Universal Mechanisms and Moral Preferences in Implementation[+]

Hitoshi Matsushima[*]

Faculty of Economics, University of Tokyo

First Version: March 4, 2002
This version: December 18, 2003

Abstract

This paper reconsiders implementation of social choice functions defined as mapping from states to consequences, where we require the uniqueness of equilibrium outcome at every state. In contrast with the standard models, we construct only mechanisms that are universal, i.e., are free from the detail of the model specification such as the set of states, and allow each agent to have small moral preference. We show that a single mechanism can implement every incentive compatible social choice function. Moral preferences serve not only to eliminate unwanted equilibria but also to make the central planner's information processing simplified as much as possible in ways that each agent will translate her indescribable private signal into the describable characteristic of the socially optimal alternative.

Keywords: Implementation, Recommendations, Moral Preferences, Indescribability, Universal Mechanisms.

JEL Classification Numbers: C72, D71, D78, H41

[+] This paper includes the contents of my old manuscripts entitled "Honesty-Proof Implementation" (Discussion Paper CIRJE-F-178, Faculty of Economics, University of Tokyo, 2002) and "Implementation and Preference for Honesty" (Discussion Paper CIRJE-F-244, Faculty of Economics, University of Tokyo, 2003). I would like to thank Mr. Daisuke Shimizu for his careful reading. All errors are mine. The research for this paper was supported by Grant-In-Aid for Scientific Research (KAKENHI 15330036) from JSPS and MEXT of the Japanese Government.

[*] Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan. Fax: +81-3-3818-7082. E-mail: hitoshi@e.u-tokyo.ac.jp

1. Introduction

This paper demonstrates a new approach to the implementation problem, where a social choice function defined as a mapping from states to consequences is said to be implementable in terms of any equilibrium concept if we can construct a mechanism in which at every state there exists the *unique* equilibrium outcome and this outcome equals the value of the social choice function. The mechanisms used in the previous works in the implementation literature depended crucially on the very detail of the model specification such as the set of states. In real situations, however, it might be impossible to describe this detail on a document, because of its complexity. Hence, transaction-cost economists sometimes criticize implementation theory, because the constructed mechanisms are difficult to put into practice.[1] Based on this observation, this paper reconsiders the implementation problem by investigating the possibility that a *single* mechanism, which is *universal*, i.e., is *not* tailored to any particular model specification, can implement a wide variety of social choice functions.

This paper considers the following public decision procedure with complete information, where which alternative is socially optimal is common knowledge among agents but is unknown to the central planner. The central planner requires each agent to make multiple announcements about which alternative is to be recommended as the public decision. The central planner then randomly picks up a message profile from their multiple announcements. If sufficiently many agents announce the same alternative that is enforceable by the central planner, then the central planner will decide on it. Otherwise, the central planner will decide on the status quo alternative. Here, we assume that it is verifiable to the court whether the recommendations by agents are enforceable or not. This procedure does not depend on the set of states, and therefore, we do *not* require the set of states to be describable.

Unfortunately, any mechanism based on this procedure fails to work in the standard models of implementation, where each agent is assumed to have preference *only* for consequences. See Moore (1992), Palfrey (1992), and Maskin and Sjostrom (2002) for the surveys of the standard models of implementation.[2] Maskin and Tirole (1999) and Tirole (1999) argued that if a social choice function is implementable in Nash equilibrium or any other equilibrium concept with complete information, then any factors of the state other than agents' preferences on which the social choice function depends must be known to the central planner, describable on a document, and verifiable to the court. On the other hand, there exist many important attempts to establish ideas on the theoretical foundations of social choice and welfare such as Rawls (1971), Dworkin (1981), and Sen (1982, 1985, 1999), all of which are based on their respective ethical factors of the state other than

[1] For the surveys on transaction-cost economics and incomplete contract theory, see Milgrom and Roberts (1992) and Hart (1995).

[2] Eliaz (2002) took into account factors other than individuals' preferences for consequences such as bounded rationality. Glazer and Rubinstein (1998) allowed agents to have non-consequentialist preferences.

individuals' preferences for consequences.[3] The relevance of these factors to social choice and welfare might be too complicated to be described on a document. In order to verify which alternative is to be socially optimal, however, it might be necessary for the central planner to describe this relevance. Hence, we can conclude that ethically important social choice functions are never implementable in the standard models of implementation.[4] In order to implement them, it might be inevitable to take into account the possibility that agents have preferences not only for consequences but also for anything non-consequential.

Based on this observation, this paper will assume that some agents have *moral* preferences in a sense that they have positive psychological costs for recommending any alternative other than the socially optimal alternative. Several works such as Erard and Feinstein (1994), Alger and Ma (2003), and Deneckere and Severinov (2001) examined the case that agents' ability to manipulate information is limited and demonstrated that including agents who have preference for honesty could significantly alter the model. These works assumed that the cost for reporting dishonestly is sufficiently large, while the present paper will allow the maximal total cost for immorality to be *as close to zero as possible*.

The results of this paper are very permissive. In particular, we show that *there exists a single mechanism with small fines that implements any social choice function in iterative dominance whenever at least one agent has moral preference*. Here, we do *not* even require the set of alternatives to be describable. All we have to require is that at every state the value of the social choice function and the status quo alternative are describable. This point is in contrast with the standard models of implementation, where in order to eliminate unwanted equilibria the central planner has to incentivize agents to announce mostly full information about their preference profile honestly. This inevitably requires the set of alternatives to be describable. Hence, agents' moral preferences will play a powerful role in making the central planner's information processing simplified as much as possible.

We will extend the above arguments to the *incomplete* information environments, where an alternative is defined as a bundle of characteristics, and every describable alternative is assumed to be enforceable. Each agent receives her private signal that has partial information about the characteristics of the socially optimal alternative. A state is defined as a profile of agents' private signals. We assume that each characteristic of the socially optimal alternative is known to at least one agent. Hence, a social choice function is described by a profile of mappings from private signals to profiles of characteristics.

We consider the following public decision procedure. The central planner requires each agent to make multiple announcements about what the true characteristics of the socially optimal alternative are. The central planner then randomly picks up a message profile from their multiple announcements as the alternative that agents jointly recommend,

---

[3] Rawls introduced primary goods. Dworkin introduced compensation and responsibility. Sen introduced liberty, functioning, and capabilities. See Basu, Pattanaik, and Suzumura (1995), Sen (1999), and Suzumura (2002) for the surveys on social choice and welfare.

[4] Moreover, Serrano and Vohra (2001) investigated the economic environments with incomplete information where agents' preferences are the same across states but their initial endowments depend on the state. They showed that no individually rational social choice function is implementable.

and will decide on it. This procedure does not depend on the set of states, and therefore, we do not require the set of states to be describable.

Similarly to the complete information environments, we assume that all agents have moral preferences in a sense that they have positive psychological costs for making dishonest announcements about the characteristics of the socially optimal alternative. Their costs can be as close to *zero* as possible. We then show that *there exists a single mechanism with small fines that implements any social choice function in Bayesian iterative dominance with respect to any probability structure if this social choice function and this probability structure satisfy incentive compatibility associated with agents' psychological costs.*

In general, as agents' psychological costs become large, the incentive compatibility constraint becomes weaker, and therefore, the range of implementable social choice functions expands. Even if their costs are close to zero, their moral preferences will play the significant role in eliminating unwanted equilibria. In fact, in the standard models incentive compatibility is not sufficient for implementability in terms of any equilibrium concept with incomplete information[5], while it is sufficient in our model. Agents' moral preferences also serve to make the central planner free from any complicated information processing because each agent is well incentivized to translate her private signal that may be indescribable into the characteristics of the socially optimal alternative that are describable. This will be the driving force of making the mechanism *free* from the detail of the model specification. In particular, the mechanisms is independent of the *probability structure*, as well as of the social choice function.

The organization of this paper is as follows. Section 2 shows the basic model with complete information. Section 3 considers the case where each agent makes only a *single* announcement. We show that with five or more agents there exists a mechanism with no fine that implements any alternative in pure strategy Nash equilibrium if all agents regard it as being socially optimal and a majority of the agents prefer it to agent 1's most favorite alternative. Sections 4 and 5 consider the case where each agent makes multiple announcements. Section 4 supposes that at least one agent has moral preference in a *minimal* sense that whenever the other agents recommend *only* the socially optimal alternative then she has positive psychological cost for immorality. We show that with four agents there exists a mechanism with small fines that implements any alternative in pure strategy Nash equilibrium if this agent regards it as being socially optimal. We argue that this result holds *even if* no agent has moral preference. Section 5 supposes that at least one agent has moral preference in the original sense. We show that with three agents there exists a mechanism with small fines that implements any alternative in iterative dominance if this agent regards it as being socially optimal. Section 6 explains the implications of the above results in the implementation literature. Section 7 investigates the incomplete information environments, and shows the possibility result by constructing mechanisms with small fines and multiple announcements.

---

[5] See Jackson (1991), Abreu and Matsushima (1992b), Matsushima (1993), Duggan (1997), Serrano and Vohra (2000), and others.

2. Basic Model

Let $N = \{1,...,n\}$ denote the set of *agents* where $n \geq 2$. Let A denote the set of *alternatives*. We may assume for a while that A is describable on a document and enforceable by the central planner. As we will argue later, however, our results do not depend on this assumption. Let Δ denote the set of *simple* lotteries over alternatives. Let M_i denote the set of *messages* for each agent $i \in N$. Let $M = \prod_{i \in N} M_i$ denote the set of message profiles.

Fix a positive real number $\varepsilon > 0$ arbitrarily. Given the set of message profiles M, a *mechanism* is defined by $G = (x,t)$, where $x : M \to \Delta$, $t = (t_i)_{i \in N}$, $t_i : M \to [-\varepsilon, \infty)$, and t satisfies the *budgetary constraint* in the sense that $\sum_{i \in N} t_i(m) \leq 0$ for all $m \in M$. When the agents announce a message profile $m = (m_i)_{i \in N} \in M$, the central planner will choose an alternative according to the lottery $x(m) \in \Delta$ and make a monetary transfer $t_i(m) \in [-\varepsilon, \infty)$ to each agent $i \in N$. We regard ε as the *upper bound* of monetary fines. We write $x(m) = a$ if $x(m)(a) = 1$.

A *utility function* for each agent $i \in N$ is defined by $u_i : A \times R \times M \to R$, where $u_i(a, t_i, m)$ denotes agent $i's$ utility when the agents announce the message profile $m \in M$ and the central planner chooses the alternative $a \in A$ and makes the monetary transfer $t_i \in R$ to agent i. Here, we will allow the agents' announcements to have intrinsic value for each agent's welfare. We assume the expected utility hypothesis with respect to alternatives. For every lottery $\alpha \in \Delta$, let $u_i(\alpha, t_i, m) = \sum_{a \in \Gamma} u_i(a, t_i, m)\alpha(a)$, where Γ is the support of α. Denote $u_i(\alpha, m) = u_i(\alpha, 0, m)$. Let $u = (u_i)_{i \in N}$ denote a utility function profile.

A message profile $m \in M$ is said to be a *Nash equilibrium in the game defined by* (G,u) if for every $i \in N$ and every $m_i' \in M_i$,

$$u_i(x(m), t_i(m), m) \geq u_i(x(m_i', m_{-i}), t_i(m_i', m_{-i}), m_i', m_{-i}).$$

An alternative $a^* \in A$ is said to be *implemented in Nash equilibrium by a mechanism* G *with respect to a utility function profile* u if there exists the unique Nash equilibrium $m \in M$ in (G,u), and this message profile satisfies

$$x(m) = a^*, \text{ and } t_i(m) = 0 \text{ for all } i \in N.$$

For every $i \in N$, let $M_i^{(0,G,u)} = M_i$. Recursively, for every $i \in N$ and every $r = 1,2,...$, let $M_i^{(r,G,u)} \subset M_i^{(r-1,G,u)}$ denote the set of messages $m_i \in M_i^{(r-1,G,u)}$ for agent i such that there exists no $m_i' \in M_i^{(r-1,G,u)}$ satisfying that for every $m_{-i} \in M_{-i}^{(r-1,G,u)}$,

$$u_i(x(m_i', m_{-i}), t_i(m_i', m_{-i}), m_i', m_{-i}) > u_i(x(m), t_i(m), m),$$

where $M_{-i}^{(r,G,u)} = \prod_{j \in N/\{i\}} M_j^{(r,G,u)}$. Let $M^{(r,G,u)} = \prod_{i \in N} M_i^{(r,G,u)}$ and $M^{(\infty,G,u)} = \bigcap_{r=1}^{\infty} M^{(r,G,u)}$. A message profile $m \in M$ is said to be *iteratively undominated in the game* (G,u) if

$$m \in M^{(\infty,G,u)}.$$

An alternative $a^* \in A$ is said to be *implemented in iterative dominance by a mechanism* G *with respect to a utility function profile* u if there exists the unique iteratively undominated message profile m in (G,u), and this message profile satisfies

$$x(m) = a^*, \text{ and } t_i(m) = 0 \text{ for all } i \in N.$$

Here, implementation in Nash equilibrium does not imply the uniqueness of *mixed* strategy Nash equilibrium, whereas implementation in iterative dominance implies this uniqueness.

In this paper except for in Subsection 6.2, we will assume that for every $i \in N$, there exists a positive integer $K_i > 0$ such that

$$M_i = A^{K_i}.$$

Hence, the central planner will require each agent $i \in N$ to make K_i announcements about which alternative to be recommended as the public decision.[6] Let $m_i = (m_i^h)_{h=1}^{K_i} \in M_i$. Fix an alternative $a^* \in A$ arbitrarily, which is regarded as the *socially optimal* alternative. Let $m^* \in M$ denote the message profile such that for every $i \in N$ and every $h \in \{1,...,K_i\}$,

$$m_i^{*h} = a^*,$$

where each agent recommends *only* the socially optimal alternative a^*.

In this paper except for in Subsection 6.2, we will confine our attentions to utility functions u_i for each agent $i \in N$ where there exist a function $v_i : A \to R$, a positive real number $c_i > 0$, and a function $r_i : M \to [0,1]$ such that

$$u_i(a,t_i,m) = v_i(a) + t_i - r_i(m)c_i \text{ for all } a \in A, \text{ all } t_i \in R, \text{ and all } m \in M,$$

and

$$r_i(m_i^*, m_{-i}) = 0 \text{ for all } m_{-i} \in M_{-i}$$

are satisfied. We regard c_i as the upper bound of agent $i's$ psychological cost for recommending any alternatives other than the socially optimal alternative a^*, which is caused by her moral sentiment. Each agent has no such cost received if she announces only the socially optimal alternative.[7] The function $v_i(\cdot)$ is regarded as agent $i's$ preference for consequences. The mechanisms constructed in this paper will not much depend on how

---

[6] Each agent simultaneously announces multiple messages at once. This may exclude any complexity of agents' psychological interaction observed in laboratory experiments. See Fehr and Schmidt (2003).

[7] We assume quasi-linearity and risk neutrality for simplicity of arguments. We can drop this assumption with only minor changes.

to specify $(v_i(\cdot))_{i\in N}$. This point is in contrast with the standard models of implementation where the construction of mechanism is tailored to particular specifications of $(v_i(\cdot))_{i\in N}$.

3. Single Recommendation with No Fines

This section assumes that $n \geq 5$, n is odd,[8] and $K_i = 1$ for all $i \in N$, i.e.,

$$M_i = A \text{ for all } i \in N.$$

Hence, each agent makes a *single* announcement about which alternative to be recommended. This section assumes that *no* fines are available.

We specify a mechanism $G^* = (x^*, t^*)$ as follows, where $t_i^*(m) = 0$ for all $i \in N$ and all $m \in M$. Fix $m \in M$ arbitrarily. If there exists $a \in A$ such that

$$m_i = a \text{ for at least } \frac{n+1}{2} \text{ agents } i \in N/\{1\} \text{ other than agent 1,}$$

then

$$x^*(m) = a.$$

If there exists $a \in A/\{m_1\}$ such that

$$m_i \in \{a, m_1\} \text{ for all } i \in N,$$

$$m_i = a \text{ for } \frac{n-1}{2} \text{ agents } i \in N/\{1\},$$

and

$$m_i = m_1 \text{ for } \frac{n-1}{2} \text{ agents } i \in N/\{1\},$$

then

$$x^*(m) = a.$$

Otherwise,

$$x^*(m) = m_1.$$

The central planner will regard agent 1 as the *dictator* with the following restrictions. If there is an alternative that is recommended by a *majority* of the agents other than agent 1, then the central planner will choose this alternative. If just $\frac{n-1}{2}$ agents other than agent 1 recommend the same alternative and all other agents agree with agent 1, then the central planner will choose this alternative.[9] Otherwise, the central planner will choose the alternative that agent 1 recommends.

For every $a^* \in A$, we define $U^*(a^*)$ as the set of utility function profiles u satisfying the following three properties.

(i) For every $i \in N$ and every $m \in M$, if $m_i \neq a^*$, then

---

[8] With minor changes, we can apply the argument of this section to the case where n is even.

[9] Hence, G^* cannot be regarded as being *majority-based* in that whenever there is an alternative recommended by a majority of agents then the central planner will choose this alternative.

$r_i(m) = 1$.

(ii) There exists $a^1 \in A$ such that

$v_1(a^1) - c_1 > v_1(a)$ for all $a \in A/\{a^1, a^*\}$.

(iii) There exist at least $\frac{n+1}{2}$ agents $i \in N/\{1\}$ such that

$v_i(a^*) > v_i(a^1) - c_i$.

Property (i) implies that every agent has moral preference in the sense that she prefers the announcement of the socially optimal alternative to any other announcement whenever the alternative and monetary transfer are unchanged. Note that for every $i \in N$, every $a \in A$, every $t_i \in R$, and every $m \in M$,

$$u_i(a, t_i, m) = v_i(a) + t_i \text{ if } m_i = a^*,$$

and

$$u_i(a, t_i, m) = v_i(a) + t_i - c_i \text{ if } m_i \neq a^*.$$

Property (ii) implies that a^1 is regarded as agent 1's *most* favorite alternative except for a^* at the expense of the cost c_1. Property (iii) implies that a^* is preferred to a^1 by a majority of agents.

Theorem 1: *Any alternative $a^* \in A$ is implemented in Nash equilibrium by G^* with respect to all $u \in U^*(a^*)$.*

Proof: Note that

$$x^*(m^*) = a^*.$$

It is clear from $n \geq 5$ that m^* is a Nash equilibrium in (G^*, u), because for every $i \in N$ and every $m_i \in M_i/\{m_i^*\}$,

$$x^*(m_i, m_{-i}^*) = a^*,$$

and therefore, it follows from property (i) that

$$u_i(x^*(m^*), m^*) = v_i(a^*) > v_i(a^*) - c_i = u_i(x^*(m_i, m_{-i}^*), m_i, m_{-i}^*).$$

Fix $\hat{m} \in M/\{m^*\}$ arbitrarily, and suppose that $\hat{m}$ is a Nash equilibrium in (G^*, u). Suppose $x^*(\hat{m}) = a^*$. Then, for every $i \in N/\{1\}$,

$$x^*(m_i^*, \hat{m}_{-i}) = a^*,$$

and therefore, it follows from property (i) that if $\hat{m}_i \neq a^*$, then

$$u_i(x^*(\hat{m}), \hat{m}) = v_i(a^*) - c_i < v_i(a^*) = u_i(x^*(m_i^*, \hat{m}_{-i}), m_i^*, \hat{m}_{-i}).$$

Hence, it must hold that $\hat{m}_i = a^*$ for all $i \in N/\{1\}$, but $\hat{m}_1 \neq a^*$. This contradicts the Nash equilibrium property, because agent 1 has incentive to announce a^* instead of $\hat{m}_1$. Hence, without loss of generality, we will assume

$x^*(\hat{m}) \neq a^*$.

First, suppose that there exist at least $\frac{n+1}{2}$ agents $i \in N/\{1\}$ such that

$\hat{m}_i = x^*(\hat{m})$.

Then, every agent who announces neither $x^*(\hat{m})$ nor a^* has incentive to announce a^* because of property (i), i.e., because for every $i \in N$, if $\hat{m}_i \notin \{x^*(\hat{m}), a^*\}$, then

$x^*(m_i^*, \hat{m}_{-i}) = x^*(\hat{m})$,

and therefore,

$u_i(x^*(\hat{m}), \hat{m}) = v_i(x^*(\hat{m})) - c_i < v_i(x^*(\hat{m})) = u_i(x^*(m_i^*, \hat{m}_{-i}), m_i^*, \hat{m}_{-i})$.

This contradicts the Nash equilibrium property. Hence, it must hold that

$\hat{m}_i \in \{x^*(\hat{m}), a^*\}$ for all $i \in N$.

Note that agent 1 has incentive to announce a^* because of property (i) and $x^*(m_1^*, \hat{m}_{-1}) = x^*(\hat{m})$. Hence, it must hold that $\hat{m}_1 = a^*$. Note that for every $i \in N/\{1\}$, $x^*(m_i^*, \hat{m}_{-i}) = x^*(\hat{m})$, and therefore, it follows from property (i) that each agent $i \in N/\{1\}$ has incentive to announce a^*. This contradicts the Nash equilibrium property.

Next, suppose that

$x^*(\hat{m}) \notin \{\hat{m}_1, a^*\}$,

$\hat{m}_i = x^*(\hat{m})$ for $\frac{n-1}{2}$ agents $i \in N/\{1\}$,

and

$\hat{m}_i = \hat{m}_1$ for $\frac{n-1}{2}$ agents $i \in N/\{1\}$.

Then, agent 1 is regarded as being dictatorial in the sense that for every $m_1 \in M_1$,

$x^*(m_1, \hat{m}_{-1}) = \hat{m}_1$ if $m_1 = x^*(\hat{m})$,

and

$x^*(m_1, \hat{m}_{-1}) = m_1$ if $m_1 \notin \{x^*(\hat{m}), \hat{m}_1\}$.

Hence, from property (ii), it must hold that $x^*(\hat{m}) = a^1$, and therefore, $a^1 \neq a^*$. This implies $\hat{m}_1 = a^*$, because if $\hat{m}_1 \neq a^*$, then every agent $i \in N/\{1\}$ who announces $\hat{m}_i = \hat{m}_1$ has incentive to announce a^* instead of $\hat{m}_1$. Note that for every $i \in N/\{1\}$, if $\hat{m}_i = a^1$, then

$x^*(m_i^*, \hat{m}_{-i}) = a^*$.

It follows from property (iii) that there exists an agent $i \in N/\{1\}$ such that $\hat{m}_i = a^1$ and

$v_i(a^*) > v_i(a^1) - c_i$.

Hence, this agent has incentive to announce a^* instead of a^1, because

$u_i(x^*(\hat{m}), \hat{m}) = v_i(a^1) - c_i < v_i(a^*) = u_i(x^*(m_i^*, \hat{m}_{-i}), m_i^*, \hat{m}_{-i})$.

This contradicts the Nash equilibrium property.

Finally, suppose that the above two suppositions do not hold. Then, it must be that

$$x^*(\hat{m}) = \hat{m}_1,$$

and

$$x^*(m_1, \hat{m}_{-1}) = m_1 \quad \text{for all} \quad m_1 \in M_1,$$

and therefore,

$$x^*(\hat{m}) = \hat{m}_1 = a^1$$

hold. Note that for every $i \in N/\{1\}$,

$$\text{either} \quad x^*(m_i^*, \hat{m}_{-i}) = a^1 \quad \text{or} \quad x^*(m_i^*, \hat{m}_{-i}) = a^*.$$

It follows from properties (i) and (iii) that there exists an agent $i \in N/\{1\}$ who announces $\hat{m}_i \neq a^*$ but has incentive to announce a^*, where

$$u_i(x^*(\hat{m}), \hat{m}_i) = v_i(a^1) - c_i$$
$$< \min[v_i(a^1), v_i(a^*)] \leq u_i(x^*(m_i^*, \hat{m}_{-i}), m_i^*, \hat{m}_{-i}).$$

This contradicts the Nash equilibrium property.

Hence, we have proved that any alternative $a^* \in A$ is implemented in Nash equilibrium by G^* with respect to all $u \in U^*(a^*)$.

Q.E.D.

Theorem 1 does not much depend on the assumption that all alternatives are describable and enforceable. *We only need the socially optimal alternative a^*, the status quo alternative $\bar{a}$, and agent 1's most favorite alternative a^1 to be describable and enforceable*. When agents could announce a message profile $m \in M$ in G^*, the alternative m_i for each $i \in N$ must be describable, and therefore, any alternative in the support of $x^*(m)$ is describable, because $x^*(m)(a) > 0$ only if $a \in \{\bar{a}, a^1, m_1, ..., m_n\}$. If an element of the support of $x^*(m)$ is not enforceable, then we may have to modify G^* by replacing it with another enforceable alternative such as $\bar{a}$, but this modification does not change the essence of the proof of Theorem 1.

The requirement implied by property (iii) that the socially optimal alternative is preferred to agent 1's most favorite by a majority of agents may be restrictive. In the next two sections, we will exclude this restriction and show more powerful possibility results than Theorem 1 by allowing small fines and multiple announcements.

4. Multiple Recommendations with Small Fines and Single Minimal Moralist

We allow the central planner to require each agent except for agent 1 to announce multiple recommendations, and allow small fines. We assume that agent 1 has moral preference in a *minimal* sense that she prefers the announcement of the socially optimal alternative to any other announcement whenever the other agents announce *only* the socially optimal alternative. We will not require the other agents to have moral preference. We then show that *there exists a mechanism that can implement any alternative in Nash equilibrium whenever agent 1 regards this alternative as being socially optimal.*

Assume that $n = 4$,[10] $K_1 = 1$, and there exists a positive integer $K > 0$ such that $K_2 = K_3 = K_4 = K$, i.e.,

$$M_1 = A \text{ and } M_2 = M_3 = M_4 = A^K.$$

Hence, the central planner requires agent 1 to make a single announcement and the other agents to make K announcements each. For every $h \in \{1,...,K\}$, let $m^h = (m_2^h, m_3^h, m_4^h)$.

Choose $\varepsilon > 0$ to be close to zero so that

(1) $c_1 > 3\varepsilon$.

Fix a positive real number $d > 0$ arbitrarily, and choose K to be large so that

(2) $K\varepsilon > d + c_i$ for all $i \in N/\{1\}$.

Note that we can choose $\varepsilon^+ \in (0, \varepsilon)$ to satisfy

(3) $K\varepsilon^+ + \rho > d + c_i$ for all $i \in N/\{1\}$,

where we denote $\rho = \varepsilon - \varepsilon^+ > 0$.

We specify a mechanism $G^+ = G^+(K, \varepsilon, d) = (x^+, t^+)$ as follows. Fix an alternative $\overline{a} \in A$ arbitrarily, which is regarded as the *status quo* alternative. We define $z : A^3 \to \Delta$ as follows. For every $\delta = (\delta_1, \delta_2, \delta_3) \in A^3$,

$$z(\delta) = a \text{ if } \delta_i = a \text{ for at least two components of } \delta,$$

and

$$z(\delta) = \overline{a} \text{ if } \delta_1 \neq \delta_2 \neq \delta_3 \neq \delta_1.$$

Fix $m \in M$ arbitrarily. Let

$$x^+(m) = \frac{\sum_{h=1}^{K} z(m^h)}{K},$$

where we regard $z(\delta)$ as a simple lottery such that $z(\delta)(a) = 1$ if $z(\delta) = a$. For every $h \in \{1,...,K\}$, with probability $\frac{1}{K}$, the central planner will choose $z(m^h)$, where for every

[10] This implies that there may exist five or more agents, but only four agents are required to participate in this decision procedure.

$a \in A$,

$$z(m^h) = a \quad \text{if} \quad m_i^h = a \quad \text{for at least two agents} \quad i \in \{2,3,4\},$$

and

$$z(m^h) = \overline{a} \quad \text{if} \quad m_2^h \neq m_3^h \neq m_4^h \neq m_2^h.$$

Note that $x^+(m)$ does not depend on m_1. For every $i \in \{2,3,4\}$, let

$$t_i^+(m) = -\varepsilon^+ - \frac{q_i(m)}{K}\rho \quad \text{if there exists} \quad h \in \{1,...,K\} \quad \text{such that}$$
$$m_2^{h'} = m_3^{h'} = m_4^{h'} = m_1 \quad \text{for all} \quad h' \in \{1,...,h-1\} \quad \text{and} \quad m_i^h \neq m_1^{K+1},$$

and

$$t_i^+(m) = -\frac{q_i(m)}{K}\rho \quad \text{otherwise,}$$

where $q_i(m) \in \{0,...,K\}$ is the number of agent $i's$ announcements that is not the same as agent 1's announcement, i.e.,

$$q_i(m) = \left|\{h \in \{1,...,K\} \mid m_i^h \neq m_1\}\right|.$$

Let

$$t_1^+(m) = -\sum_{i \in N/\{1\}} t_i^+(m),$$

and therefore, t^+ is *budget balancing* in that $\sum_{i \in N} t_i^+(m) = 0$ for all $m \in M$. Each agent $i \in \{2,3,4\}$ pays the monetary amount ε^+ to agent 1, in addition to $\frac{q_i}{K}\rho$, if and only if she is the first agent(s) to make a different announcement from agent 1's announcement.[11]

For every $a^* \in A$ and every positive real number $d > 0$, we define $U^+(a^*, d)$ as the set of utility function profiles u satisfying the following three properties.

(iv) For every $i \in \{2,3,4\}$, every $a \in A$, and every $a' \in A/\{a\}$,
$$\left|v_i(a) - v_i(a')\right| \leq d.$$

(v) For every $m_1 \in M_1/\{a^*\}$,
$$r_1(m_1, m_{-1}^*) = 1.$$

(vi) For every $i \in \{2,3,4\}$ and every $m \in M$,
$$r_i(m) \leq \frac{w_i(m)}{K},$$
where $w_i(m)$ is the number of agent $i's$ announcements that is not the same as the socially optimal alternative, i.e.,
$$w_i(m) = \left|\{h \in \{1,...,K_i\} \mid m_i^h \neq a^*\}\right|.$$

[11] We assume imperfect information.

Property (iv) implies that d is the upper bound of the utility differences for all agents except for agent 1. Property (v) implies that agent 1 has moral preference in the *minimal* sense that she prefers the announcement of the socially optimal alternative to any other announcement as long as the other agents recommend *only* the socially optimal alternative. Property (vi) allows each agent other than agent 1 to have *no* moral preference.

Theorem 2: *Suppose that inequalities (1) and (2) hold. Then, any alternative* $a^* \in A$ *is implemented in Nash equilibrium by* $G^+(K,\varepsilon,d)$ *with respect to all* $u \in U^+(a^*,d)$.

Proof: Note that

$$x^+(m^*) = a^*,$$

and

$$t_i^+(m^*) = 0 \quad \text{for all} \quad i \in N.$$

Fix $a \in A$ arbitrarily. Fix $h \in \{1,...,K\}$ and $m \in M$ arbitrarily, where

$$m_1 = a,$$

and

$$m_i^{h'} = a \quad \text{for all} \quad i \in N \quad \text{and all} \quad h' \in \{1,...,h-1\}.$$

First, consider any agent $i \in \{2,3,4\}$. Suppose $m_i^h \neq a$. Let $m_i' \in M_i$ be the message for agent i defined by

$$m_i'^h = a,$$

and

$$m_i'^{h'} = m_i^{h'} \quad \text{for all} \quad h' \in \{1,...,K\}/\{h\}.$$

If $m_j^h = a$ for all $j \in N/\{1,i\}$, then it follows that $x^+(m)$ is independent of m_i^h and $t_i^+(m_i', m_{-i}) - t_i^+(m) \geq \frac{\rho}{K} > 0$, which implies that agent i has incentive to announce m_i' instead of m_i. If $m_j^h \neq a$ for some $j \neq i$, then it follows that $t_i^+(m_i', m_{-i}) - t_i^+(m') = \varepsilon^+ + \frac{\rho}{K}$, which, together with properties (iv) and (vi) and the inequalities (3), implies that agent i has incentive to announce m_i' instead of m_i, where

$$u_i(x^+(m), t_i^+(m), m) - u_i(x^+(m_i', m_{-i}), t_i^+(m_i', m_{-i}), m_i', m_{-i})$$
$$\leq -\varepsilon^+ - \frac{\rho}{K} + \frac{1}{K}\{v_i(\overline{\alpha}) - v_i(z(m_i'^h, m_{-i}^h)) + c_i\} < -\varepsilon^+ - \frac{\rho}{K} + \frac{d + c_i}{K} < 0.$$ [12]

Next, suppose that

$$m_i^h = a \quad \text{for all} \quad i \in \{2,3,4\} \quad \text{and all} \quad h \in \{1,...,K\},$$

[12] This argument is related to Abreu and Matsushima (1992a, 1992b), which explored a similar idea of iterative removal of dominated strategies. We will use this idea also in the proofs of Theorems 3 and 6.

and

$$m_1 = a\,.$$

If $a \neq a^*$, then

$$u_1(x^+(m_1^*, m_{-1}), t_1^+(m_1^*, m_{-1}), m_1^*, m_{-1}) = v_1(x^+(m)) + 3\varepsilon$$
$$> v_1(x^+(m)) \geq u_1(x^+(m), t_1^+(m), m)\,,$$

which implies that agent 1 does not have incentive to announce $m_1 = a$. If $a = a^*$, then $m_1 = m_1^*$, and therefore, it follows from inequality (1) that for every $m_1' \in M_1 / \{a^*\}$,

$$u_1(x^+(m), t_1^+(m), m) = v_1(x^+(m))$$
$$> v_1(x^+(m)) + 3\varepsilon - c_1 = u_1(x^+(m_1', m_{-1}), t_1^+(m_1', m_{-1}), m_1', m_{-1})\,,$$

which implies that agent 1 has incentive to announce $m_1 = a^*$.

The above arguments imply that m^* is the unique Nash equilibrium in (G^+, u). Hence, we have proved that any alternative $a^* \in A$ is implemented in Nash equilibrium by G^+ with respect to all $u \in U^+(a^*, d)$.

Q.E.D.

The logical core of Theorem 2 is as follows. Since no agents other than agent 1 want to be the first deviant(s), they have incentive to announce only the same recommendation as agent 1's. Agent 1, however, can receive the monetary gain $3\varepsilon > 0$ by announcing differently from the other agents' recommendations. This interrupts any message profile other than m^* from being a Nash equilibrium. On the other hand, agent 1 has no incentive to deviate from m^*, because she can save the psychological cost c_1 for immorality, which is greater than the monetary gain 3ε. This is why m^* is the unique pure strategy Nash equilibrium in G^+.

Theorem 2 does not much depend on the assumption that the set of alternatives is describable and enforceable. We only need a^* and $\bar{a}$ to be describable and enforceable, which is *weaker* than the mechanism G^* in Section 3. Hence, from the viewpoint of contractual incompleteness, the mechanisms with small fines and multiple announcements may have advantage over the mechanism with no fine and single announcements.

We can check that m^* is the unique mixed strategy Nash equilibrium in (G^+, u) with a *refinement* device that agent 1 never announces any message m_1 that is *weakly dominated* by m_1^* in that for every $m_{-1} \in M_{-1}$,

$$u_1(x^+(m_1^*, m_{-1}), t_1^+(m_1^*, m_{-1}), m_1^*, m_{-1}) \geq u_1(x^+(m), t_1^+(m), m)\,,$$

and the strict inequality holds for some $m_{-1} \in M_{-1}$. In fact, every $m_1 \in M_1 / \{m_1^*\}$ is weakly dominated by m_1^*, because for every $m_{-1} \in M_{-1}$,

$$u_1(x^+(m_1^*, m_{-1}), t_1^+(m_1^*, m_{-1}), m_1^*, m_{-1}) \geq u_1(x^+(m), t_1^+(m), m)\,,$$

and the strict inequality holds for $m_{-1} = m_{-1}^*$. Hence, agent 1 will announce only m_1^* in this case. Since m^* is the unique mixed strategy Nash equilibrium where agent 1 chooses

m_1^*, we have proved that m^* is the unique mixed strategy Nash equilibrium with the above refinement device.

There, however, may exist unwanted mixed strategy Nash equilibria in the game (G^+, u), where with positive probability agent 1 announces messages that are weakly dominated by m_1^*. In the next section, we will show the possibility that any alternative is implementable even in terms of *mixed* strategy Nash equilibrium.

The result of this section holds even if no agent has moral preference. Suppose that for every $a \in A/\{a^*\}$,

$$3\varepsilon > u_1(a, m) - \max_{m_1' \neq a} u_1(a, m_1', m_{-1}),$$

and

$$u_1(a^*, m^*) - \max_{m_1' \neq a^*} u_1(a^*, m_1', m_{-1}^*) > u_1(a, m) - \max_{m_1' \neq a} u_1(a, m_1', m_{-1}).$$

Hence, agent 1's gain from the same announcement as the other agents' common immoral announcements is less than 3ε. It is also less than agent 1's gain from the moral announcement when the other agents make the honest announcements. Here, *we do not require agent 1 to prefer the moral announcement the most*. Modify t_1^+ in ways that agent 1 receives from the first deviants a value between $\dfrac{u_1(a^*, m^*) - \max_{m_1' \neq a^*} u_1(a^*, m_1', m_{-1}^*)}{3}$ and $\dfrac{u_1(a, m) - \max_{m_1' \neq a} u_1(a, m_1', m_{-1})}{3}$ instead of ε^+. In the same ways as Theorem 2, we can check that a^* is implemented in Nash equilibrium by this modified mechanism.

5. Iterative Dominance

This section shows that there exists a mechanism with small fines that can implement any alternative in *iterative dominance*, and therefore in *mixed* strategy Nash equilibrium, whenever agent 1 regards this alternative as being socially optimal and has moral preference in the original sense. Here, the other agents are not required to have moral preferences. This section assumes that $n = 3$,[13] and there exists a positive integer K such that $K_1 = K + 1$ and $K_2 = K_3 = K$, i.e.,

$$M_1 = A^{K+1} \text{ and } M_2 = M_3 = A^K .$$

Hence, the central planner requires agent 1 to make $K+1$ announcements and the other agents to make K announcements each. For every $h \in \{1,...,K\}$, let $m^h = (m_1^h, m_2^h, m_3^h)$.

Fix positive real numbers $\varepsilon > 0$ and $d > 0$ arbitrarily. Choose K to be large so that

(4) $$K\varepsilon > d .$$

Note that we can choose $\varepsilon^{++} \in (0, \varepsilon)$ to satisfy

(5) $$K\varepsilon^{++} + \rho > d ,$$

where we denote $\rho = \varepsilon - \varepsilon^{++} > 0$.

We specify a mechanism $G^{++} = G^{++}(K, \varepsilon, d) = (x^{++}, t^{++})$ as follows. Fix $m \in M$ arbitrarily. Let

$$x^{++}(m) = \frac{\sum_{h=1}^{K} z(m^h)}{K} .$$

For every $h \in \{1,...,K\}$, with probability $\frac{1}{K}$, the central planner will choose $z(m^h)$, where for every $a \in A$,

$$z(m^h) = a \text{ if there exist at least two agents } i \in \{1,2,3\} \text{ such that } m_i^h = a ,$$

and

$$z(m^h) = \overline{a} \text{ if } m_1^h \neq m_2^h \neq m_3^h \neq m_1^h .$$

Note that $x^{++}(m)$ does not depend on m_1^{K+1}. For every $i \in \{2,3\}$, let

$$t_i^{++}(m) = -\varepsilon^{++} - \frac{q_i(m)}{K}\rho \text{ if there exists } h \in \{1,...,K\} \text{ such that}$$
$$m_1^{h'} = m_2^{h'} = m_3^{h'} = m_1^{K+1} \text{ for all } h' \in \{1,...,h-1\} \text{ and } m_i^h \neq m_1^{K+1},$$

and

$$t_i^{++}(m) = -\frac{q_i(m)}{K}\rho \text{ otherwise.}$$

---

[13] This implies that there may exist four or more agents but only three agents are required to participate in this decision procedure.

Each agent $i \in \{2,3\}$ is fined the monetary amount ε^{++}, in addition to $\frac{q_i(m)}{K}\rho$, if and only if she is the first agent to announce differently from agent 1's $(K+1)-th$ announcement. Let

$$t_1^{++}(m) = 0 \quad \text{for all} \quad m \in M .$$

Hence, agent 1 is never fined or rewarded.

For every $a^* \in A$ and every $d > 0$, we define $U^{++}(a^*, d)$ as the set of utility function profiles u satisfying the following two properties.

(vii) For every $i \in \{2,3\}$, every $a \in A$, and every $a' \in A/\{a\}$,

$$\left|v_i(a) - v_i(a')\right| \le d .$$

(viii) For every $m \in M$,

$$r_1(m) = \frac{w_1(m)}{K+1},$$

where $w_1(m) = \left|\{h \in \{1,...,K_1\} \mid m_1^h \neq a^*\}\right|$.

Property (vii) implies that d is the upper bound of the utility differences for all agents except for agent 1. Property (viii) implies that agent 1 always prefers the announcement of the socially optimal alternative to any other announcement. Note that for every $a \in A$, every $t_i \in R$, and every $m \in M$,

$$u_1(a, t_1, m) = v_1(a) + t_1 - \frac{w_1(m)}{K+1}c_1 .$$

We do not require each agent other than agent 1 to have moral preference.

Theorem 3: *Suppose that inequality (4) holds. Then, any alternative $a^* \in A$ is implemented in iterative dominance by $G^{++}(K, \varepsilon, d)$ with respect to all $u \in U^{++}(a^*, d)$.*

Proof: Note that

$$x^{++}(m^*) = a^* ,$$

and

$$t_i^{++}(m^*) = 0 \quad \text{for all} \quad i \in N .$$

Agent 1 has incentive to announce $m_1^{K+1} = a^*$, because both $x^{++}(m)$ and $t_1^{++}(m)$ are independent of m_1^{K+1} and because of property (viii).

Fix $h \in \{1,...,K\}$ and $m \in M$ arbitrarily, where

$$m_1^{K+1} = a^* ,$$

and

$$m_i^{h'} = a^* \quad \text{for all} \quad i \in N \quad \text{and all} \quad h' \in \{1,...,h-1\} .$$

First, consider any agent $i \in \{2,3\}$. Suppose $m_i^h \neq a^*$. Let $m_i' \in M_i$ be the message for agent i defined by

$$m_i'^h = a^*,$$

and

$$m_i'^{h'} = m_i^{h'} \text{ for all } h' \in \{1,...,K\}/\{h\}.$$

If $m_j^h = a^*$ for all $j \in N/\{i\}$, then, it follows that $x^{++}(m)$ is independent of m_i^h and $t_i^{++}(m_i', m_{-i}) - t_i^{++}(m) \geq \frac{\rho}{K} > 0$, which implies that agent i has incentive to announce m_i' instead of m_i. If $m_j^h \neq a^*$ for some $j \neq i$, then it follows that $t_i^{++}(m_i', m_{-i}) - t_i^{++}(m) = \varepsilon^{++} + \frac{\rho}{K}$, which, together with property (vii) and inequality (5), implies that agent i has incentive to announce m_i' instead of m_i, where

$$u_i(x^+(m), t_i^+(m), m) - u_i(x^+(m_i', m_{-i}), t_i^+(m_i', m_{-i}), m_i', m_{-i})$$
$$\leq -\varepsilon^{++} - \frac{\rho}{K} + \frac{1}{K}\{v_i(\overline{\alpha}) - v_i(z(m_i'^h, m_{-i}^h))\} < -\varepsilon^{++} - \frac{\rho}{K} + \frac{d}{K} < 0.$$

Next, suppose that

$$m_i^h = a^* \text{ for each } i \in \{2,3\},$$

and

$$m_1^h \neq a^*.$$

Let $m_1' \in M_1$ be the message for agent 1 defined by

$$m_1'^h = a^*,$$

and

$$m_1'^{h'} = m_1^{h'} \text{ for all } h' \in \{1,...,K\}/\{h\}.$$

Note that $x^{++}(m)$ is independent of m_1^h and $t_1^{++}(m_1', m_{-1}) = t_1^{++}(m) = 0$, which, together with property (viii), implies that agent 1 has incentive to announce m_1' instead of m_1.

The above arguments imply that m^* is the unique iteratively undominated message profile in (G^{++}, u). Hence, we have proved that any alternative $a^* \in A$ is implemented in iterative dominance by G^{++} with respect to all $u \in U^{++}(a^*, d)$.

Q.E.D.

The logical core of the proof of Theorem 3 is as follows. Since m_1^{K+1} does not influence $x^{++}(m)$ and $t_1^{++}(m)$, it follows from moral preference that agent 1 always prefers to announce $m_1^{K+1} = a^*$. In the same way as the idea of iterative removal of undominated strategies originated in Abreu and Matsushima (1992b), every agent dislikes to announce differently from m_1^{K+1}.

In contrast with Theorems 1 and 2, we do not need any restriction on the sizes of the

costs $c_i > 0$ such as properties (ii) and (iii) and inequalities (1) and (2). This implies that the same mechanism G^{++} works for any specification of $(c_i)_{i \in N}$.

Theorem 3 does not much depend on the assumption that the set of alternatives is describable and enforceable. In the same way as the mechanism G^{+} in Section 4, we can check that *we only need* a^* *and* $\overline{a}$ *to be describable and enforceable*.

6. Implementation of Social Choice Functions

We will show that each of the mechanisms G^*, G^+, and G^{++} can implement a wide variety of social choice functions. We do not require the set of states to be describable. The mechanisms do not depend on the detail of a particular model specification such as the set of states and the social choice function. Let Ω denote the set of states. A *social choice function* $f:\Omega \to A$ is defined as a mapping from states to alternatives. Let F denote the set of social choice functions. A *state-contingent* utility function profile is given by $\mu = (u^\omega)_{\omega\in\Omega}$, where $u^\omega = (u_i^\omega)_{i\in N}$ and $u_i^\omega : A\times R\times M \to R$. A social choice function f is said to *be implemented in Nash equilibrium (iterative dominance) by a mechanism* G *with respect to a state-contingent utility function profile* μ if for every $\omega\in\Omega$, $f(\omega)$ is implemented in Nash equilibrium (iterative dominance, respectively) by G with respect to u^ω.

Subsection 6.1 will show that a wide variety of social choice functions are implementable even if the set of states are indescribable. Subsection 6.2 will show that whenever a single mechanism can implement multiple social choice functions, agents' preferences must depend on the social choice function. Moral preference is regarded as a special case that their preferences depend on the social choice function.

6.1. Indescribability

The mechanisms G^*, G^+, and G^{++} do not depend on how the set of states and the social choice function are to be specified. These mechanisms each, nevertheless, can implement a wide variety of social choice functions. The following theorem is straightforward from Theorems 1, 2, and 3.

Theorem 4: *Suppose that* $n\geq 5$, n *is odd, and* $M_i = A$ *for all* $i\in N$. *Then, a social choice function* f *is implemented in Nash equilibrium by* G^* *with respect to* μ *if*

$$u^\omega \in U^*(f(\omega)) \text{ for all } \omega\in\Omega.$$

Suppose that $n=4$, $K_1=1$, *and there exists a positive integer* K *such that* $M_2=M_3=M_4=A^K$. *Then, a social choice function* f *is implemented in Nash equilibrium by* G^+ *with respect to* μ *if*

$$u^\omega \in U^+(f(\omega),d) \text{ for all } \omega\in\Omega.$$

Suppose that $n=3$, *and there exists a positive integer* K *such that* $M_1=A^{K+1}$ *and* $M_2=M_3=A^K$. *Then, a social choice function* f *is implemented in iterative dominance by* G^{++} *with respect to* μ *if*

$$u^\omega \in U^{++}(f(\omega),d) \text{ for all } \omega\in\Omega.$$

Theorem 4 implies that even if the set of states is indescribable, a wide variety of

social choice functions are implementable. Theorem 4 does not much depend on the describability and enforceability of alternatives. In order for G^* to work, we only need that at every state $\omega \in \Omega$, $f(\omega)$, $\overline{a}$, and agent 1's most favorite alternative $a^1 = a^1(\omega)$, which may depend on the state, are describable and enforceable. In order for G^+ and G^{++} to work, we only need that at every state $\omega \in \Omega$, $f(\omega)$ and $\overline{a}$ are describable and enforceable.

6.2. Dependence of Preferences on Social Choice Function

This subsection will allow the set of states and the social choice function to be describable, and therefore allow a mechanism to depend on the set of states Ω and the social choice function f. A subset of social choice functions is denoted by $\widetilde{F} \subset F$. Fix a set of message profiles M arbitrarily. Let $(\mu^f)_{f \in \widetilde{F}}$ denote a collection of state-contingent utility function profiles, where $\mu^f = (u^{\omega,f})_{\omega \in \Omega}$ and $u^{\omega,f} = (u_i^{\omega,f})_{i \in N}$. *A subset of social choice functions $\widetilde{F} \subset F$ is said to be implemented in Nash equilibrium (iterative dominance) by a mechanism G with respect to $(\mu^f)_{f \in \widetilde{F}}$* if every $f \in \widetilde{F}$ is implemented in Nash equilibrium (iterative dominance, respectively) by G with respect to μ^f.

Note that there exist no subset of social choice functions $\widetilde{F} \subset F$ that is not a singleton, no collection of state-contingent utility function profiles $(\mu^f)_{f \in \widetilde{F}}$ that is constant with respect to f, and no mechanism G that implements $\widetilde{F}$ in Nash equilibrium with respect to $(\mu^f)_{f \in \widetilde{F}}$. This implies that in order for a single mechanism to implement any subset of social choice function that is not a singleton, the state-contingent utility function profile must depend on the social choice function. This result hold true when we replace the Nash equilibrium concept with any equilibrium concept with complete information.

Let $M_i = A^{K_i}$ for all $i \in N$, where K_i is a positive integer. Let $m_i^{\omega,f} = (m_i^{\omega,f,h})_{h=1}^{K_i} \in M_i$ denote the *moral* message profile defined by

$$m_i^{\omega,f,h} = f(\omega) \text{ for all } h \in \{1,...,K_i\}.$$

The following theorem is straightforward from Theorem 4.

Theorem 5: *Suppose that $n \geq 5$, n is odd, and $M_i = A$ for all $i \in N$. Then, $\widetilde{F} \subset F$ is implemented in Nash equilibrium by G^* with respect to $(\mu^f)_{f \in \widetilde{F}}$ if*

$$u^{\omega,f} \in U^*(f(\omega)) \text{ for all } f \in \widetilde{F} \text{ and all } \omega \in \Omega.$$

Suppose that $n = 4$, $K_1 = 1$, and there exists a positive integer K such that

$M_2 = M_3 = M_4 = A^K$. *Then,* $\widetilde{F} \subset F$ *is implemented in Nash equilibrium by* G^+ *with respect to* $(\mu^f)_{f \in \widetilde{F}}$ *if*

$$u^{\omega, f} \in U^+(f(\omega), d) \text{ for all } f \in \widetilde{F} \text{ and } \omega \in \Omega.$$

Suppose that $n = 3$, *and there exists a positive integer* K *such that* $M_1 = A^{K+1}$ *and* $M_2 = M_3 = A^K$. *Then,* $\widetilde{F} \subset F$ *is implemented in iterative dominance by* G^{++} *with respect to* $(\mu^f)_{f \in \widetilde{F}}$ *if*

$$u^{\omega} \in U^{++}(f(\omega), d) \text{ for all } f \in \widetilde{F} \text{ and all } \omega \in \Omega.$$

Note that moral preferences are regarded as a special case of the dependence that agents' preferences depend on the social choice function. Theorem 5 implies that when agents' preferences depend on the social choice function in this way, a single mechanism can implement a wide variety of social choice functions.

7. Incomplete Information

This section investigates the *incomplete information* environments. We assume $n \geq 2$. Each agent receives her *private signal* denoted by ω_i. Let Ω_i denote the set of private signals for agent $i \in N$. Let $\Omega = \prod_{i \in N} \Omega_i$. Let $p : \Psi \to [0,1]$ denote a probability measure on (Ω, Ψ) where Ψ is a $\sigma-$field. Let P denote the set of probability measures. A *message rule for each agent* $i \in N$ is defined as a function $\eta_i : \Omega_i \to M_i$. Let Ξ_i denote the set of all message rules for agent i. We denote by $\eta = (\eta_i)_{i \in N}$ a message rule profile. Let $\Xi \equiv \prod_{i \in N} \Xi_i$, $\eta(\omega) = (\eta_i(\omega_i))_{i \in N}$, and $\eta_{-i}(\omega_{-i}) = (\eta_j(\omega_j))_{j \in N/\{i\}}$.

Let $\Xi_i^{(0,G,p,\mu)} = \Xi_i$ and $\Xi^{(0,G,p,\mu)} = \prod_{i \in N} \Xi_i^{(0,G,p,\mu)}$. Recursively, for every $r = 1,2,\ldots$, let $\Xi_i^{(r,G,p,\mu)}$ denote the set of message rules $\eta_i \in \Xi_i^{(r-1,G,p,\mu)}$ for each agent i such that there exist no $m_i \in M_i$ and no $\omega_i \in \Omega_i$ satisfying that for every $\eta_{-i} \in \Xi_{-i}^{(r-1,G,p,\mu)}$,

$$E[u_i^{\omega}(g(\eta(\omega)), x_i(\eta(\omega)), \eta(\omega)) \mid p, \omega_i]$$
$$< E[u_i^{\omega}(g(m_i, \eta_{-i}(\omega_{-i})), x_i(m_i, \eta_{-i}(\omega_{-i})), m_i, \eta_{-i}(\omega_{-i})) \mid p, \omega_i],$$

where $\Xi_{-i}^{(r-1,G,p,\mu)} = \prod_{j \in N/\{i\}} \Xi_j^{(r-1,G,p,\mu)}$, and $E[\cdot \mid p, \omega_i]$ implies the expected value conditional on agent $i's$ private signal ω_i with respect to the probability measure p. Let $\Xi^{(r,G,p,\mu)} = \prod_{i \in N} \Xi_i^{(r,G,p,\mu)}$ and $\Xi^{(\infty,G,p,\mu)} = \bigcap_{r=0}^{\infty} \Xi^{(r,G,p,\mu)}$. A message rule profile $\eta \in \Xi$ is said to be *iteratively undominated in the Bayesian game defined by* (G, p, μ) if $\eta \in \Xi^{(\infty,G,p,\mu)}$. A social choice function $f \in F$ is said to *be implemented in iterative dominance by a Bayesian game* (G, p, μ) if there exists the unique iteratively undominated message rule profile η in (G, p, μ), and this message rule profile satisfies that for every $\omega \in \Omega$,

$$g(\eta(\omega)) = f(\omega), \text{ and } x_i(\eta(\omega)) = 0 \text{ for all } i \in N.$$

Note that if η is the unique iteratively undominated message rule profile, then it is the unique Bayesian Nash equilibrium.

Let A_i denote the set of possible characteristics of the socially optimal alternative that agent i may know. We assume that a profile of characteristics $(a_i)_{i \in N} \in \prod_{i \in N} A_i$ uniquely determines an alternative. We denote $A = \prod_{i \in N} A_i$ and $a = (a_i)_{i \in N}$. Hence, a social choice function f is decomposable in the sense that there exists $(f_i)_{i \in N}$ such that

$$f_i : \Omega_i \to A_i \text{ for all } i \in N,$$

and

$$f(\omega) = (f_i(\omega_i))_{i \in N} \text{ for all } \omega \in \Omega.$$

We assume that there exists a positive integer $K > 0$ such that

$$M_i = A_i^K \text{ for all } i \in N.$$

Each agent $i \in N$ makes K announcements about what the characteristic of the socially optimal alternative that she knows is. Let $M_i = M_i^1 \times \cdots \times M_i^K$ where $M_i^k = A_i$. Let $\hat{\eta}$ denote the *moral* message rule profile such that for every $i \in N$,

$$\hat{\eta}_i^k(\omega_i) = f_i(\omega_i) \text{ for all } k \in \{1,...,K\} \text{ and all } \omega_i \in \Omega_i.$$

Fix a positive real number $d > 0$ and a positive integer $\hat{K} \in \{1,...,K-1\}$ arbitrarily, and choose $\varepsilon > 0$ to be close to zero so that

$$\frac{\hat{K}}{K} c_i > \varepsilon \text{ for all } i \in N, \tag{6}$$

and

$$(K - \hat{K})\varepsilon > d. \tag{7}$$

We specify a mechanism $\hat{G} = \hat{G}(K, \hat{K}, \varepsilon, d) = (\hat{x}, \hat{t})$ as follows. For every $m \in M$,

$$\hat{x}(m) = \frac{\sum_{h=\hat{K}+1}^{K} f(m^h)}{K - \hat{K}},$$

where we regard $f(\omega)$ as the simple lottery such that $f(\omega)(a) = 1$ if $f(\omega) = a$. For every $k \in \{\hat{K}+1,...,K\}$, with probability $\frac{1}{K-\hat{K}}$, the central planner will choose $f(m^k)$. Note that $\hat{x}(m)$ does not depend on agents' first $\hat{K}$ announcements $(m^1,...,m^{\hat{K}})$. This independence will play an important role in establishing a reference point to check whether each agent made the moral announcements or not in the incomplete information environments.

For every $i \in N$ and every $m \in M$,

$\hat{t}_i(m) = -\varepsilon$ if there exist $k \in \{2,...,K\}$ such that $m_i^k \neq m_i^1$, and $m^h = m^1$ for all $h \in \{1,...,k-1\}$.

and

$\hat{t}_i(m) = 0$ if there exists no such k.

Each agent $i \in N$ is fined if and only if she is the first agent whose announcement is inconsistent with her first announcement.

When the agents announce the moral message profile $\hat{\eta}(\omega)$, the central planner will choose $f(\omega)$ and no agents are fined, i.e., for every $\omega \in \Omega$,

$$\hat{x}(\hat{\eta}(\omega)) = f(\omega), \text{ and } \hat{t}_i(\hat{\eta}(\omega)) = 0 \text{ for all } i \in N.$$

We shall confine our attentions to state-dependent utility function profiles $\mu = (u^\omega)_{\omega \in \Omega}$ such that

$$\max_{(a,a',\omega,i) \in A^2 \times \Omega \times N} \left| v_i(a,\omega) - v_i(a',\omega) \right| \leq d,$$

and every agent i has *moral* preference in the sense that she has a positive psychological cost $\frac{c_i}{K} > 0$ for announcing any characteristic other than $f_i(\omega_i)$, i.e., for every $\omega \in \Omega$, every $(a, t_i) \in A \times [-\varepsilon, 0]$, and every $m \in M$,

$$u_i^{\omega}(a, t_i, m) = v_i(a, \omega) + t_i - \frac{q_i(m_i, f_i(\omega_i))}{K} c_i,$$

where $q_i(m_i, a_i) \in \{0, ..., K\}$ denotes the number of $k \in \{1, ..., K\}$ satisfying $m_i^k \neq a_i$. Let $W(d)$ denote the set of all such state-dependent utility function profiles.

Note that the mechanism $\hat{G}$ is not tailored to a particular model specification. In fact, it is independent of the *probability structure*, as well as of the preference structure, the set of states, and the social choice function, as mentioned before.

Theorem 6: *Suppose that inequalities (6) and (7) hold. Then, For every* $p \in P$ *and every* $\mu \in W(d)$, *any social choice function* $f \in F$ *is implemented in iterative dominance by* $(\hat{G}, p, \mu)$ *if for every* $i \in N$, *every* $\omega_i \in \Omega_i$, *and every* $\omega_i' \in \Omega_i$,

$$E[v_i(f(\omega), \omega) \mid p, \omega_i] \geq E[v_i(f(\omega_i', \omega_{-i}), \omega) \mid p, \omega_i] - \frac{(K - \hat{K})}{K} c_i. \tag{8}$$

Proof: Fix $\eta \in \Xi$ and $i \in N$ arbitrarily. Fix $\omega \in \Omega$ arbitrarily. Suppose that

$$\eta_j^k(\omega_j) \neq \eta_j^{k-1}(\omega_j) \text{ for some } j \in N / \{i\} \text{ and some } k \in \{2, ..., \hat{K}\}.$$

Then, agent i is never fined whenever she announces

$$m_i^k = f_i(\omega_i) \text{ for all } k \in \{1, ..., \hat{K}\}.$$

Next, suppose that

$$\eta_j^k(\omega_j) = \eta_j^{k-1}(\omega_j) \text{ for all } k \in \{2, ..., \hat{K}\} \text{ and all } j \in N.$$

If $\eta_i^k(\omega_i) \neq f_i(\omega_i)$ for all $k \in \{1, ..., \hat{K}\}$, then, by announcing $m_i^k = f_i(\omega_i)$ for all $k \in \{1, ..., \hat{K}\}$ instead, agent i can save the amount $\frac{\hat{K}}{K} c$ of her psychological cost. This amount is greater than the monetary fine ε, because of inequality (6). If $\eta_i^k(\omega_i) \neq \eta_i^{k-1}(\omega_i)$ for some $k \in \{2, ..., \hat{K}\}$, then the central planner will fine agent i. Since she has moral preference and her first $\hat{K}$ announcements do not influence the central planner's alternative choice, the above arguments imply that agent i is willing to replace the first $\hat{K}$ announcements $(\eta_i^k(\omega_i))_{k=1}^{\hat{K}}$ with the moral announcements $(\hat{\eta}_i^k(\omega_i))_{k=1}^{\hat{K}}$. Hence, we have proved that for every $i \in N$, if η_i is iteratively undominated, then it must hold that

$$\eta_i^k = \hat{\eta}_i^k \text{ for all } k \in \{1, ..., \hat{K}\}.$$

Fix $\bar{k} \in \{\hat{K}+1,...,K\}$ arbitrarily. Suppose that

$$\eta_j^k = \hat{\eta}_j^k \quad \text{for all} \quad j \in N \quad \text{and all} \quad k \in \{1,...,\bar{k}-1\}.$$

Fix $\omega_i \in \Omega_i$ arbitrarily, and suppose that $\eta_i^{\bar{k}}(\omega_i) \neq f_i(\omega_i)$. Let $m_i \in M_i$ denote the message for agent i defined by

$$m_i^k = \hat{\eta}_i^k(\omega_i) \quad \text{for all} \quad k \in \{1,...,\bar{k}\},$$

and

$$m_i^k = \eta_i^k(\omega_i) \quad \text{for all} \quad k \in \{\bar{k}+1,...,K\}.$$

Suppose that $\eta_j^{\bar{k}}(\omega_j) \neq f_i(\omega_j)$ for some $j \in N/\{i\}$. Then,

$$\hat{t}_i(\eta(\omega)) = -\varepsilon \quad \text{and} \quad \hat{t}_i(m_i, \eta_{-i}(\omega_{-i})) = 0.$$

Inequality (7) implies that the expected value of utility difference for alternative between the messages $\eta_i(\omega_i)$ and m_i is less than ε. Hence, agent i strictly prefers announcing m_i instead of $\eta_i(\omega_i)$.

Next, suppose that $\eta_j^{\bar{k}}(\omega_j) = f_j(\omega_j)$ for all $j \in N/\{i\}$. Then,

$$\hat{t}_i(\eta(\omega)) = -\varepsilon \quad \text{and} \quad \hat{t}_i(m_i, \eta_{-i}(\omega_{-i})) \geq -\varepsilon.$$

Inequality (8), together with moral preference, implies that agent i has strict incentive to make the moral announcement when the other agents make the moral announcements. Hence, agent i strictly prefers to announce m_i instead of $\eta_i(\omega_i)$.

From the above arguments, we have proved that if η is an iteratively undominated message rule profile, then $\eta = \hat{\eta}$ must hold. Since the set of iteratively undominated message rule profiles $\Xi^{(\infty,G,p,\mu)}$ is nonempty, we have completed the proof of Theorem 6.

Q.E.D.

We can choose K, $\hat{K} \in \{1,...,K-1\}$, and $\varepsilon > 0$ such that inequalities (6) and (7) hold and $\frac{\hat{K}}{K}$ is as close to zero as possible. Hence, it follows from Theorem 6 that for every $p \in P$ and every $\mu \in W(d)$, every social choice function $f \in F$ is implemented in iterative dominance by $\hat{G}$ if for every $i \in N$, every $\omega_i \in \Omega_i$, and every $\omega_i' \in \Omega_i$,

$$E[v_i(f(\omega),\omega) \mid p, \omega_i] \geq E[v_i(f(\omega_i', \omega_{-i}),\omega) \mid p, \omega_i] - c_i.$$

We do not need the set of alternatives to be describable. When agents could announce a message profile $m \in M$, the alternative given by m^h for each $h \in \{1,...,K\}$ must be describable, and therefore, any alternative in the support of $\hat{x}(m)$ is describable, because whenever $\hat{x}(m)(a) > 0$ then $a = m^h$ for some $h \in \{\hat{K}+1,...,K\}$. Hence, we only need that at every state $\omega \in \Omega$, the value of the social choice function $f(\omega)$ is describable, provided that any describable alternative is enforceable. If some describable alternatives are not enforceable, then we may have to modify the incentive compatibility condition by

replacing it with any describable alternative such as the status quo alternative $\bar{a}$.

References

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

Abreu, D. and H. Matsushima (1992b): "Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information," mimeo.

Alger, I. and C. A. Ma (2003): "Moral Hazard, Insurance, and Some Collusion," *Journal of Economic Behavior and Organization* 50, 225-247.

Basu, K., P. Pattanaik, and K. Suzumura (1995): *Choice, Welfare, and Development*, Oxford: Clarendon Press.

Deneckere, R. and S. Severinov (2001): "Mechanism Design and Communication Costs," mimeo.

Duggan, J. (1997): "Virtual Bayesian Implementation," *Econometrica* 65, 1175-1199.

Dworkin, R. (1981): "What is equality ? Part 1: Equality of Welfare, Part 2: Equality of Resources," *Philosophy and Public Affairs* 10, 185-246, 283-345.

Eliaz, K. (2002): "Fault Tolerant Implementation," *Review of Economic Studies* 69, 589-610.

Erard, B. and J. Feinstein (1994): "Honesty and Evasion in the Tax Compliance Game," *RAND Journal of Economics* 25, 1-19.

Fehr, E. and K. Schmidt (2003): "Theories of Fairness and Reciprocity: Evidence and Economic Applications," in *Advances in Economics and Econometrics: Eighth World Congress*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky.

Glazer, J. and A. Rubinstein (1998): "Motives and Implementation: On the Design of Mechanisms to Elicit Opinions," *Journal of Economic Theory* 79, 157-173.

Hart, O. (1995): *Firms, Contracts, and Financial Structure*, Oxford: Oxford University Press.

Jackson, M. (1991): "Bayesian Implementation," *Econometrica* 59, 461-477.

Maskin, E. and T. Sjostrom (2002): "Implementation Theory," in *Handbook of Social Choice and Welfare Volume 1*, ed. by K. Arrow, A. Sen, and K. Suzumura. Elsevier.

Maskin, E. and J. Tirole (1999): "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies* 66, 83-114.

Matsushima, H. (1993): "Bayesian Monotonicity with Side Payments," *Journal of Economic Theory* 59, 107-121.

Milgrom, P. and J. Roberts (1992): *Economics, Organizations and Management*, Englewood Cliffs, NJ: Prentice Hall.

Moore, J. (1992): "Implementation in Environments with Complete Information," in *Advances in Economic Theory: Sixth World Congress*, ed. by J.-J. Laffont. Cambridge University Press.

Palfrey, T. (1992): "Implementation in Bayesian Equilibrium: the Multiple Equilibrium Problem in Mechanism Design," in *Advances in Economic Theory: Sixth World Congress*, ed. by J.-J. Laffont, Cambridge University Press.

Rawls, J. (1971): *A Theory of Justice*, Cambridge: Harvard: Harvard University Press.

Sen, A. (1982): *Choice, Welfare and Measurement*, Oxford: Blackwell.

Sen, A. (1985): *Commodities and Capabilities*, Amsterdam: North-Holland.
Sen, A. (1999): "The Possibility of Social Choice," *American Economic Review* 89, 349-378.
Serrano, R. and R. Vohra (2000): "Type Diversity and Virtual Bayesian Implementation," Working Paper No. 00-16, Department of Economics, Brown University.
Serrano, R. and R. Vohra (2001): "Some Limitations of Virtual Bayesian Implementation," *Econometrica* 69, 785-792.
Suzumura, K. (2002): "Introduction," in *Handbook of Social choice and Welfare, Volume 1*, ed. by K. Arrow, A. Sen, and K. Suzumura, Elsevier Science.
Tirole, J. (1999): "Incomplete Contracts: Where do we stand?," *Econometrica* 67, 741-781.