturbations considered in this note, 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 (2010) 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 (2010) in a multi-product monopoly setup for a related investigation).

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

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

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

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

[7] Gibbard, A. (1973), “Manipulation of Voting Schemes”, *Econometrica* **41**, 587-601.

[8] Groves T. (1973), “Incentives in Teams,” *Econometrica* **41**, 617-631.

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

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

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

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

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

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

[15] Lopomo, G., L. Rigotti, C. Shannon (2009), “Uncertainty in Mechanism Design”, *mimeo*

[16] Meyer-ter-Vehn, M. and S. Morris (2010), “The Robustness of Robust Implementation”, *Princeton ETC working paper 002*

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

[18] Milgrom, P., I. Segal (2002), “Envelope Theorems for Arbitrary Choice Sets”, *Econometrica* **70**, 583-601

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

[20] Oury, M., O. Tercieux (2009), “Continuous Implementation”, mimeo

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

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

[23] 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.

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