RATIONALIZABLE IMPLEMENTATION

By

Dirk Bergemann, Stephen Morris and Olivier Tercieux

May 2009
Revised January 2010

COWLES FOUNDATION DISCUSSION PAPER NO. 1697R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Rationalizable Implementation*

Dirk Bergemann[†] Stephen Morris[‡] Olivier Tercieux[§]

January 2010

Abstract

We consider the implementation of social choice functions under complete information in rationalizable strategies. A strict (and thus stronger) version of the monotonicity condition introduced by Maskin (1999) is necessary under the solution concept of rationalizability. Assuming the social choice function is responsive (i.e., it never selects the same outcome in two distinct states), we show that it is also sufficient under a mild "no worst alternative" condition. In particular, no economic condition is required. We also discuss how our results extend when the social choice function is not responsive.

KEYWORDS: Implementation, Complete Information, Rationalizability, Maskin Monotonicity.
JEL CLASSIFICATION: C79, D82

*We are grateful to Navin Kartik, Ludovic Renou and Roberto Serrano for comments on an earlier draft. This research is supported by NSF Grant #SES-0851200.

[†]Department of Economics, Yale University, New Haven, CT 06511, U.S.A., dirk.bergemann@yale.edu
[‡]Department of Economics, Princeton University, Princeton, NJ 08544, U.S.A., smorris@princeton.edu
[§]Paris School of Economics, Paris, France, tercieux@pse.ens.fr

1 Introduction

We consider the implementation of social choice functions under complete information in rationalizable strategies. We say that a social choice function f is rationalizably implemented if there exists a mechanism such that every rationalizable strategy profile leads to the realization of the social choice function f. A priori, implementation in rationalizable strategies does not require the existence of a (pure or mixed) Nash equilibrium that leads to the realization of f, and hence this implementation notion is neither stronger nor weaker than that of Nash implementation. However, we establish that a strict (and thus stronger) version of the monotonicity condition shown by Maskin (1999) to be necessary for Nash implementation is necessary under the more stringent solution concept of rationalizability. Assuming the social choice function is responsive (i.e., it never picks the same outcome in two distinct states), we show that it is also sufficient under a "no worst alternative" (NWA) condition. In particular, no economic condition is required.

We are able to obtain this strong result because - like much of the classical implementation literature - we allow infinite mechanisms (including "integer games"); and - unlike the classical implementation literature - we allow for stochastic mechanisms.

In earlier work (Bergemann and Morris (2008), (2009)), two of us established necessary and sufficient conditions for "robust implementation" in incomplete information environments. There we showed that a social choice function f can be Bayesian equilibrium implemented for all possible beliefs and higher order beliefs if and only if f is implementable under an incomplete information version of rationalizability. The results here are obtained by refining and further developing the rationalizability arguments for the complete information environment. We can establish stronger necessary and sufficient conditions than in the incomplete information environment. We can also dispense with an economic condition on the environment. In turn, we establish necessary conditions and sufficient conditions almost equivalent to Nash equilibrium implementation when the social choice function is responsive. The augmented mechanism which establishes the sufficiency result permits each agent to propose a menu of allocations. This construction already appeared in Maskin (1999) and Maskin and Sjostrom (2004) to establish complete information implementation in the presence of mixed strategies. The sufficiency arguments for Nash equilibrium implementation typically rely on a no-veto property of the social choice function. In contrast, we use a weak condition, introduced as "no worst alternative" by Cabrales and Serrano (2008), to establish the sufficiency argument. This condition requires that in state θ and for every agent i, the social choice $f(\theta)$ is not the worst alternative among all possible allocations. The no worst alternative property plays a role in our proof that is quite distinct from the no veto property in the classic Nash equilibrium results. The no worst alternative property guarantees that in the augmented mechanism, any report in state θ in which an agent expresses his disagreement with the remaining agents cannot be a rationalizable report. By contrast, the no veto property guaranteed that if an agent were to express his disagreement, then further disagreement by other agents would only be possible in equilibrium if it would lead to the same equilibrium allocation as prescribed by $f(\theta)$.

These results narrow an open question in the literature. The existing literature shows that Maskin monotonicity is necessary for Nash implementation in any mechanism (even if stochastic mechanisms are allowed[1]). Abreu and Matsushima (1992) shows that if implementation is made easier by (i) requiring only virtual implementation; and (ii) imposing a weak domain restriction ruling out identical preferences; then implementation is always possible even if it is made harder by (iii) requiring finite mechanisms; and (iv) requiring the stronger solution concept of rationalizability. Our result shows that it is possible to exactly implement a social choice function, in rationalizable strategies, even if domain restriction (ii) fails, as long as infinite, stochastic, mechanisms are allowed.

2 Setup

The *environment* consists of a collection of I agents (we write $\mathcal{I}$ for the set of agents); a finite set of possible states Θ; a countable set of pure allocations Z (we write $Y \equiv \Delta(Z)$ for the set of lotteries on Z); and, to each state, we associate for each player i a von Neumann-Morgenstern utility function $u_i : Z \times \Theta \to \mathbb{R}$, extended to lotteries as $u_i : Y \times \Theta \to \mathbb{R}$ with

$$u_i(y, \theta) = \sum_{z \in Z} y_z u_i(z, \theta).$$

Thus at two distinct states θ and θ', all agents can have the same ordinal preferences; this contrasts with some of the literature that associates a state with a profile of ordinal preferences (e.g. Maskin (1999)). A *mechanism* $\mathcal{M}$ is given by $\mathcal{M} = \left((M_i)_{i=1}^{I}, g\right)$, where each M_i is countable, $M = M_1 \times \cdots \times M_I$ and $g : M \to Y$.

The environment and the mechanism together describe a game of complete information for each $\theta \in \Theta$. We will use (correlated) rationalizability as a solution concept.[2] Our formal definition will coincide with the standard definition with finite or compact message spaces. But we will also allow infinite, non-compact, message spaces; in this case, our definition is equivalent to one introduced

[1] In such a case, Maskin monotonicity (that is usually defined on the set of pure allocations) has to be stated on the set of lotteries on pure allocations.

[2] The original definition of rationalizability of Bernheim (1984) and Pearce (1984) required agents' conjectures over their opponents' play to be independent. We follow the convention of some of the recent literature (e.g., Osborne and Rubinstein (1994)) in using "rationalizability" for the correlated version of rationalizability (see Brandenburger and Dekel (1987) for an early definition and discussion). Our results do not rely on the use of the correlated version of rationalizability.

in Lipman (1994). Let a message set profile $S = (S_1, ..., S_I)$, where each $S_i \in 2^{M_i}$, and we write $\mathcal{S}$ for the collection of message set profiles. The collection $\mathcal{S}$ is a lattice with the natural ordering of set inclusion: $S \leq S'$ if $S_i \subseteq S'_i$ for all i. The largest element is $\overline{S} = (M_1, ..., M_I)$. The smallest element is $\underline{S} = (\varnothing, \varnothing, ..., \varnothing)$.

We define an operator $b^\theta : \mathcal{S} \to \mathcal{S}$ to iteratively eliminate never best responses with $b^\theta = \left(b_1^\theta, ...b_i^\theta, ..., b_I^\theta\right)$ and b_i^θ is defined by:

$$b_i^\theta(S) = \left\{ m_i \in M_i \;\middle|\; \begin{array}{l} \text{there exists } \lambda_i \in \Delta(M_{-i}) \text{ such that} \\ (1)\ \lambda_i(m_{-i}) > 0 \Rightarrow m_j \in S_j \text{ for each } j \neq i, \\ (2)\ m_i \in \underset{m'_i \in M_i}{\arg\max} \sum\limits_{m_{-i} \in M_{-i}} \lambda_i(m_{-i})\, u_i(g(m'_i, m_{-i}), \theta), \end{array} \right\}.$$

We observe that b^θ is increasing by definition: i.e., $S \leq S' \Rightarrow b^\theta(S) \leq b^\theta(S')$. By Tarski's fixed point theorem, there is a largest fixed point of b^θ, which we label $S^{\mathcal{M},\theta}$. Thus (i) $b^\theta\left(S^{\mathcal{M},\theta}\right) = S^{\mathcal{M},\theta}$ and (ii) $b^\theta(S) = S \Rightarrow S \leq S^{\mathcal{M},\theta}$. If $m_i \in S_i^{\mathcal{M},\theta}$, we say that message m_i is rationalizable in (the complete information game parameterized by) state θ.

We can also construct the fixed point $S^{\mathcal{M},\theta}$ by starting with $\overline{S}$ - the largest element of the lattice - and iteratively applying the operator b^θ. If the message sets are finite, we have

$$S_i^{\mathcal{M},\theta} \triangleq \bigcap_{n \geq 0} b_i^\theta\left(\left[b^\theta\right]^n(\overline{S})\right).$$

In this case, the solution concept is equivalent to iterated deletion of strictly dominated strategies (see Brandenburger and Dekel (1987)). But if the mechanism $\mathcal{M}$ is infinite, transfinite induction may be necessary to reach the fixed point.[3] We will also sometimes use the following notation

$$S_{i,k}^{\mathcal{M},\theta} \triangleq b_i^\theta\left(\left[b^\theta\right]^{k-1}(\overline{S})\right),$$

again using transfinite induction if necessary. Thus $S_i^{\mathcal{M},\theta}$ is the set of messages surviving (transfinite) iterated deletion of never best responses. It is possible to show formally that $S_i^{\mathcal{M},\theta}$ is the set of messages that agent i might send consistent with common certainty of rationality and the fact that payoffs are given by θ (Lipman (1994)). Finally, we will say that a message set profile $S = (S_1, ..., S_I)$ has the best response property in state θ if $S \subseteq b^\theta(S)$, or equivalently, if for each player i and message $m_i \in S_i$, there exists $\lambda_i \in \Delta(M_{-i})$ such that $\lambda_i(m_{-i}) > 0 \Rightarrow m_j \in S_j$ for each $j \neq i$, and

$$m_i \in \underset{m'_i \in M_i}{\arg\max} \sum_{m_{-i} \in M_{-i}} \lambda_i(m_{-i})\, u_i\left(g\left(m'_i, m_{-i}\right), \theta\right).$$

[3] Lipman (1994) contains a formal description of the transfinite induction required. As he notes "we remove strategies which are never a best reply, taking limits where needed".

It is easy to check that if S has the best response property in state θ, then $S \subseteq S^{\mathcal{M},\theta}$.

Now a *social choice function* (SCF) f is given by $f : \Theta \to Y$. Mechanism $\mathcal{M}$ *implements* f *in rationalizable strategies* if there exists $\mathcal{M}$ such that, for all θ, $S^{\mathcal{M},\theta} \neq \varnothing$ and $m \in S^{\mathcal{M},\theta} \Rightarrow g(m) = f(\theta)$. SCF f is *implementable in rationalizable strategies* if there exists $\mathcal{M}$ such that $\mathcal{M}$ implements f in rationalizable strategies. The definition of rationalizable implementation does not require the existence of a (pure or mixed) Nash equilibrium that leads the realization of the social choice function f. Hence, a priori, rationalizable implementation need not be stronger (neither weaker) than Nash implementation. However, in the next Section, we provide necessary conditions and sufficient conditions for rationalizable implementation almost equivalent to Nash implementation.

3 Main Result

We first recall the definition of Maskin monotonicity restricted to social choice functions:

Definition 1 (Maskin Monotonicity)
Social choice function f satisfies Maskin monotonicity if

1. $f(\theta) = f(\theta')$ *whenever*

$$u_i(f(\theta), \theta) \geq u_i(y, \theta) \Rightarrow u_i(f(\theta), \theta') \geq u_i(y, \theta') \text{ for all } i \text{ and } y;$$

or, equivalently,

2. $f(\theta) \neq f(\theta')$ *implies*

$$u_i(f(\theta), \theta) \geq u_i(y, \theta) \text{ and } u_i(y, \theta') > u_i(f(\theta), \theta') \text{ for some } i \text{ and } y.$$

The latter condition states that in case the desired alternative differs at state θ and θ', there must exist at least one agent who, if the true state were θ' and she expected other agents to claim the state is θ, could be offered a reward y that would give her a strict incentive to "report" the deviation of other agents, where the reward y would not tempt her if the true state was in fact θ i.e. she would have a (weak) incentive to "report truthfully". The strengthening of Maskin monotonicity we will use, reinforces the latter statement, requiring that the reward y gives a strict incentive to "report truthfully" if the true state were θ.

Definition 2 (Strict Maskin Monotonicity)
Social choice function f satisfies strict Maskin monotonicity if

1. $f(\theta) = f(\theta')$ *whenever*

$$u_i(f(\theta), \theta) > u_i(y, \theta) \Rightarrow u_i(f(\theta), \theta') \geq u_i(y, \theta') \text{ for all } i \text{ and } y; \tag{1}$$

   or, equivalently,

2. $f(\theta) \neq f(\theta')$ *implies*

$$u_i(f(\theta), \theta) > u_i(y, \theta) \text{ and } u_i(y, \theta') > u_i(f(\theta), \theta') \text{ for some } i \text{ and } y. \tag{2}$$

Maskin monotonicity, which is necessary for Nash implementation, is weaker than strict Maskin monotonicity. We show in the following proposition that strict Maskin monotonicity (and hence Maskin monotonicity) is necessary for rationalizable implementation.

Proposition 1 (Necessary Conditions)
If f is implementable in rationalizable strategies, then f satisfies strict Maskin monotonicity.

Proposition 1 is a consequence of the following Lemma. In words, it states that, given a social choice function f, if θ and θ' satisfy condition (1) in the definition of strict Maskin monotonicity and, in addition, f is implementable by a mechanism $\mathcal{M}$, then the set of rationalizable message profiles must be the same in state θ and θ'.

Lemma 1
Pick θ and θ' satisfying condition (1). If mechanism $\mathcal{M} = \left((M_i)_{i=1}^I, g\right)$ implements f in rationalizable strategies, then we have $S^{\mathcal{M},\theta} = S^{\mathcal{M},\theta'}$.

Proof. Pick θ and θ' satisfying condition (1) and fix any mechanism $\mathcal{M} = \left((M_i)_{i=1}^I, g\right)$ that implements f in rationalizable strategies.

We first show that $S^{\mathcal{M},\theta} \subseteq S^{\mathcal{M},\theta'}$. Because $b^\theta\left(S^{\mathcal{M},\theta}\right) = S^{\mathcal{M},\theta}$, $S^{\mathcal{M},\theta}$ has the best response property in state θ (i.e., for each player i and all $m_i \in S_i^{\mathcal{M},\theta}$, there exists $\lambda_i^{m_i,\theta} \in \Delta(M_{-i})$ such that $\lambda_i^{m_i,\theta}(m_{-i}) > 0 \Rightarrow m_j \in S_j^{\mathcal{M},\theta}$ for each $j \neq i$), and

$$\sum_{m_{-i} \in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i, m_{-i}), \theta) \geq \sum_{m_{-i} \in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i', m_{-i}), \theta) \tag{3}$$

for all $m_i' \in M_i$. We want to show that m_i is also a best response against $\lambda_i^{m_i,\theta}$ in state θ'. Since i and $m_i \in S_i^{\mathcal{M},\theta}$ have been fixed arbitrarily, this will prove that $S^{\mathcal{M},\theta}$ has the best response property in state θ' and so that $S^{\mathcal{M},\theta} \subseteq S^{\mathcal{M},\theta'}$ as claimed. Note first that for any m_{-i} such that $\lambda_i^{m_i,\theta}(m_{-i}) > 0$, $m_{-i} \in S_{-i}^{\mathcal{M},\theta}$ and so because $m_i \in S_i^{\mathcal{M},\theta}$, we have $g(m_i, m_{-i}) = f(\theta)$. Thus,

$$\sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i, m_{-i}), \theta') = u_i(f(\theta), \theta') = \sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i', m_{-i}), \theta') \quad (4)$$

for all $m_i' \in S_i^{\mathcal{M},\theta}$. In addition, we claim that

$$\sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i, m_{-i}), \theta) = u_i(f(\theta), \theta) > \sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i', m_{-i}), \theta)$$

for any $m_i' \notin S_i^{\mathcal{M},\theta}$. By (3), the above is true with a weak inequality. Now if an equality were to hold, some $m_i' \notin S_i^{\mathcal{M},\theta}$ would be a best response against $\lambda_i^{m_i,\theta}$ in state θ and so $(\{m_i'\} \cup S_i^{\mathcal{M},\theta}) \times S_{-i}^{\mathcal{M},\theta}$ would have the best response property in state θ implying that $m_i' \in S_i^{\mathcal{M},\theta}$ which is false by assumption. Now we know that

$$u_i(f(\theta), \theta) > u_i(y, \theta) \Rightarrow u_i(f(\theta), \theta') \geq u_i(y, \theta') \text{ for all } i \text{ and } y,$$

and so applying this to the lotteries $y \triangleq \sum_{m_{-i}} \lambda_i^{m_i,\theta}(m_{-i})g(m_i', m_{-i})$, we get that

$$\sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i, m_{-i}), \theta') = u_i(f(\theta), \theta') \geq \sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i', m_{-i}), \theta') \quad (5)$$

for any $m_i' \notin S_i^{\mathcal{M},\theta}$. Finally, (4) and (5) ensure that m_i is also a best response against $\lambda_i^{m_i,\theta}$ in state θ'.

Now, let us show that $S^{\mathcal{M},\theta} \supseteq S^{\mathcal{M},\theta'}$. Since f is implementable in rationalizable strategies by $\mathcal{M}$ and $S^{\mathcal{M},\theta} \subseteq S^{\mathcal{M},\theta'}$, we have $f(\theta) = f(\theta')$. Take any player i and any $m_i^* \in S_i^{\mathcal{M},\theta'}$ we show that $m_i^* \in S_i^{\mathcal{M},\theta}$. Pick any $m_i \in S_i^{\mathcal{M},\theta}$ and $\lambda_i^{m_i,\theta} \in \Delta(M_{-i})$ satisfying $\lambda_i^{m_i,\theta}(m_{-i}) > 0 \Rightarrow m_j \in S_j^{\mathcal{M},\theta}$ for all $j \neq i$, and

$$u_i(f(\theta), \theta) = \sum_{m_{-i}} \lambda_i^{m_i,\theta}(m_{-i})\, u_i(g(m_i, m_{-i}), \theta) \geq \sum_{m_{-i}} \lambda_i^{m_i,\theta}(m_{-i})\, u_i\left(g\left(m_i', m_{-i}\right), \theta\right) \quad (6)$$

for any $m_i' \in M_i$. Note that since $S^{\mathcal{M},\theta} \subseteq S^{\mathcal{M},\theta'}$, we have that $\lambda_i^{m_i,\theta}(m_{-i}) > 0 \Rightarrow m_j \in S_j^{\mathcal{M},\theta'}$ for all $j \neq i$; in addition, $m_i^* \in S_i^{\mathcal{M},\theta'}$ and so $g(m_i^*, m_{-i}) = f(\theta') = f(\theta)$ for any m_{-i} such that $\lambda_i^{m_i,\theta}(m_{-i}) > 0$. Hence

$$\sum_{m_{-i}} \lambda_i^{m_i,\theta}(m_{-i})\, u_i(g(m_i^*, m_{-i}), \theta) = u_i(f(\theta), \theta) = \sum_{m_{-i}} \lambda_i^{m_i,\theta}(m_{-i})\, u_i(g(m_i, m_{-i}), \theta), \quad (7)$$

and thus (6) and (7) together show that m_i^* is a best response against $\lambda_i^{m_i,\theta}$ in state θ which proves that $m_i^* \in S_i^{\mathcal{M},\theta}$, as claimed. ■

It is clear that Proposition 1 is obtained as a corollary of Lemma 1.

Oury and Tercieux (2009) have shown that Maskin monotonicity is a necessary condition for "continuous" partial implementation of a social choice function, where "continuous" means that the direct mechanism itself must work for types that are close to the complete information types in the product topology. They also show that full implementation in rationalizable strategies is necessary. Hence, an alternative way to prove the necessity of Maskin monotonicity would be to use this latter result and Proposition 1.

We need two extra conditions for the sufficiency result.

Definition 3 (Responsive Social Choice Function)
Social choice function f is responsive if $\theta \neq \theta' \Rightarrow f(\theta) \neq f(\theta')$.

The notion of responsiveness requires that the social choice function "responds" to a change in the state with a change in the social allocation.

Definition 4 (No Worst Alternative)
Social choice function f satisfies "no worst alternative" (NWA) if, for each i and θ, there exists $\underline{y}_i(\theta)$ such that

$$u_i(f(\theta), \theta) > u_i\left(\underline{y}_i(\theta), \theta\right). \tag{8}$$

Property NWA requires that an agent never gets his worst outcome under the social choice function. The NWA property appears in Cabrales and Serrano (2008) as a sufficient condition to guarantee implementation in best-response dynamics. Given the set of allocations $\{\underline{y}_i(\theta)\}_{\theta\in\Theta}$, we define the average allocation $\underline{y}_i$ of this set by setting

$$\underline{y}_i \triangleq \frac{1}{\#\Theta}\sum_{\theta\in\Theta}\underline{y}_i(\theta). \tag{9}$$

Note that under NWA, for all θ and all i, there exists $y_i(\theta)$ such that

$$u_i(y_i(\theta), \theta) > u_i(\underline{y}_i, \theta); \tag{10}$$

this can be established by defining $y_i(\theta)$ as follows:

$$y_i(\theta) \triangleq \frac{1}{\#\Theta}\sum_{\hat{\theta}\neq\theta}\underline{y}_i(\hat{\theta}) + \frac{1}{\#\Theta}f(\theta).$$

We also define the average allocation $\underline{y}$ of the set $\{\underline{y}_i\}_{i\in\mathcal{I}}$ by setting

$$\underline{y} \triangleq \frac{1}{I}\sum_{i\in\mathcal{I}}\underline{y}_i.$$

Here again, we note that under NWA, for all θ and all i, there exists $y_i^*(\theta)$ such that

$$u_i(y_i^*(\theta),\theta) > u_i(\underline{y},\theta), \tag{11}$$

where the above inequality clearly holds after defining $y_i^*(\theta)$ as follows:

$$y_i^*(\theta) \triangleq \frac{1}{I}\sum_{j\neq i}\underline{y}_j + \frac{1}{I}y_i(\theta).$$

We now construct an auxiliary set of allocations, denoted by $\{z_i(\theta,\theta')\}_{\theta,\theta'}$, which uses the existence of the allocations $\{\underline{y}_i(\theta)\}_{\theta\in\Theta}$. The allocations $\{z_i(\theta,\theta')\}_{\theta,\theta'}$ are going to appear in the canonical mechanism to be defined shortly where they guarantee the existence of better response for agent i should the remaining agents choose to misreport the true state. In particular, the following Lemma establishes that for agent i the allocation $z_i(\theta,\theta')$ represents an improvement if the true state is θ but the other agents misreport it to be θ'. It also establishes that $z_i(\theta,\theta')$ would not constitute an improvement relative to $f(\theta')$ if the true state were indeed θ'.

Lemma 2

If social choice function f satisfies "no worst alternative" (NWA) then for each player i, there exists a collection of lotteries $\{z_i(\theta,\theta')\}_{\theta,\theta'}$ such that for all θ,θ' :

$$u_i(f(\theta'),\theta') > u_i(z_i(\theta,\theta'),\theta'), \tag{12}$$

and for $\theta\neq\theta'$:

$$u_i(z_i(\theta,\theta'),\theta) > u_i(z_i(\theta',\theta'),\theta). \tag{13}$$

Proof. Based on the allocations $\{\underline{y}_i(\theta)\}_{\theta\in\Theta}$ from Definition 4, we define our collection of lotteries as follows. First, for all θ':

$$z_i(\theta',\theta') \triangleq (1-\varepsilon)\,\underline{y}_i(\theta') + \varepsilon\underline{y}_i,$$

with $\underline{y}_i$ as defined in (9), and for all θ,θ' with $\theta\neq\theta'$:

$$z_i(\theta,\theta') \triangleq (1-\varepsilon)\,\underline{y}_i(\theta') + \frac{\varepsilon}{\#\Theta}\left(\sum_{\hat{\theta}\neq\theta}\underline{y}_i(\hat{\theta}) + f(\theta)\right).$$

By NWA and the finiteness of the state space Θ, we can find a sufficiently small, but positive, $\varepsilon > 0$ such that for all θ and θ': $u_i(f(\theta'),\theta') > u_i(z_i(\theta,\theta'),\theta')$ which establishes inequality (12). Now we observe that the only difference between $z_i(\theta',\theta')$ and $z_i(\theta,\theta')$ is the fact that the lottery $\underline{y}_i(\theta)$

is replaced by the lottery $f(\theta)$. But now by NWA, this is clearly increasing the expected utility of agent i in state θ, and hence we have for all θ, θ' with $\theta \neq \theta'$:

$$u_i\left(z_i\left(\theta, \theta'\right), \theta\right) > u_i\left(z_i\left(\theta', \theta'\right), \theta\right),$$

which establishes the strict inequality (13). ■

We establish the sufficient conditions for implementation in rationalizable strategies by means of a canonical mechanism. The canonical mechanism shares many basic features with the implementation mechanism suggested by Maskin and Sjostrom (2004) to establish complete information implementation in the presence of mixed strategies, and is a modification of the original mechanism suggested by Maskin (1999). The aforementioned allocations $\left\{z_i\left(\theta, \theta'\right)\right\}_{\theta, \theta'}$ appear in the mechanism if agent i reports a state θ different from the reported state θ' by all the other agents. In this case, the allocation $z_i\left(\theta', \theta'\right)$ is chosen with positive probability, yet this probability can be lowered by a suitable message of agent i and be replaced by a more favorable allocation $z_i\left(\theta, \theta'\right)$. In the Proposition below, we show that Maskin monotonicity together with NWA are sufficient for rationalizable implementation. The fact that we do not refer to strict Maskin monotonicity in this statement may seem surprising given that in Proposition 1 we showed that strict Maskin monotonicity is a necessary condition for rationalizable implementation. This is due to the simple fact that under NWA, strict Maskin monotonicity and Maskin monotonicity are equivalent.[4]

Proposition 2 (Sufficient Conditions)
If $I \geq 3$, f is responsive, satisfies Maskin monotonicity and NWA, then f is implementable in rationalizable strategies.

Proof. We establish the result by constructing an implementing mechanism $\mathcal{M} = (M, g)$. First, recall that by definition of Maskin monotonicity, for all θ and θ' such that $f(\theta) \neq f\left(\theta'\right)$, there exist i and $y(\theta, \theta') \in Y$ with $u_i(f(\theta), \theta) \geq u_i\left(y(\theta, \theta'), \theta\right)$ and $u_i\left(y(\theta, \theta'), \theta'\right) > u_i\left(f(\theta), \theta'\right)$. We define the following finite set of lotteries:

$$\mathcal{Y} = \left\{z_i\left(\theta, \theta'\right)\right\}_{i, \theta, \theta'} \cup \left\{y\left(\theta, \theta'\right)\right\}_{\{\theta, \theta' \mid f(\theta) \neq f(\theta')\}} \cup \{y_i^*(\theta)\}_{i, \theta},$$

where the collection $\left\{z_i\left(\theta, \theta'\right)\right\}_{i, \theta, \theta'}$ has been defined in Lemma 2 while the collection $\{y_i^*(\theta)\}_{i, \theta}$ has been established in (11).

[4] To see this just note that if f is Maskin monotonic then $f(\theta) \neq f(\theta')$ implies the existence of some i and y satisfying $u_i(f(\theta), \theta) \geq u_i(y, \theta)$ and $u_i(y, \theta') > u_i(f(\theta), \theta')$. Now, under NWA, there exists $\underline{y}_i(\theta)$ such that $u_i(f(\theta), \theta) > u_i\left(\underline{y}_i(\theta), \theta\right)$. Now if one sets $\tilde{y} = \varepsilon \underline{y}_i(\theta) + (1 - \varepsilon) y$, for ε small enough we get $u_i(f(\theta), \theta) > u_i(\tilde{y}, \theta)$ and $u_i(y, \theta') > u_i(f(\theta), \theta')$, showing that f is strict Maskin monotonic.

Each agent i sends a message $m_i = \left(m_i^1, m_i^2, m_i^3, m_i^4\right)$, where $m_i^1 \in \Theta$, $m_i^2 \in \mathbb{Z}_+$, $m_i^3 : \Theta \to \mathcal{Y}, m_i^4 \in \mathcal{Y}$. The third component of the message profile will allow agent i to suggest an allocation $m_i^3(\theta)$ contingent on all the other agents $j \neq i$ reporting $m_j^1 = \theta$. The outcome function will make use of the "uniformly worse outcome" defined earlier by $\underline{y}$. Now the outcome $g(m)$ is determined by the following rules:

Rule 1: If $m_i^1 = \theta$ and $m_i^2 = 1$ for all i, pick $f(\theta)$.
Rule 2: If there exists $i \in I$ – called the deviating player – such that $\left(m_j^1, m_j^2\right) = (\theta, 1)$ for all $j \neq i$ and $\left(m_i^1, m_i^2\right) \neq (\theta, 1)$, then we go to two subrules:
(i): if $u_i(f(\theta), \theta) \geq u_i\left(m_i^3(\theta), \theta\right)$, pick $m_i^3(\theta)$ with probability $1 - 1/(m_i^2 + 1)$ and $z_i(\theta, \theta)$ with probability $1/(m_i^2 + 1)$;
(ii): if $u_i(f(\theta), \theta) < u_i\left(m_i^3(\theta), \theta\right)$, pick $z_i(\theta, \theta)$ with probability 1.
Rule 3: In all other cases, we identify a pivotal agent i by requiring that $m_i^2 \geq m_j^2$ for all $j \in I$ and that if for $j \neq i$, $m_i^2 = m_j^2$, then $i < j$. The rule then requires that with probability $1 - 1/(m_i^2 + 1)$ we pick m_i^4, and with probability $1/(m_i^2 + 1)$ we pick $\underline{y}$.
Claim 1. It is never a best reply for agent i to send a message with $m_i^2 > 1$ (i.e., $m_i \in b_i^\theta\left(\overline{S}\right) \Rightarrow m_i^2 = 1$).
Proof of Claim 1. We proceed by contradiction and suppose that $m_i = \left(m_i^1, m_i^2, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$ and $m_i^2 > 1$. Then for any profile of messages m_{-i} that player i's opponents may play, (m_i, m_{-i}) will trigger either Rule 2 or Rule 3. But in this case, whatever agent i's beliefs $\lambda_i \in \Delta(M_{-i})$ about the other agents' messages, his payoff can be increased by modifying m_i appropriately, in particular by increasing the integer choice from m_i^2. To see this, denote the set of messages of all agents excluding i in which Rule 2 is triggered by:

$$M_{-i}^2 \triangleq \left\{m_{-i} \in M_{-i} \left| m_j^1 = \theta' \text{ and } m_j^2 = 1 \text{ for some } \theta' \text{ for all } j \neq i \right.\right\}, \tag{14}$$

and the set of messages of all agents excluding i in which Rule 3 is triggered as the complement set:

$$M_{-i}^3 \triangleq M_{-i} \setminus M_{-i}^2. \tag{15}$$

Suppose first that agent i has a belief $\lambda_i \in \Delta(M_{-i})$ under which Rule 3 is triggered with positive probability, so that $\lambda_i\left(M_{-i}^3\right) > 0$. Note that if agent i plays m_i, with strictly positive probability $\underline{y}$ is provided. Hence, because from (11), $y_i^*(\theta) \in \mathcal{Y}$ is such that $u_i(y_i^*(\theta), \theta) > u_i(\underline{y}, \theta)$, i's expected utility conditional on Rule 3 satisfies:

$$\sum_{m_{-i} \in M_{-i}^3} \lambda_i(m_{-i})\, u_i(g(m_i, m_{-i}), \theta) < \sum_{m_{-i} \in M_{-i}^3} \lambda_i(m_{-i}) \max_{y \in \mathcal{Y}} u_i(y, \theta).$$

Now, if i deviates to $\widehat{m}_i = (\widehat{m}_i^1, \widehat{m}_i^2, \widehat{m}_i^3, \widehat{m}_i^4)$ where $\widehat{m}_i^4 \in \arg\max_{y \in \mathcal{Y}} u_i(y, \theta)$, it is easily checked that i's expected utility conditional on Rule 3 tends to

$$\sum_{m_{-i} \in M_{-i}^3} \lambda_i(m_{-i}) \max_{y \in \mathcal{Y}} u_i(y, \theta)$$

as $\widehat{m}_i^2$ tends to infinity. Thus, player i can always improve his expected payoff conditional on Rule 3 by deviating from m_i to $\widehat{m}_i$ and announcing $\widehat{m}_i^2$ large enough.

Now suppose that agent i believes that Rule 2 will be triggered with positive probability, so that $\lambda_i\left(M_{-i}^2\right) > 0$. We again consider a deviation to $\widehat{m}_i = (\widehat{m}_i^1, \widehat{m}_i^2, \widehat{m}_i^3, \widehat{m}_i^4)$ and observe that the choice of $\widehat{m}_i^4$ does not affect the outcome of the mechanism conditional on Rule 2. We also note that for any $m_{-i} \in M_{-i}^2$ such that $\lambda_i(m_{-i}) > 0$, (m_i, m_{-i}) does not trigger Rule 2(ii). Indeed, if it were the case, we would have $u_i(g(m_i, m_{-i}), \theta) = u_i(z_i(m_{-i}^1, m_{-i}^1), \theta)$. We have to distinguish two cases: whether players $j \neq i$ send message θ or not. First, consider the case where $m_{-i}^1 \neq \theta$.[5] Now, player i could change m_i to $\widehat{m}_i$ having $\widehat{m}_i^3(m_{-i}^1) = z_i(\theta, m_{-i}^1)$ and keeping m_i unchanged otherwise. By Lemma 2, $u_i\left(f\left(m_{-i}^1\right), m_{-i}^1\right) > u_i(z_i(\theta, m_{-i}^1), m_{-i}^1) = u_i\left(\widehat{m}_i^3\left(m_{-i}^1\right), m_{-i}^1\right)$, and so by construction of the mechanism, $(\widehat{m}_i, m_{-i})$ now triggers Rule 2(i). Again using Lemma 2 and the fact that $m_{-i}^1 \neq \theta$, we get

$$\begin{aligned} & u_i(g(m_i, m_{-i}), \theta) = u_i(z_i(m_{-i}^1, m_{-i}^1), \theta) \\ < \; & \left(1 - 1/(m_i^2 + 1)\right) u_i(z_i(\theta, m_{-i}^1), \theta) + \left(1/(m_i^2 + 1)\right) u_i(z_i(m_{-i}^1, m_{-i}^1), \theta) = u_i(g(\widehat{m}_i, m_{-i}), \theta). \end{aligned}$$

Hence, the expected utility of player i would strictly increase, which yields the contradiction. Consider the second case where $m_{-i}^1 = \theta$, player i could change m_i to $\widehat{m}_i$ having $\widehat{m}_i(m_{-i}^1) = f(\theta)$ and keeping m_i unchanged otherwise. It is clear that by construction of the mechanism, $(\widehat{m}_i, m_{-i})$ now triggers Rule 2(i). Since Lemma 2 gives us

$$\begin{aligned} & u_i(g(m_i, m_{-i}), \theta) = u_i(z_i(m_{-i}^1, m_{-i}^1), \theta) \\ < \; & \left(1 - 1/(m_i^2 + 1)\right) u_i(f(\theta), \theta) + \left(1/(m_i^2 + 1)\right) u_i(z_i\left(\theta, \theta\right), \theta) = u_i(g(\widehat{m}_i, m_{-i}), \theta), \end{aligned}$$

the expected utility of player i would strictly increase, which here again yields a contradiction. So now we know that for any $m_{-i} \in M_{-i}^2$ such that $\lambda_i(m_{-i}) > 0$, (m_i, m_{-i}) does not trigger Rule 2 (ii). Using a similar reasoning, it is easily shown that for any $m_{-i} \in M_{-i}^2$ such that $\lambda_i(m_{-i}) > 0$, we must have $u_i(m_i^3(m_{-i}^1), \theta) > u_i(z_i(m_{-i}^1, m_{-i}^1), \theta)$, hence the expected payoff conditional on Rule 2 from m_i :

$$\sum_{m_{-i} \in M_{-i}^2} \lambda_i(m_{-i}) \left[\left(1 - 1/(m_i^2 + 1)\right) u_i(m_i^3(m_{-i}^1), \theta) + (1/(m_i^2 + 1)) u_i(z_i(m_{-i}^1, m_{-i}^1), \theta)\right]$$

[5] We sometimes abuse notations and write $m_{-i}^1 = \theta$ whenever $m_j^1 = \theta$ for all $j \neq i$.

is strictly increasing in m_i^2. It follows that the choice of $\widehat{m}_i$ with $\widehat{m}_i^2$ large and strictly larger than m_i^2 strictly improves the expected utility of agent i if either Rule 2 or 3 is triggered, which yields the desired contradiction.

Claim 2. $\left(\theta, 1, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$ for all i, θ, m_i^3, m_i^4.

Proof of Claim 2. Suppose that player i in state θ puts probability 1 on each other agent j sending a message of the form $(\theta, 1, m_j^3, m_j^4)$. If player i announces a message of the form $\left(\theta, 1, m_i^3, m_i^4\right)$, he gets payoff $u_i\left(f\left(\theta\right), \theta\right)$. If he announces a message not of this form, the outcome is determined by Rule 2. Since by Lemma 2, $u_i(z_i(\theta, \theta), \theta) < u_i(f(\theta), \theta)$, it is clear that by construction of the mechanism, his payoff from invoking Rule 2 is bounded above by $u_i\left(f\left(\theta\right), \theta\right)$.

Claim 3. If $m_i = \left(\theta', 1, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$, then $\theta' = \theta$.

Proof of Claim 3. Suppose $m_i = \left(\theta', 1, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$. Given the message m_i, we can define the set of messages of the remaining agents which trigger Rule 1, 2 or 3, respectively. In particular, we define M_{-i}^1 as the set of $m_{-i} \in M_{-i}$ such that (m_i, m_{-i}) triggers Rule 1. Similarly, $M_{-i}^{2,i}$ is defined as the set of $m_{-i} \in M_{-i}$ such that (m_i, m_{-i}) triggers Rule 2 where player i is the deviating player. Now consider a given belief λ_i of agent i. If $\lambda_i\left(\left\{m_{-i} \in M_{-i}^1\left(m_i\right)\right\}\right) = 0$, then Rule 2 or 3 will be triggered with probability one. Although, Rule 2 can now be triggered with a "deviating player" being different of i, it is easily checked that a similar argument as in Claim 1 applies and so the message m_i cannot be a best reply by agent i. Suppose now that the belief λ_i of agent i is such that:

$$0 < \lambda_i\left(\left\{m_{-i} \in M_{-i}^1\right\}\right) < 1. \tag{16}$$

While we still argue that agent i can strictly increase his expected utility by selecting an integer $\widehat{m}_i^2 > 1$, we observe that a complication arises as with λ_i given by (16), a choice of $\widehat{m}_i^2 > 1$ leads from an allocation determined by Rule 1 to an allocation determined by Rule 2, and hence the realization of an unfavorable allocation $\underline{y}$ with positive probability. But now we observe that by selecting $\widehat{m}_i$ such that:

$$\widehat{m}_i^3\left(\hat{\theta}\right) = \begin{cases} f\left(\theta'\right), & \text{if} \quad \hat{\theta} = \theta', \\ m_i^3\left(\hat{\theta}\right), & \text{if} \quad \text{otherwise}, \end{cases}$$

$\widehat{m}_i^4 \in \arg\max_{y \in \mathcal{Y}} u_i(y, \theta)$ and by choosing an integer $\widehat{m}_i^2$ sufficiently large, the small loss in Rule 2 can always be offset by a gain in Rule 3 relative to the allocation achieved under $g\left(m_i, m_{-i}\right)$. More formally, for $m_i = \left(\theta', 1, m_i^3, m_i^4\right)$, since $0 < \lambda_i\left(\left\{m_{-i} \in M_{-i}^1\right\}\right) < 1$ and since – as claimed before – for all $m_{-i} \in M_{-i}^{2,i}$ such that $\lambda_i(m_{-i}) > 0$, $u_i(m_i^3(m_{-i}^1), \theta) > u_i(z_i(m_{-i}^1, m_{-i}^1), \theta)$, i's expected payoff from playing m_i is strictly lower than

$$\sum_{m_{-i} \in M_{-i}^1} \lambda_i\left(m_{-i}\right) u_i(f(\theta'), \theta) + \sum_{m_{-i} \in M_{-i}^{2,i}} \lambda_i\left(m_{-i}\right) u_i(m_i^3(m_{-i}^1), \theta) + \sum_{m_{-i} \notin M_{-i}^1 \cup M_{-i}^{2,i}} \lambda_i\left(m_{-i}\right) \max_{y \in \mathcal{Y}} u_i(y, \theta)$$

while for $\widehat{m}_i = (\theta', \widehat{m}_i^2, \widehat{m}_i^3, \widehat{m}_i^4)$, it is easily checked that as $\widehat{m}_i^2$ tends to infinity, i's expected payoffs tend toward the expression above. Hence, choosing $\widehat{m}_i^2$ large enough, $\widehat{m}_i$ is a better response against λ_i for player i than m_i, a contradiction.

So if $m_i = \left(\theta', 1, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$, it follows player i must be convinced that each other player must be choosing a message of the form $\left(\theta', 1, m_j^3, m_j^4\right)$, and hence $\lambda_i\left(\left\{m_{-i} \in M_{-i}^1\right\}\right) = 1$. Thus there must exist a message of the form $m_j = \left(\theta', 1, m_j^3, m_j^4\right) \in S_j^{\mathcal{M},\theta}$ for all j. Now, proceed by contradiction and assume that $f(\theta') \neq f(\theta)$. By Maskin monotonicity, we know that there exist j with $u_j\left(f\left(\theta'\right), \theta'\right) \geq u_j\left(y(\theta', \theta), \theta'\right)$ and $u_j\left(y(\theta', \theta), \theta\right) > u_j\left(f\left(\theta'\right), \theta\right)$. By the above argument, we know that player j's belief against which $m_j = \left(\theta', 1, m_j^3, m_j^4\right)$ is a best reply assigns probability one to each player $l \neq j$ sending a message of the form $m_l = \left(\theta', 1, m_l^3, m_l^4\right)$. Hence, player j's expected payoff from playing m_j is $u_j(f(\theta'), \theta)$, while if j deviates to $\hat{m}_j = \left(\theta', \hat{m}_j^2, \hat{m}_j^3, m_j^4\right)$, where $\hat{m}_j^2 > 1$ and

$$\hat{m}_j^3(\hat{\theta}) = \begin{cases} y(\theta', \theta), & \text{if} \quad \hat{\theta} = \theta', \\ m_i^3\left(\hat{\theta}\right), & \text{if} \quad \text{otherwise}, \end{cases}$$

player j believes with probability one that Rule 2(i) will be triggered. Hence, player j's expected payoff would be

$$\left(1 - 1/(\widehat{m}_j^2 + 1)\right) u_j(y(\theta', \theta), \theta) + (1/(\widehat{m}_j^2 + 1)) u_j(z_j(\theta', \theta'), \theta).$$

Note that as $\hat{m}_j^2$ tends to infinity, this expression tends to $u_j(y(\theta', \theta), \theta)$ which is strictly larger than $u_j\left(f\left(\theta'\right), \theta\right)$. Hence for $\hat{m}_j^2$ large enough, $\hat{m}_j$ is better response for player j than $\left(\theta', 1, m_j^3, m_j^4\right)$, a contradiction. Thus $f\left(\theta'\right) = f\left(\theta\right)$. Since the social choice function has been assumed to be responsive, we get $\theta' = \theta$ as claimed.

Completion of proof. Claims 1, 2 and 3 together imply that for each $\theta : S_i^{\mathcal{M},\theta} \neq \varnothing$ and $m_i \in S_i^{\mathcal{M},\theta} \Rightarrow m_i^2 = 1$ and $m_i^1 = \theta$. Thus $S^{\mathcal{M},\theta} \neq \varnothing$ and $m \in S^{\mathcal{M},\theta} \Rightarrow g\left(m\right) = f\left(\theta\right)$. ■

The mechanism $\mathcal{M}$ used here allows each agent to propose a menu of choices $m_i^3 = \left\{m_i^3\left(\theta\right)\right\}_{\theta \in \Theta}$. The menu m_i^3 gives agent i the opportunity to select an appropriate allocation in case that Rule 2 is triggered. In our sufficiency argument, the NWA property replaces the no veto property which commonly appears in the sufficiency argument for implementation in Nash equilibrium. Yet, in terms of the proof, the role of the NWA property is quite distinct from the no veto property. The NWA property guarantees that in the augmented mechanism, any report in state θ in which an agent expresses his disagreement with the remaining agents (i.e. $m_i^2 > 1$) cannot be a rationalizable report. By contrast, the no veto property guaranteed that if an agent were to express his disagreement, then further disagreement by other agents would only be possible in equilibrium if it would lead to the same equilibrium allocation as prescribed $f\left(\theta\right)$.

We note that our mechanism not only implements in rationalizable messages but also implements in Nash equilibrium (the proof of Claim 2 above indeed establishes the existence of a pure Nash equilibrium at each state). In recent work, Bochet (2007) and Benoit and Ok (2008) report sufficient conditions for implementation in Nash equilibrium strategies using stochastic mechanisms. Their conditions, the *top strict difference* condition and the *top coincidence* condition, respectively, do not imply nor are they implied by the NWA property required for sufficiency. In related work, Serrano and Vohra (2007) have used stochastic implementing mechanisms to provide weak sufficient conditions for Bayesian implementation in *mixed strategy* Bayes Nash equilibrium.

4 The Non-Responsive Case

In this section, we discuss extensions of our results to the cases when the social choice function is not responsive. We will provide a strengthening of strict Maskin monotonicity that can be shown to be sufficient (together with a strengthening of the NWA) even if the social choice function is not responsive. We also show that the strengthening of strict Maskin monotonicity is actually necessary for rationalizable implementation given a weak condition on the class of mechanisms to be considered. This weak condition is trivially satisfied when the social choice function is responsive.

Now, given a social choice function f, let us consider the unique partition of $\Theta : \mathcal{P}_f = \{\Theta_z\}_{z \in f(\Theta)}$ such that

$$\Theta_z = \{\theta \in \Theta \,|\, f(\theta) = z\}. \tag{17}$$

We now introduce the following notion which reduces to strict Maskin monotonicity in case f is responsive.

Definition 5 (Strict Maskin Monotonicity*)
Social choice function f satisfies strict Maskin monotonicity* if there exists a partition $\mathcal{P}$ of Θ finer than $\mathcal{P}_f$ s.t. for any θ :

1. *$\theta' \in \mathcal{P}(\theta)$ whenever for all i and y*

$$\left[\text{for all } \widehat{\theta} \in \mathcal{P}(\theta) : u_i\left(f(\theta), \widehat{\theta}\right) > u_i\left(y, \widehat{\theta}\right)\right] \Rightarrow \left[u_i\left(f(\theta), \theta'\right) \geq u_i\left(y, \theta'\right)\right]; \tag{18}$$

or, equivalently,

2. *$\theta' \notin \mathcal{P}(\theta)$ implies for some i and y*

$$u_i\left(y, \theta'\right) > u_i\left(f(\theta), \theta'\right) \text{ and } u_i\left(f(\theta), \widehat{\theta}\right) > u_i\left(y, \widehat{\theta}\right) \text{ for all } \widehat{\theta} \in \mathcal{P}(\theta). \tag{19}$$

Before we establish the necessary and sufficient conditions, we briefly describe the complications that arise with a non-responsive social choice function. By definition, under a non-responsive social choice function there are at least two states, θ and θ', that lead to the same social choice: $f(\theta) = f(\theta') = z$. Now, a priori, the principal would not need to know whether it is the state θ or θ' which leads to the realization of the social choice z. In fact, it would appear that it would be sufficient to learn that the realized state belongs to the set Θ_z of states which lead to the social choice z. Now, such a coarse reporting protocol as suggested by the above partition $\mathcal{P}_f$ would be sufficient if the agents were known to report truthfully, yet a problem arises if they might not report truthfully. For, if an agent now alleges collusive behavior of the remaining agents, the principal may lack the information to verify whether the whistle-blower himself is behaving in an incentive compatible manner. After all, the principal would merely know that the reported state is in some set Θ_z but would not know the identity of the state itself. Thus, while it might not be useful to distinguish between any two states $\theta, \theta' \in \Theta_z$ if the agents were to report truthfully, it might be critical to distinguish between θ and θ' in order to fend off undesirable equilibrium play by the agents. This discussion might therefore suggest that the inequalities (18), or alternatively (19), should be satisfied for the finest possible partition of states. But, as we argue next, such a condition would (i) require too much to constitute a necessary condition, and (ii) be impossible to satisfy by any implementing mechanism.

The first observation is straightforward to establish. Consider for the moment the strict Maskin monotonicity* condition in the version of (19), which we might refer to as the whistle-blower inequality. Now suppose that the social choice problem is such that the inequalities (19) are satisfied even for the coarse partition $\mathcal{P}_f$ itself. In this case, we would find that the principal would not need to distinguish between any two states $\theta, \theta' \in \Theta_z$, either for truthtelling or, by condition (19), for whistle-blowing behavior.

The second observation stems from an earlier result. Lemma 1 gave a sufficient condition under which the set of rationalizable actions for any pair of states, θ and θ', have to be identical for all agents. For the purpose here we can restrict attention to any two states with $\theta, \theta' \in \Theta_z$. In this case, the condition (1) reads as follows:

$$u_i(z, \theta) > u_i(y, \theta) \Rightarrow u_i(z, \theta') \geq u_i(y, \theta') \text{ for all } i \text{ and } y.$$

In words, if for every agent i, the upper contour set (relative to the allocation $f(\theta) = f(\theta') = z$), at one state, say θ', is included in the upper contour set of the other state, say θ, then the sets of rationalizable actions have to coincide. But of course, once the sets of rationalizable actions have to agree, it will be impossible to distinguish behavior in state θ from behavior in state θ'. The inclusion property of the upper contour sets, given by condition (1), thus imposes an upper bound on how fine the partition $\mathcal{P}$ can be chosen while remaining compatible with rationalizable behavior. We finally observe that the partition $\mathcal{P}$ may yet have to be coarser than is indicated by the pairwise inclusion property. To see this, consider $\theta, \theta', \theta'' \in \Theta_z$, and suppose that the upper contour sets (relative to the allocation z) in state θ' as well as in state θ'' are included in the upper contour sets in state θ, but that the upper contour sets in the state θ' and θ'' themselves do not display an inclusive relationship. Now, Lemma 1 tells us that $S^{\mathcal{M},\theta} = S^{\mathcal{M},\theta'}$ and that $S^{\mathcal{M},\theta} = S^{\mathcal{M},\theta''}$ which of course implies that $S^{\mathcal{M},\theta'} = S^{\mathcal{M},\theta''}$ even though the condition (1) does not apply to the states θ' and θ'' themselves.

As we already stated, we can prove that strict Maskin monotonicity* is necessary under a weak condition on the class of mechanisms we consider. This condition states that for any state θ and any rationalizable message m_i of any player i in this state, the message m_i is also best-response to some belief with support in the set of rationalizable actions of the other players and for any state $\widehat{\theta}$ such that $S^{\mathcal{M},\widehat{\theta}} = S^{\mathcal{M},\theta}$, best responses against this belief are non-empty.

Definition 6 (Best Response Property)

Given a social choice function f, a mechanism $\mathcal{M}$ has the best-response property *if for all θ and all $m_i \in S_i^{\mathcal{M},\theta}$, there exists $\lambda_i^{m_i,\theta} \in \Delta(M_{-i})$ satisfying $\lambda_i^{m_i,\theta}(m_{-i}) > 0 \Rightarrow m_j \in S_j^{\mathcal{M},\theta}$ for each $j \neq i$, and such that m_i is a best response against $\lambda_i^{m_i,\theta}$ in state θ and*

$$\arg\max_{m_i'} \sum_{m_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i', m_{-i}), \widehat{\theta}) \neq \emptyset$$

for all $\widehat{\theta}$ such that $S^{\mathcal{M},\widehat{\theta}} = S^{\mathcal{M},\theta}$.

Note that if f is responsive then any implementing mechanism must satisfy $S^{\mathcal{M},\widehat{\theta}} = S^{\mathcal{M},\theta} \Rightarrow \widehat{\theta} = \theta$ and so any implementing mechanism must have the best-response property. Moreover, the best-response property also holds for any pair θ, θ' which are directly related through the inclusion property (1). The best-response property then secures that it applies also to profiles which are indirectly related as in the example of $\theta, \theta', \theta'' \in \Theta_z$ discussed above. Hence, the subsequent Proposition 3 generalizes Proposition 1 above.

Proposition 3 (Necessary Conditions)

If f is implementable in rationalizable strategies by a mechanism $\mathcal{M}$ having the best-response property, then f satisfies strict Maskin monotonicity.*

In order to show this, we prove the following Lemma that generalizes Lemma 1.

Lemma 3

Assume the existence of a mechanism $\mathcal{M} = \left((M_i)_{i=1}^{I}, g \right)$*, that has the best-response property and that implements* f *in rationalizable strategies. Pick* θ *and* θ' *satisfying condition (18) where the partition* $\mathcal{P}$ *is assumed to be* $\mathcal{P}(\theta'') = \left\{ \tilde{\theta} \in \Theta \middle| S^{\mathcal{M},\theta''} = S^{\mathcal{M},\tilde{\theta}} \right\}$ *for any* θ''*. We have* $S^{\mathcal{M},\theta} = S^{\mathcal{M},\theta'}$*.*

Proof. Fix any mechanism $\mathcal{M} = \left((M_i)_{i=1}^{I}, g \right)$ that has the best-response property and that implements f and pick θ and θ' satisfying condition (18) for $\mathcal{P}(\theta'') = \left\{ \tilde{\theta} \in \Theta \middle| S^{\mathcal{M},\theta''} = S^{\mathcal{M},\tilde{\theta}} \right\}$ i.e. for all i and y

$$\left[\text{for all } \widehat{\theta} \in \mathcal{P}(\theta) : u_i\left(f(\theta), \widehat{\theta} \right) > u_i\left(y, \widehat{\theta} \right) \right] \Rightarrow \left[u_i\left(f(\theta), \theta' \right) \geq u_i\left(y, \theta' \right) \right].$$

Note that by construction, $\widehat{\theta} \in \mathcal{P}(\theta) \Rightarrow S^{\mathcal{M},\widehat{\theta}} = S^{\mathcal{M},\theta}$. In addition, since $\mathcal{M}$ implements f in rationalizable strategies, for any state in $\mathcal{P}(\theta'')$, f picks the outcome $f(\theta'')$ and so $\mathcal{P}$ is finer than $\mathcal{P}_f$.

We first show that $S^{\mathcal{M},\theta} \subseteq S^{\mathcal{M},\theta'}$. Because $b^{\theta}\left(S^{\mathcal{M},\theta} \right) = S^{\mathcal{M},\theta}$, $S^{\mathcal{M},\theta}$ has the best response property in state θ i.e. for all player i and all $m_i \in S_i^{\mathcal{M},\theta}$, there exists $\lambda_i^{m_i,\theta} \in \Delta(M_{-i})$ such that $\lambda_i^{m_i,\theta}(m_{-i}) > 0 \Rightarrow m_j \in S_j^{\mathcal{M},\theta}$ for each $j \neq i$, and

$$\sum_{m_{-i} \in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i, m_{-i}), \theta) \geq \sum_{m_{-i} \in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i', m_{-i}), \theta)$$

for all $m_i' \in M_i$. In addition, since $\mathcal{M}$ has the best response property, $\lambda_i^{m_i,\theta}$ can be chosen so that:

$$\arg\max_{m_i'} \sum_{m_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i', m_{-i}), \widehat{\theta}) \neq \emptyset,$$

for all $\widehat{\theta} \in \mathcal{P}(\theta)$. This in turn implies that for all $\widehat{\theta} \in \mathcal{P}(\theta)$:

$$\sum_{m_{-i} \in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i, m_{-i}), \widehat{\theta}) \geq \sum_{m_{-i} \in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i', m_{-i}), \widehat{\theta}) \tag{20}$$

for all $m_i' \in M_i$. To see this, observe that if it were not true, we would have for some $\widehat{\theta} \in \mathcal{P}(\theta)$:

$$\sum_{m_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i^*, m_{-i}), \widehat{\theta}) > \sum_{m_{-i} \in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i}) u_i(g(m_i, m_{-i}), \widehat{\theta}) = u_i(f(\theta), \widehat{\theta})$$

where m_i^* denotes a best response to $\lambda_i^{m_i,\theta}$ in state $\widehat{\theta}$ and so $g(m_i^*, m_{-i}) \neq f(\theta)$ for some $(m_i^*, m_{-i}) \in S^{\mathcal{M},\hat{\theta}}$ which contradicts the fact that $\mathcal{M} = \left((M_i)_{i=1}^{I}, g \right)$ implements f.

Now, we want to show that m_i is also a best response against $\lambda_i^{m_i,\theta}$ in state θ'. Since i and $m_i \in S_i^{\mathcal{M},\theta}$ have been fixed arbitrarily, this will prove that $S^{\mathcal{M},\theta}$ has the best response property in state θ' and so that $S^{\mathcal{M},\theta} \subseteq S^{\mathcal{M},\theta'}$ as claimed. Note first that for any m_{-i} such that $\lambda_i^{m_i,\theta}(m_{-i}) > 0$, $m_{-i} \in S_{-i}^{\mathcal{M},\theta}$ and so because $m_i \in S_i^{\mathcal{M},\theta}$, we have $g(m_i, m_{-i}) = f(\theta)$. Thus,

$$\sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i, m_{-i}), \theta') = u_i(f(\theta), \theta') = \sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i', m_{-i}), \theta') \quad (21)$$

for all $m_i' \in S_i^{\mathcal{M},\theta}$. In addition, we claim that for all $\widehat{\theta} \in \mathcal{P}(\theta)$:

$$\sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i, m_{-i}), \widehat{\theta}) = u_i(f(\theta), \widehat{\theta}) > \sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i', m_{-i}), \widehat{\theta}) \quad (22)$$

for any $m_i' \notin S_i^{\mathcal{M},\theta}$. Indeed, by (20), the above is true with a weak inequality. Now if an equality were to hold, some $m_i' \notin S_i^{\mathcal{M},\theta}$ would be a best response against $\lambda_i^{m_i,\theta}$ in some state $\widehat{\theta}$. Thus the set $\left(\{m_i'\} \cup S_i^{\mathcal{M},\widehat{\theta}}\right) \times S_{-i}^{\mathcal{M},\widehat{\theta}} = \left(\{m_i'\} \cup S_i^{\mathcal{M},\theta}\right) \times S_{-i}^{\mathcal{M},\theta}$ would have the best response property in state $\widehat{\theta}$ implying that $m_i' \in S_i^{\mathcal{M},\widehat{\theta}} = S_i^{\mathcal{M},\theta}$ which is false by assumption.

Now, by assumption, we know that

$$\left[\widehat{\theta} \in \mathcal{P}(\theta) : u_i\left(f(\theta), \widehat{\theta}\right) > u_i\left(y, \widehat{\theta}\right)\right] \Rightarrow \left[u_i\left(f(\theta), \theta'\right) \geq u_i\left(y, \theta'\right)\right] \text{ for all } i \text{ and } y$$

and so applying this to the lotteries $y \triangleq \sum_{m_{-i}} \lambda_i^{m_i,\theta}(m_{-i})g(m_i', m_{-i})$, equation (22) yields

$$\sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i, m_{-i}), \theta') = u_i(f(\theta), \theta') \geq \sum_{m_{-i}\in M_{-i}} \lambda_i^{m_i,\theta}(m_{-i})u_i(g(m_i', m_{-i}), \theta') \quad (23)$$

for any $m_i' \notin S_i^{\mathcal{M},\theta}$. Finally, (21) and (23) ensure that m_i is also a best response against $\lambda_i^{m_i,\theta}$ in state θ'. Hence, $S^{\mathcal{M},\theta} \subseteq S^{\mathcal{M},\theta'}$ as claimed.

To complete the proof, we have to show that $S^{\mathcal{M},\theta} \supseteq S^{\mathcal{M},\theta'}$. The argument is the same as in Lemma 1. ■

Note that if f is implementable by a mechanism $\mathcal{M}$ that has the best response property, then if one were to pick the partition $\mathcal{P}$ given by $\mathcal{P}(\theta'') = \left\{\tilde{\theta} \in \Theta \,\middle|\, S^{\mathcal{M},\theta''} = S^{\mathcal{M},\tilde{\theta}}\right\}$ for any θ'', then whenever θ and θ' satisfy condition (18), by Lemma 3, we must have $S^{\mathcal{M},\theta} = S^{\mathcal{M},\theta'}$ and so, by definition, $\theta' \in \mathcal{P}(\theta)$. Hence, Proposition 3 is obtained as a corollary of Lemma 3.

As mentioned earlier, it is easily checked that our sufficiency argument can be extended to this setting provided that a strengthening of NWA is used. To be more specific, if one assumes that f is (strict) Maskin monotonic and that for any state θ, there exists some outcome that is worse than the outcome selected by f at any state in the partition cell $\mathcal{P}(\theta)$, then we can build a mechanism similar to the one built in the proof of Proposition 2.[6] In the revised mechanism each player is

[6] As for Proposition 2, the proof would go through if we just considered Maskin monotonicity* instead of strict Maskin monotonicity*.

asked to report a partition cell P in $\mathcal{P}$, an integer, a mapping from $\mathcal{P}$ to $\mathcal{Y}$ and a lottery in $\mathcal{Y}$. Essentially, everything would go as if we were replacing each state θ by the partition cell containing θ. In particular, as in the responsive case, we can show that for any rationalizable message, using condition (19) in the definition of Maskin monotonicity*, each agent will report truthfully, i.e., will report $\mathcal{P}(\theta)$ whenever the true state is θ and announce an integer equal to 1. The modified notions of strict Maskin monotonicity and NWA as well as the sufficiency argument itself are presented in detail in the appendix.

5 Concluding Remarks

We conclude with a few observations. First, this paper focused on social choice functions, let us briefly discuss the case of social choice correspondences. In Proposition 1 and 2 we reported results for social choice functions only. A social choice correspondence defines a set of permissible allocations and rationalizability is a set-based solution concept. Thus, there are a number of plausible extensions of the definition of rationalizable implementation to social choice correspondences. The extensions basically vary to the extent that one wishes to restrict attention to selections in the set of outcome profiles.[7] We now show that Maskin monotonicity may not even be a necessary condition for implementation in rationalizable strategies (according to at least one natural definition of these terms).[8] We describe the difficulty of social choice correspondences with the following approach (and subsequent example). A *(pure outcome) social choice correspondence* (SCC) is a mapping $F : \Theta \to 2^{Z} / \varnothing$. A social choice correspondence F is implementable in rationalizable strategies if there exists a mechanism $\mathcal{M}$ with $g\left[S^{\mathcal{M},\theta}\right] = F(\theta)$ for all $\theta \in \Theta$. A SCC F is *Maskin monotonic* if: whenever $z^* \in F(\theta)$ and $u_i(z^*, \theta) \geq u_i(z, \theta) \Rightarrow u_i(z^*, \theta') \geq u_i(z, \theta')$ for all i and z; then $z^* \in F(\theta')$. Note that this definition is given in terms of pure outcomes. Now consider the following example. There are 2 agents; $\Theta = \{\alpha, \beta\}$; $Z = \{a, b, c, d\}$; payoffs are given by the following table:

$u(\cdot,\cdot)$	a	b	c	d
α	$1+\varepsilon, 0$	$0, 1+\varepsilon$	$1, 1$	$1+2\varepsilon, 1+2\varepsilon$
β	$1+\varepsilon, 0$	$0, \varepsilon$	$1, 1$	$1+2\varepsilon, 1+2\varepsilon$

The social choice correspondence is $F^*(\alpha) = \{a, b, c, d\}$ and $F^*(\beta) = \{d\}$. Now we demonstrate that F^* is not Maskin monotonic. To see why, note that $a \in F^*(\alpha)$ and that $u_i(a, \alpha) \geq u_i(z, \alpha) \Rightarrow$

[7] The issue already appears in incomplete information implementation literature, where it is common to use a social choice set, a selection, rather than the social choice correspondence.

[8] As shown in Mezzetti and Renou (2009), a similar issue appears when one considers implementation in mixed Nash equilibrium where – contrary to the usual requirement – implementation does not ask for each alternative in the set of desired alternatives to be the outcome of a *pure* Nash equilibrium.

$u_i(a,\beta) \geq u_i(z,\beta)$ for all i and z. So Maskin monotonicity would require $a \in F^*(\beta)$. But F^* is implementable in rationalizable strategies. Consider the mechanism $\mathcal{M}$ with $M_i = \{m_i^1, m_i^2, m_i^3\}$ and deterministic g given by the following matrix:

$g(\cdot)$	m_2^1	m_2^2	m_2^3
m_1^1	a	b	c
m_1^2	b	a	c
m_1^3	c	c	d

Now, for each i, $S_{i,k}^{\mathcal{M},\alpha} = M_i$ for all k and thus $S_i^{\mathcal{M},\alpha} = M_i$. Thus $g[S^{\mathcal{M},\alpha}] = \{a,b,c,d\} = F^*(\alpha)$. But in state β, we have in subsequent rounds of elimination:

$$
\begin{aligned}
S_{1,0}^{\mathcal{M},\beta} &= \{m_1^1, m_1^2, m_1^3\} \text{ and } S_{2,0}^{\mathcal{M},\beta} = \{m_2^1, m_2^2, m_2^3\}, \\
S_{1,1}^{\mathcal{M},\beta} &= \{m_1^1, m_1^2, m_1^3\} \text{ and } S_{2,1}^{\mathcal{M},\beta} = \{m_2^3\}, \\
S_{1,2}^{\mathcal{M},\beta} &= \{m_1^3\} \text{ and } S_{2,2}^{\mathcal{M},\beta} = \{m_2^3\},
\end{aligned}
$$

and thus $g[S^{\mathcal{M},\beta}] = \{d\} = F^*(\beta)$. We thus showed that F^* is implementable in rationalizable strategies, yet did not satisfy Maskin monotonicity.

Second, Proposition 1 exhibits a necessary condition for rationalizable implementation that is strictly stronger than the usual one for Nash implementation. Here we provide an example of a social choice function that is not rationalizable implementable but which is Nash-implementable. There are 3 agents; $\Theta = \{\alpha, \beta\}$; $Z = \{a,b,c,d\}$; payoffs are given by the following table:

$u(\cdot,\cdot)$	a	b	c	d
α	$0,0,0$	$0,1,0$	$1,0,0$	$0,0,1$
β	$0,0,0$	$1,1,1$	$0,0,0$	$0,0,0$

The social choice correspondence is $f(\alpha) = a$ and $f(\beta) = b$. It is easily checked that f is Maskin monotonic ($u_1(f(\alpha),\alpha) = u_1(b,\alpha)$ but $u_1(f(\alpha),\beta) < u_1(b,\beta)$; similarly, $u_1(f(\beta),\beta) > u_1(c,\beta)$ but $u_1(f(\beta),\beta) < u_1(c,\beta)$) and satisfies no-veto-power. Hence, standard arguments (see Maskin (1999) and Maskin and Sjostrom (2004)) show that f is implementable in (pure or mixed) Nash equilibrium. However, for any player i and $y \in \Delta(Z) : u_i(f(\alpha),\alpha) \leq u_i(y,\alpha)$ and so this social choice function cannot be strict Maskin monotonic, and so it is not implementable in rationalizable strategies.

Finally, from a purely game-theoretic point of view, the results presented in Proposition 1 and 2 may appear surprisingly strong. Given that we are investigating a social choice function, the notion of full implementation is akin to requiring that the game has a unique equilibrium (outcome). The present implementation results then say that – provided that the social choice function is responsive – a unique rationalizable outcome arises under (almost) the same conditions as a unique Nash equilibrium outcome. This is noteworthy as the necessary and almost sufficient condition of Maskin monotonicity is much weaker than the well-known conditions under which there are close connections between Nash equilibrium and rationalizability, such as supermodular or concave games. The Nash equilibrium results indicate the strength of the implementation approach to reduce the number of equilibria. The arguments presented here complement and extend these results. By using infinite message spaces and stochastic allocations, we strengthen the positive implementation results to the weaker solution concept of rationalizability.

References

Abreu, D., and H. Matsushima (1992): "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica*, 60, 993–1008.

Benoit, J.-P., and E. Ok (2008): "Nash Implementation Without No-Veto Power," *Games and Economic Behavior*, 64, 51–67.

Bergemann, D., and S. Morris (2008): "Robust Implementation in General Mechanisms," Discussion Paper 1666, Cowles Foundation, Yale University.

——— (2009): "Robust Implementation in Direct Mechanisms," *Review of Economic Studies*, 76, 1175–1206.

Bernheim, D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007–1028.

Bochet, O. (2007): "Nash Implementation with Lottery Mechanisms," *Social Choice and Welfare*, 28, 111–125.

Brandenburger, A., and E. Dekel (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391–1402.

Cabrales, A., and R. Serrano (2008): "Implementation in Adaptive Better-Response Dynamics," Discussion paper, Universitat Carlos III de Madrid and Brown University.

Lipman, B. (1994): "A Note on the Implications of Common Knowledge of Rationality," *Games and Economic Behavior*, 6, 114–129.

Maskin, E. (1999): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, 23–38.

Maskin, E., and T. Sjostrom (2004): "Implementation Theory," in *Handbook of Social Choice and Welfare*, ed. by K. Arrow, A. Sen, and K. Suzumura, vol. 1. North-Holland, Amsterdam.

Mezzetti, C., and L. Renou (2009): "Implementation in Mixed Nash Equilibrium," Discussion Paper, University of Leicester.

Osborne, M., and A. Rubinstein (1994): *A Course in Game Theory*. MIT Press, Cambridge.

Oury, M., and O. Tercieux (2009): "Continuous Implementation," Discussion paper, Paris School of Economics.

Pearce, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050.

Serrano, R., and R. Vohra (2007): "Multiplicity of Mixed Equilibria in Mechanisms: A Unified Approach to Exact and Approximate Implementation," Discussion paper, Brown University.

6 Appendix

In the appendix we provide a proof of our sufficiency result for the case that the social choice function is not responsive. Assume that f satisfies strict Maskin monotonicity* with partition $\mathcal{P}$. We say that f satisfies the "no worst alternative*", not only when agents never get their worst outcome under the social choice function but if for any state θ, there is some outcome that is worst than the outcome selected by f at any state in the element $\mathcal{P}(\theta)$ of the partition $\mathcal{P}$. In the sequel, we will write $[\theta]$ for $\mathcal{P}(\theta)$ and sometimes abuse notations writing $f([\theta])$ for $f(\theta)$.

Definition 7

Social choice function f satisfies "no worst alternative" (NWA*) if, for each i and $[\theta]$, there exists $\underline{y}_i([\theta])$ such that*

$$u_i\left(f(\theta),\widehat{\theta}\right) > u_i\left(\underline{y}_i([\theta]),\widehat{\theta}\right)$$

for each $\widehat{\theta} \in [\theta]$.

Given a set of allocations $\{\underline{y}_i([\theta])\}_{\theta\in\Theta}$, it is useful to define the average allocation $\underline{y}_i$ of this set by setting

$$\underline{y}_i \triangleq \frac{1}{\#\mathcal{P}} \sum_{[\theta]\in\mathcal{P}} \underline{y}_i([\theta]).$$

Note that under NWA*, for all $[\theta]$ and all i, there exists a $y_i([\theta])$ such that

$$u_i(y(i,[\theta]),\hat{\theta}) > u_i(\underline{y}_i,\hat{\theta}), \tag{24}$$

for all $\hat{\theta} \in [\theta]$, it is indeed easily checked that this is true for $y_i([\theta]) = \frac{1}{\#\mathcal{P}} \sum_{[\tilde{\theta}]\in\mathcal{P}\setminus[\theta]} \underline{y}_i([\tilde{\theta}]) + \frac{1}{\#\mathcal{P}} f(\theta)$. It will also be useful to define the average allocation $\underline{y}$ of the set $\{\underline{y}_i\}_{i\in\mathcal{I}}$ by setting

$$\underline{y} \triangleq \frac{1}{I} \sum_{i\in\mathcal{I}} \underline{y}_i.$$

Here again, we note that under NWA*, for all $[\theta]$ and all i, there exists a $y^*(i,[\theta])$ such that

$$u_i(y^*(i,[\theta]),\hat{\theta}) > u_i(\underline{y},\hat{\theta}), \tag{25}$$

for all $\hat{\theta} \in [\theta]$. It is indeed easily checked that this is true for $y_i^*([\theta]) = \frac{1}{I}\sum_{j\neq i} \underline{y}_j + \frac{1}{I} y_i([\theta])$.

We now construct an auxiliary set of allocations, denoted by $\{z_i\left([\theta],[\theta']\right)\}_{\theta,\theta'}$, which uses the existence of the allocations $\{\underline{y}_i([\theta])\}_{\theta\in\Theta}$. Here is an analogous lemma to Lemma 2.

Lemma 4

If social choice function f satisfies NWA*, then for each player i, there exists a collection of lotteries $\{z_i([\theta],[\theta'])\}_{\theta,\theta'}$ such that for all $[\theta],[\theta']$:

$$u_i\left(f(\theta'),\hat{\theta}\right) > u_i\left(z_i([\theta],[\theta']),\hat{\theta}\right) \text{ for all } \hat{\theta}\in[\theta'] \tag{26}$$

and for $[\theta]\neq[\theta']$:

$$u_i\left(z_i([\theta],[\theta']),\theta\right) > u_i\left(z_i([\theta'],[\theta']),\theta\right). \tag{27}$$

Proof. Based on the allocations $\{\underline{y}_i([\theta])\}_{\theta\in\Theta}$, we define our collection of lotteries as follows. First, for all $[\theta']$:

$$z_i([\theta'],[\theta']) \triangleq (1-\varepsilon)\,\underline{y}_i([\theta']) + \varepsilon \underline{y}_i,$$

with $\underline{y}_i$ as defined in (24). In addition, for all $[\theta],[\theta']$ with $[\theta]\neq[\theta']$:

$$z_i([\theta],[\theta']) \triangleq (1-\varepsilon)\,\underline{y}_i([\theta']) + \varepsilon\left(\frac{1}{\#\mathcal{P}}\sum_{[\hat{\theta}]\in\mathcal{P}\setminus[\theta]}\underline{y}_i([\hat{\theta}]) + \frac{1}{\#\mathcal{P}}f(\theta)\right).$$

By NWA* and the finiteness of the state space Θ, we can find a sufficiently small, but positive, $\varepsilon>0$ such that for all $[\theta]$ and $[\theta']$: $u_i\left(f(\theta'),\hat{\theta}\right) > u_i\left(z_i([\theta],[\theta']),\hat{\theta}\right)$ for all $\hat{\theta}\in[\theta']$ which establishes inequality (26). Now we observe that the only difference between $z_i([\theta'],[\theta'])$ and $z_i([\theta],[\theta'])$ is the fact that the lottery $\underline{y}_i([\theta])$ is replaced by the lottery $f(\theta)$. But now by NWA*, this is clearly increasing the expected utility of agent i in state θ, and hence we have for all $[\theta],[\theta']$ with $[\theta]\neq[\theta']$:

$$u_i\left(z_i([\theta],[\theta']),\theta\right) > u_i\left(z_i([\theta'],[\theta']),\theta\right),$$

which establishes the strict inequality (27). ■

We are now in a position to prove our sufficiency result.

Proposition 4 (Sufficient Conditions)

If $I\geq 3$ and f satisfies strict Maskin monotonicity* and NWA*, then f is implementable in rationalizable strategies.

Proof. We establish the result by constructing an implementing mechanism $\mathcal{M}=(M,g)$. First, recall that by definition of strict Maskin monotonicity*, for all θ and θ' such that $[\theta']\neq[\theta]$, there exists i and $y([\theta],\theta')\in Y$ with $u_i\left(y([\theta],\theta'),\theta'\right) > u_i\left(f(\theta),\theta'\right)$ and for all $\widehat{\theta}\in[\theta]$: $u_i\left(f(\theta),\widehat{\theta}\right) > u_i\left(y([\theta],\theta'),\widehat{\theta}\right)$. We define the following finite set of lotteries:

$$\mathcal{Y}=\left\{z_i([\theta],[\theta'])\right\}_{i,\theta,\theta'} \cup \left\{y([\theta],\theta')\right\}_{\theta,\theta':[\theta]\neq[\theta']} \cup \{y^*(i,[\theta])\}_{i,\theta}$$

where the collection $\left\{z_i\left([\theta],[\theta']\right)\right\}_{i,\theta,\theta'}$ has been defined in Lemma 4 while the collection $\{y^*(i,[\theta])\}_{i,\theta}$ has been established in (25).

Each agent i sends a message $m_i = \left(m_i^1, m_i^2, m_i^3, m_i^4\right)$, where $m_i^1 \in \mathcal{P}$, $m_i^2 \in \mathbb{Z}_+$, $m_i^3 : \mathcal{P} \rightarrow \mathcal{Y}, m_i^4 \in \mathcal{Y}$. The third component of the message profile will allow agent i to suggest an allocation $m_i^3([\theta])$ contingent on all the other agents $j \neq i$ reporting $m_j^1 = [\theta]$. The outcome function will make use of the "uniformly worst outcome" defined earlier by $\underline{y}$. Now the outcome $g(m)$ is determined by the following rules:

Rule 1: If $m_i^1 = [\theta]$ and $m_i^2 = 1$ for all i, pick $f(\theta)$.[9]
Rule 2: If there exists $i \in I$ – called the deviating player – such that $\left(m_j^1, m_j^2\right) = ([\theta], 1)$ for all $j \neq i$ and $\left(m_i^1, m_i^2\right) \neq ([\theta], 1)$, then we go to two subrules:
(i): if for all $\widehat{\theta} \in [\theta] : u_i\left(f(\theta), \widehat{\theta}\right) \geq u_i\left(m_i^3([\theta]), \widehat{\theta}\right)$, pick $m_i^3([\theta])$ with probability $1 - 1/(m_i^2 + 1)$ and $z_i([\theta],[\theta])$ with probability $1/(m_i^2 + 1)$;
(ii): if for some $\widehat{\theta} \in [\theta] : u_i\left(f(\theta), \widehat{\theta}\right) < u_i\left(m_i^3([\theta]), \widehat{\theta}\right)$, pick $z_i([\theta],[\theta])$ with probability 1.
Rule 3: In all other cases, we identify a pivotal agent i by requiring that $m_i^2 \geq m_j^2$ for all $j \in I$ and that if for $j \neq i$, $m_i^2 = m_j^2$, then $i < j$. The rule then requires that with probability $1 - 1/(m_i^2 + 1)$ we pick m_i^4, and with probability $1/(m_i^2 + 1)$ we pick $\underline{y}$.
Claim 1. It is never a best reply for agent i to send a message with $m_i^2 > 1$ (i.e., $m_i \in b_i^\theta\left(\overline{S}\right) \Rightarrow m_i^2 = 1$).
Proof for Claim 1. Proceed by contradiction and suppose $m_i = \left(m_i^1, m_i^2, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$ and $m_i^2 > 1$. Then for any profile of messages m_{-i} player i's opponents may play, (m_i, m_{-i}) will trigger either Rule 2 or Rule 3. But in this case, whatever agent i's beliefs $\lambda_i \in \Delta(M_{-i})$ about the other agents' messages, his payoff can be increased by modifying m_i appropriately, in particular by increasing the integer choice from m_i^2. To see this, denote the set of messages of the remaining agents in which Rule 2 is triggered by:

$$M_{-i}^2 \triangleq \left\{m_{-i} \in M_{-i} \left| m_j^1 = [\theta'] \text{ and } m_j^2 = 1 \text{ for some } [\theta'] \text{ for all } j \neq i \right.\right\},$$

and the set of messages of the remaining agents in which Rule 3 is triggered is the complement set, defined by:

$$M_{-i}^3 \triangleq M_{-i} \setminus M_{-i}^2.$$

Suppose first that agent i has a belief $\lambda_i \in \Delta(M_{-i})$ under which Rule 3 is triggered with positive probability, so that $\lambda_i\left(M_{-i}^3\right) > 0$. Note that if agent i plays m_i, with strictly positive probability $\underline{y}$ is provided. Hence, because from (25), $y^*(i,[\theta]) \in \mathcal{Y}$ is such that $u_i(y^*(i,[\theta]),\theta) > u_i(\underline{y},\theta)$, i's expected utility conditional on Rule 3 i.e., $\sum_{m_{-i} \in M_{-i}^3} \lambda_i(m_{-i})\, u_i(g(m_i, m_{-i}),\theta)$, is strictly smaller

[9] This rule is well-defined because $\mathcal{P}$ is finer than $\mathcal{P}_f$.

than $\sum_{m_{-i}\in M^3_{-i}} \lambda_i(m_{-i}) \max_{y\in\mathcal{Y}} u_i(y,\theta)$. Now, if i deviates to $\widehat{m}_i = (\widehat{m}_i^1, \widehat{m}_i^2, \widehat{m}_i^3, \widehat{m}_i^4)$ where $\widehat{m}_i^4 \in \arg\max_{y\in\mathcal{Y}} u_i(y,\theta)$, it is easily checked that i's expected utility conditional on Rule 3 tends to

$$\sum_{m_{-i}\in M^3_{-i}} \lambda_i(m_{-i}) \max_{y\in\mathcal{Y}} u_i(y,\theta)$$

as $\widehat{m}_i^2$ tends to infinity. Thus, player i can always improve his expected payoff conditional on Rule 3 deviating from m_i to $\widehat{m}_i$ and announcing $\widehat{m}_i^2$ large enough.

Now suppose that agent i believes that Rule 2 will be triggered with positive probability, so that $\lambda_i\left(M^2_{-i}\right) > 0$. We again consider a deviation to $\widehat{m}_i = (\widehat{m}_i^1, \widehat{m}_i^2, \widehat{m}_i^3, \widehat{m}_i^4)$ and observe that the choice of $\widehat{m}_i^4$ does not affect the outcome of the mechanism conditional on Rule 2. We also note that for any $m_{-i} \in M^2_{-i}$ such that $\lambda_i(m_{-i}) > 0$, (m_i, m_{-i}) does not trigger Rule 2(ii). Indeed, if it were the case, we would have $u_i(g(m_i,m_{-i}),\theta) = u_i(z_i(m^1_{-i}, m^1_{-i}),\theta)$. We have to distinguish two cases: whether players $j \neq i$ send message $[\theta]$ or not. First, consider the case where[10] $m^1_{-i} \neq [\theta]$, player i could change m_i to $\widehat{m}_i$ having $\widehat{m}_i^3(m^1_{-i}) = z_i([\theta], m^1_{-i})$ and keeping m_i unchanged otherwise. By Lemma 4, $u_i\left(f\left(m^1_{-i}\right), \widehat{\theta}\right) > u_i(z_i([\theta], m^1_{-i}), \widehat{\theta}) = u_i\left(\widehat{m}_i^3\left(m^1_{-i}\right), \widehat{\theta}\right)$ for all $\widehat{\theta} \in m^1_{-i}$, and so by construction of the mechanism, $(\widehat{m}_i, m_{-i})$ now triggers Rule 2(i). Again using Lemma 4 and the fact that $m^1_{-i} \neq [\theta]$, we get

$$\begin{aligned} & u_i(g(m_i,m_{-i}),\theta) = u_i(z_i(m^1_{-i}, m^1_{-i}),\theta) \\ < \;& \left(1 - 1/(m_i^2+1)\right) u_i(z_i([\theta], m^1_{-i}),\theta) + \left(1/(m_i^2+1)\right) u_i(z_i(m^1_{-i}, m^1_{-i}),\theta) = u_i(g(\widehat{m}_i, m_{-i}),\theta). \end{aligned}$$

Hence, the expected utility of player i would strictly increase, which yields contradiction. Consider the second case where $m^1_{-i} = [\theta]$, player i could change m_i to $\widehat{m}_i$ having $\widehat{m}_i(m^1_{-i}) = f(\theta)$ and keeping m_i unchanged otherwise. It is clear that by construction of the mechanism, $(\widehat{m}_i, m_{-i})$ now triggers Rule 2(i). Since Lemma 4 gives us

$$\begin{aligned} & u_i(g(m_i,m_{-i}),\theta) = u_i(z_i(m^1_{-i}, m^1_{-i}),\theta) \\ < \;& \left(1 - 1/(m_i^2+1)\right) u_i(f(\theta),\theta) + \left(1/(m_i^2+1)\right) u_i\left(z_i\left([\theta],[\theta]\right),\theta\right) = u_i(g(\widehat{m}_i, m_{-i}),\theta) \end{aligned}$$

the expected utility of player i would strictly increase, which here again yields a contradiction. So now we know that for any $m_{-i} \in M^2_{-i}$ such that $\lambda_i(m_{-i}) > 0$, (m_i, m_{-i}) does not trigger Rule 2(ii). Using a similar reasoning, it is easily shown that for any $m_{-i} \in M^2_{-i}$ such that $\lambda_i(m_{-i}) > 0$, we must have $u_i(m_i^3(m^1_{-i}),\theta) > u_i(z_i(m^1_{-i}, m^1_{-i}),\theta)$, hence the expected payoff conditional on Rule 2 from m_i

$$\sum_{m_{-i}\in M^2_{-i}} \lambda_i(m_{-i}) \left[\left(1 - 1/(m_i^2+1)\right) u_i(m_i^3(m^1_{-i}),\theta) + (1/(m_i^2+1)) u_i(z_i(m^1_{-i}, m^1_{-i}),\theta)\right]$$

[10] Whenever $m^1_j = [\theta]$ for all $j \neq i$, we sometimes abuse notations and write $m^1_{-i} = [\theta]$.

is strictly increasing in m_i^2. It follows that the choice of $\widehat{m}_i$ with $\widehat{m}_i^2$ large and strictly larger than m_i^2 strictly improves the expected utility of agent i in case Rule 2 or 3 is triggered, which yields the desired contradiction.

Claim 2. $\left([\theta], 1, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$ for all i, θ, m_i^3, m_i^4.

Proof of Claim 2. Suppose that player i in state θ puts probability 1 on each other agent j sending a message of the form $([\theta], 1, m_j^3, m_j^4)$. If player i announces a message of the form $\left([\theta], 1, m_i^3, m_i^4\right)$, he gets payoff $u_i\left(f\left(\theta\right), \theta\right)$. If he announces a message not of this form, the outcome is determined by Rule 2. Since by Lemma 4, $u_i(z_i([\theta], [\theta]), \theta) < u_i(f(\theta), \theta)$, it is clear that by construction of the mechanism, his payoff from invoking Rule 2 is bounded above by $u_i\left(f\left(\theta\right), \theta\right)$.

Claim 3. If $m_i = \left([\theta'], 1, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$, then $[\theta'] = [\theta]$.

Proof of Claim 3. Suppose $m_i = \left([\theta'], 1, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$. Given the message m_i, we can define the set of messages of the remaining agents which trigger Rule 1, 2 or 3, respectively. In particular, we define M_{-i}^1 for the set of $m_{-i} \in M_{-i}$ such that (m_i, m_{-i}) triggers Rule 1. Similarly, $M_{-i}^{2,i}$ is defined as the set of $m_{-i} \in M_{-i}$ such that (m_i, m_{-i}) triggers Rule 2 where player i is the deviating player. Now consider a given belief λ_i of agent i. If $\lambda_i\left(\left\{m_{-i} \in M_{-i}^1\left(m_i\right)\right\}\right) = 0$, then Rule 2 or 3 will be triggered with probability one. Although, Rule 2 can now be triggered with a "deviating player" being different of i, it is easily checked that a similar argument as in Claim 1 applies and so the message m_i cannot be a best reply by agent i. Suppose now that the belief λ_i of agent i is such that:

$$0 < \lambda_i\left(\left\{m_{-i} \in M_{-i}^1\right\}\right) < 1. \tag{28}$$

While we still argue that agent i can strictly increase his expected utility by selecting an integer $\widehat{m}_i^2 > 1$, we observe that a complication arises as with λ_i given by (28), a choice of $\widehat{m}_i^2 > 1$ leads from an allocation determined by Rule 1 to an allocation determined by Rule 2, and hence the realization of an unfavorable allocation $\underline{y}$ with positive probability. But now we observe that by selecting $\widehat{m}_i$ such that:

$$\widehat{m}_i^3\left([\hat{\theta}]\right) = \begin{cases} f\left(\theta'\right) \text{ if } [\hat{\theta}] = [\theta'] \\ m_i^3\left([\hat{\theta}]\right) \text{ otherwise,} \end{cases}$$

$\widehat{m}_i^4 \in \arg\max_{y \in \mathcal{Y}} u_i(y, \theta)$ and by choosing an integer $\widehat{m}_i^2$ sufficiently large, the small loss in Rule 2 can always be offset by a gain in Rule 3 relative to the allocation achieved under $g\left(m_i, m_{-i}\right)$. More formally, for $m_i = \left([\theta'], 1, m_i^3, m_i^4\right)$, since $0 < \lambda_i\left(\left\{m_{-i} \in M_{-i}^1\right\}\right) < 1$ and since – as claimed before – for all $m_{-i} \in M_{-i}^{2,i}$ such that $\lambda_i(m_{-i}) > 0$, $u_i(m_i^3(m_{-i}^1), \theta) > u_i(z_i(m_{-i}^1, m_{-i}^1), \theta)$, i's expected payoff from playing m_i is strictly lower than

$$\sum_{m_{-i} \in M_{-i}^1} \lambda_i\left(m_{-i}\right) u_i(f(\theta'), \theta) + \sum_{m_{-i} \in M_{-i}^{2,i}} \lambda_i\left(m_{-i}\right) u_i(m_i^3(m_{-i}^1), \theta) + \sum_{m_{-i} \notin M_{-i}^1 \cup M_{-i}^{2,i}} \lambda_i\left(m_{-i}\right) \max_{y \in \mathcal{Y}} u_i(y, \theta)$$

while for $\widehat{m}_i = ([\theta'], \widehat{m}_i^2, \widehat{m}_i^3, \widehat{m}_i^4)$, it is easily checked that as $\widehat{m}_i^2$ tends to infinity, i's expected payoffs tend toward the expression above. Hence, choosing $\widehat{m}_i^2$ large enough, $\widehat{m}_i$ is a better response against λ_i for player i than m_i, a contradiction.

So if $m_i = \left([\theta'], 1, m_i^3, m_i^4\right) \in S_i^{\mathcal{M},\theta}$, it follows player i must be convinced that each other player must be choosing a message of the form $\left([\theta'], 1, m_j^3, m_j^4\right)$, and hence

$$\lambda_i\left(\left\{m_{-i} \in M_{-i}^1\right\}\right) = 1.$$

Thus there must exist a message of the form $m_j = \left([\theta'], 1, m_j^3, m_j^4\right) \in S_j^{\mathcal{M},\theta}$ for all j. Now, proceed by contradiction and assume that $[\theta'] \neq [\theta]$ and so that $\theta \notin [\theta']$. By strict Maskin monotonicity*, we know that there exist j such that for all $\widehat{\theta} \in [\theta'] : u_j\left(f\left(\theta'\right), \widehat{\theta}\right) > u_j\left(y([\theta'], \theta), \widehat{\theta}\right)$ and $u_j\left(y([\theta'], \theta), \theta\right) > u_j\left(f\left(\theta'\right), \theta\right)$. By the above argument, we know that player j's belief against which $m_j = \left([\theta'], 1, m_j^3, m_j^4\right)$ is a best reply assigns probability one to each player $l \neq j$ sending a message of the form $m_l = \left([\theta'], 1, m_l^3, m_l^4\right)$. Hence, player j's expected payoff from playing m_j is $u_j(f(\theta'), \theta)$ while if j deviates to $\hat{m}_j = \left([\theta'], \hat{m}_j^2, \hat{m}_j^3, m_j^4\right)$ where $\hat{m}_j^2 > 1$ and

$$\hat{m}_j^3([\hat{\theta}]) = \begin{cases} y([\theta'], \theta) \text{ if } [\hat{\theta}] = [\theta'] \\ m_i^3\left([\hat{\theta}]\right) \text{ otherwise,} \end{cases}$$

player j believes with probability one that Rule 2 (i) will be triggered. Hence, player j's expected payoff would be

$$\left(1 - 1/(\widehat{m}_j^2 + 1)\right) u_j(y([\theta'], \theta), \theta) + (1/(\widehat{m}_j^2 + 1)) u_j(z_j([\theta'], [\theta']), \theta).$$

Note that as $\hat{m}_j^2$ tends to infinity, this expression tends to $u_j(y([\theta'], \theta), \theta)$ which is strictly larger than $u_j\left(f\left(\theta'\right), \theta\right)$. Hence for $\hat{m}_j^2$ large enough, $\hat{m}_j$ is better response for player j than $\left([\theta'], 1, m_j^3, m_j^4\right)$, a contradiction. Thus $[\theta'] = [\theta]$ as claimed.

Completion of proof. Claims 1, 2 and 3 together imply that for each $\theta : S_i^{\mathcal{M},\theta} \neq \varnothing$ and $m_i \in S_i^{\mathcal{M},\theta} \Rightarrow m_i^2 = 1$ and $m_i^1 = [\theta]$. Thus $S^{\mathcal{M},\theta} \neq \varnothing$ and $m \in S^{\mathcal{M},\theta} \Rightarrow g(m) = f(\theta)$. ■