CARF Working Paper

CARF-F-140

Implementation and Social Influence

Hitoshi Matsushima
University of Tokyo

October,2008

CARF is presently supported by AIG, Bank of Tokyo-Mitsubishi UFJ, Ltd., Citigroup, Dai-ichi Mutual Life Insurance Company, Meiji Yasuda Life Insurance Company, Mizuho Financial Group, Inc., Nippon Life Insurance Company, Nomura Holdings, Inc. and Sumitomo Mitsui Banking Corporation (in alphabetical order). This financial support enables us to issue CARF Working Papers.

Implementation and Social Influence[+]

Hitoshi Matsushima[*]

Faculty of Economics, University of Tokyo

December 8, 2008

[+] The research for this paper was supported by a Grant-In-Aid for Scientific Research (KAKENHI 18330035) from the Japan Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of the Japanese government.

[*] Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Tel: +81-3-5841-5506. Fax:+81-3-5841-5521. E-Mail: hitoshi at mark e.u-tokyo.ac.jp

Abstract

This paper incorporates social psychology into implementation theory. Real individuals care not only about their material benefits but also about their social influence in terms of obedience and conformity. Using a continuous time horizon, we demonstrate a method of manipulating the decision-making process, according to which, an uninformed principal utilizes her/his power of social influence to incentivize multiple informed agents to make honest announcements. Following this method, we show that with incentive compatibility, the principal can implement any alternative as she/he wishes as the unique Nash equilibrium outcome, even if her/his power is limited and no contractual devices are available.

Keywords: Implementation, Uniqueness, Obedience, Conformity, Small Guilt-Aversion, Permissive Results.

Journal of Economic Literature Classification Numbers: C72, D78, D81, D86

1. Introduction

This paper investigates a decision problem in which a principal schemes to select the alternative that is desirable in light of her/his wishes, although she/he is not aware of what this alternative might be. Besides the principal, there exist multiple agents who possess their private signals about this potential alternative. The principal therefore requires these agents to make announcements to her/him about these private signals. In order to put the agents' announcements to good use in choosing the desired alternative, the principal has to come up with various ways to incentivize each agent to reveal any information that is honest and useful in light of the principal's wishes. In this case, it is insufficient for their honest announcements to satisfy incentive compatibility, since there may also exist self-enforcing, but dishonest, announcements that prevent the principal from arriving at her/his desirable alternative. Hence, in addition to incentive compatibility, the principal has to utilize additional incentive devices that function in eliminating unwanted equilibria, that is, the principal needs to obtain their honest announcements as the *unique* Nash equilibrium or similar to this.

The issue of uniqueness has been studied intensively in the standard theory of implementation; it was generally assumed that any agent is motivated by her/his material benefit and is purely rational; she/he cares only about her/his intrinsic utility that is directly derived from the alternative choice, and enjoys full autonomy in making her/his announcements. Following these assumptions, the authors in literature pertaining to this field have generally confined their attention to inventing new concepts of binding *contractual* devices such as the modulo mechanisms (Maskin [1977/1990], Matsushima [1988], Abreu and Sen [1990]) and the Abreu-Matsushima mechanisms (Abreu and

Matsushima [1992a, 1992b, 1994]), that implement, at least in the virtual sense, any value of the fixed social choice function as the unique Nash equilibrium outcome or similar in compensation for artificial tailoring[1].

In contrast to the standard theory, any real person does care about not only her/his material benefit but also about any factor of *social influence*; she/he often feels guilty about disobeying an authority figure's wishes. This feeling of guilt is especially strengthened when she/he expects all members of her/his reference group to obey these wishes. In this respect, several experimental studies in social psychology, such as the Eichmann test by Milgram (1974), the prison experiment by Zimbardo et al. (1977), and the hospital experiment by Hofling (1966), have commonly reported that the subjects in laboratories and fields tended to be extremely obedient in the presence of authority figures.[2] There also exist experimental studies such as Ash (1955), that report that the subjects tended to seek conformity to their reference groups' modes of behavior.[3] From these rich stores of knowledge in social psychology, it is natural to infer that the aforementioned principal is thinking pragmatically of utilizing social influence so as to change the agents' announcements.[4]

On the basis of the above arguments, this paper demonstrates the principal's method of implementing the desirable alternative by making full use of her/his limited power of social influence. Using the continuous time horizon, the principal will design

[1] For the surveys on the standard theory of implementation, see Moore (1992), Palfrey (1992), Osborne and Rubinstein (1994, Chapter 10), and Maskin and Sjöström (2002).

[2] These experiments assumed that the authority figures' wishes are not prosocial; the subjects may be extremely obedient even if the authority figures disturb social order.

[3] For issues on social influence in general, see also Cialdini (2001).

[4] Attempts to incorporate social psychology into economics are not new but are an area of increasing interest. See Akerlof and Dickens (1982), Geanakoplos, Pearce, and Stacchetti (1989), Bernheim (1994), Gneezy (2005), Charness and Dufwenberg (2006), Bébabou (2007) and so on.

the decision function that makes the agents' honest announcements incentive compatible, while also manipulating the announcement procedure in the following manner.

(i) The agents make their initial announcements.

(ii) Any agent is permitted to *change* her/his announcement at any time, and even many times, whenever she/he wants.

(iii) This procedure is *randomly* terminated at a constant hazard rate. According to the specified decision function, the principal selects the alternative that corresponds to their final announcements that are effective at the terminal time.

(iv) During this procedure, each agent is prohibited from monitoring the other agents' announcements.

Apart from the decision function, the principal does not use any other contractual device that is tailored to the details of the model specification. This implies that the agents' announcements, except for the ones that are effective at the terminal time, do not need to be verifiable to the public.

The result is quite permissive from the principal's point of view; with the assumption of complete information, the agents' announcing honestly at all times is the unique Nash equilibrium. We can replace the Nash equilibrium with other less-demanding solution concepts and we may extend our model to the incomplete information case.

Our permissive results depend on the psychological assumption of obedience and conformity that, at any time, each agent feels guilty about her/his dishonest announcement if she/he expects that the other agents have never announced dishonestly.

It is of particular importance to point out that our permissive results are almost *irrelevant* to the degree to which each agent feels guilty; even if the principal's power of social influence is too limited to make dishonest agents feel very guilty, she/he can considerably control their announcements in her/his own way by manipulating the procedure in the above manner. Each agent can lessen the psychological cost just a little by waiting for someone else to announce dishonestly before she/he does. This tiny cost reduction, along with incentive compatibility and random termination, is sufficient to trigger a tail-chasing competition among the agents, eliminating all their dishonest announcements at any time in due order.

The earlier works by Matsushima (2008a, 2008b) took into account behavioral aspects of agents in the theory of implementation[5] and showed that the presence of the small psychological cost of lying simplifies the method of designing the Abreu-Matsushima mechanisms. These works, however, treated the subjects' behavioral motives in a very naïve way. For instance, these works did not take account of any aspect of conformity, in that each agent's psychological cost of lying decreases once anyone else has lied. This naïveté is the central reason why the principal in these works still needed a contractual device that fines the first liars, which is the heart of the basic concept of the Abreu-Matsushima mechanism, in order to trigger their tail-chasing competition. In contrast, the principal in this paper can instigate the agents to their tail-chasing competition by resorting to their feelings of guilt, without utilizing such à la Abreu-Matsushima contractual devices.

Ash (1955) found in his famous conformity experiment that any subject feels less guilty if the other members of her/his peer group are not unanimous in conforming to

[5] There are a few other works in the study of implementation that took into account behavioral aspects, such as Eliaz (2002) and Glazer and Rubinstein (1998).

the collective norms. This finding in social psychology is consistent with our conformity assumption.

We should make a further comment on the differences between the present paper and the standard theory of implementation as follows. In the standard theory, the public can correctly infer the principal's wishes from the mechanism, or the procedure; the agents who are motivated only by material benefits surely play the associated unique Nash equilibrium, whose outcome is set equal to the value of the social choice function, that is, her/his desirable alternative. The public can understand her/his wishes just by calculating this unique Nash equilibrium. On the other hand, the public in the present paper cannot infer the principal's wishes from the procedure alone. The procedure is kept independent of the principal's wishes, and the agents' behaviors are influenced by the principal's wishes. Hence, the public cannot calculate their equilibrium behavior as long as the public is not informed of her/his wishes in advance.

The above comment may draw the following pessimistic view: the principal generally prefers to utilize her/his power of influence rather than contractual devices, when she/he wishes to undertake any action that cannot necessarily win the approval of the public. By keeping her/his wishes concealed from the public, but drawing out the agents' tiny feeling of guilt, even dogmatic principals can make their antisocial decisions as they wish.

The organization of this paper is as follows. Section 2 shows the model with complete information. Section 3 introduces psychological assumptions on social influence. Section 4 defines a solution concept named mutual dominance, which is a modification of iterative dominance and a generalization of mixed strategy Nash equilibrium. Section 5 shows the main theorem that it is the unique behavior consistent with mutual dominance that the agents announce honestly at all times. Section 6 extends the model to the incomplete information case.

2. The Model

Let $N \equiv \{1,2,...,n\}$ denote the set of *agents*, where $n \geq 2$. Let A denote the set of *alternatives*. Let us consider a decision problem with the continuous time horizon $[0,\infty)$, in which, a *principal* makes an alternative choice according to the following agents' announcement procedure, denoted by $\Gamma \equiv (M,g,r)$. Let M_i denote the set of *messages* for each agent $i \in N$. Let $M \equiv \underset{i\in N}{\times} M_i$ denote the set of message profiles. At the initial time 0, the principal requires each agent i to announce any message $m_i \in M_i$. At any time after the initial time 0, and many times, she/he can change her/his message. It is assumed that at any time, each agent cannot observe the other agents' announcements, and therefore, cannot make her/his alternative choice contingent on the other agents' past announcements. Based on this assumption, we define a *strategy* for agent i as a function $s_i : [0,\infty) \to M_i$, where $s_i(t) \in M_i$ implies the message that agent i announces at time t, that is, implies the last message that she/he has changed until time t. We assume that s_i is right-continuous, that is, for every $t > 0$, either

$$s_i(\tilde{t}) = s_i(t) \quad \text{for all } \tilde{t} \geq t,$$

or there exists $t' > t$ such that

$$s_i(t') \neq s_i(t), \text{ and}$$

$$s_i(\tilde{t}) = s_i(t) \quad \text{for all } \tilde{t} \in [t,t').$$

Let S_i denote the set of strategies for agent i. Let $S \equiv \underset{i\in N}{\times} S_i$ denote the set of strategy profiles. Let $S_{-i} \equiv \underset{j\in N\setminus\{i\}}{\times} S_j$ for each $i \in N$.

The principal *randomly* terminates this announcement procedure at a constant *hazard* rate $r \in (0,\infty)$. Hence, for every $t \in [0,\infty)$, the probability that the announcement procedure terminates at or after any time t is given by

$$\exp(-rt).$$

When the announcement procedure terminates at any time t, the principal makes an alternative choice on the basis of the message profile $s(t) = (s_i(t))_{i \in N} \in M$ that has been announced just at this terminal time t; she/he selects the alternative $g(s(t)) \in A$ according to the *decision function* $g : M \to A$, along with the message profile announced at the terminal time.

Additional accounts for this announcement procedure are as follows. Before the initial time 0, the principal explains to each agent her/his *wishes* for this decision problem, such as "I wish to aid the poorest persons."[6] The principal then requires each agent to give as her/his message any relevant information that is unknown to the principal, such as the answer to the question of "where the poorest persons live?" Given that the agents have announced any message profile $m \in M$ at the randomly determined terminal time, the principal will regard the corresponding alternative $g(m) \in A$ as being the desirable one in light of her/his wishes.

For each $i \in N$, let us fix a message $m_i^* \in M_i$ as the *honest message* for agent i, which implies the best answer by agent i in line with the principal's wishes. Let $m^* = (m_i^*)_{i \in N} \in M$ denote the *honest message profile*. We define the *honest strategy* $s_i^* \in S_i$ for agent i by

$$s_i^*(t) = m_i^* \text{ for all } t \geq 0.$$

According to s_i^*, agent i announces honestly at all times. Let $s^* = (s_i^*)_{i \in N} \in S$ denote

[6] This paper does not depend on whether the principal's wishes are prosocial, antisocial, or neither. See Footnote 2.

the *honest strategy profile*.

For every $s_i \in S_i / \{s_i^*\}$, we denote by $t_i(s_i) \in [0, \infty)$ the *first* time at which agent i announces *dishonestly*, where

$$s_i(t_i(s_i)) \neq m_i^*, \text{ and}$$

$$s_i(\tilde{t}) = m_i^* \text{ for all } \tilde{t} < t_i(s_i).$$

For convenience, let us define $t_i(s_i^*) = \infty$. For every $t > 0$ and every strategy $s_i \in S_i / \{s_i^*\}$ for agent i, we define another strategy $s_{i,t} \in S_i$ for agent i by

$$s_{i,t}(\tilde{t}) = m_i^* \text{ for all } \tilde{t} \in [0, t), \text{ and}$$

$$s_{i,t}(\tilde{t}) = s_i(\tilde{t}) \text{ for all } \tilde{t} \geq t.$$

According to $s_{i,t}$, agent i announces honestly before time t, whereas she/he follows s_i at or after time t.

3. Obedience and Conformity

We denote by $U_i : S \to R$ the *payoff* function for agent i, where $U_i(s)$ implies the payoff for agent i when she/he follows the strategy $s_i \in S_i$ and expects the other agents to follow the profile of strategies $s_{-i} \in S_{-i}$. We define a *game* as a combination of the announcement procedure and the profile of the payoff functions $(\Gamma, (U_i)_{i \in N})$. We assume that the payoff $U_i(s)$ for agent i is separated into two parts;

$$U_i(s) = V_i(s) - W_i(s).$$

The first part $V_i(s)$ is called the *material* payoff, whereas the second part $W_i(s)$ is called the *psychological* cost. The material payoff $V_i(s)$ implies the expected value of the intrinsic utility $v_i(a) \in R$ that is derived directly from the alternative choice, that is,

$$V_i(s) \equiv \int_{t=0}^{\infty} v_i(g(s(t))) d[1 - \exp(-rt)].$$

We introduce an assumption on $v_i(a)$, which implies *incentive compatibility in terms of intrinsic utilities derived directly from the alternative choices*, as follows.

Assumption 1: For every $i \in N$,

(1) $$v_i(g(m^*)) \geq v_i(g(m^* / m_i)) \text{ for all } m_i \in M_i.$$

Assumption 1 implies that each agent can maximize her/his intrinsic utility derived directly from the alternative choice by announcing honestly, provided the other agents announce honestly. Clearly from Assumption 1, for every $i \in N$,

(2) $$V_i(s^*) \geq V_i(s^* / s_i) \text{ for all } s_i \in S_i,$$

which implies *incentive compatibility in terms of material payoffs*, where each agent can maximize her/his material utility by announcing honestly at all times, provided she/he expects the other agents to announce honestly at all times.

In addition to the impact of the alternative choices on her/his material payoff $V_i(s)$, each agent cares about social influences, such as *obedience* to the principal's wishes, and *conformity* to the other agents' mode of behavior, which determine her/his psychological cost $W_i(s)$. Let us introduce two assumptions on $W_i(s)$ as follows.

Assumption 2: For every $i \in N$,

(3) $$W_i(s^*) < W_i(s^* / s_i) \text{ for all } s_i \in S_i \setminus \{s_i^*\}.$$

Assumption 2 implies that whenever each agent expects the other agents to announce honestly at all times, then she/he feels guilty about her/his dishonest announcements. In this case, she/he can save her/his psychological cost by announcing honestly at all times. It is, however, implicit in Assumption 2 that the degree to which each agent i can save her/his psychological cost is very limited.

Assumption 3: For every $i \in N$, every $j \in N \setminus \{i\}$, and every $s \in S$, if

$$t_j(s_j) < \infty \text{, and}$$

$$t_i(s_i) \le t_j(s_j) \le t_h(s_h) \text{ for all } h \in N \setminus \{i, j\},$$

then

(4) $$\lim_{\varepsilon \downarrow 0} \frac{W_i(s) - W_i(s / s_{i, t_j(s_j)+\varepsilon})}{\varepsilon} > L_i r \exp(-r t_j(s_j)),$$

where $L_i \equiv \max_{(a,a')\in A^2} |v_i(a) - v_i(a')|$ denotes the upper bound of differences in intrinsic utilities for agent i.

Assumption 3 implies that at any time, each agent's psychological cost of announcing dishonestly is smaller if she/he expects that there exists another agent who has already announced dishonestly. Hence, any agent can save her/his psychological cost by waiting for any other agent to announce dishonestly earlier than she/he, that is, by *avoiding being one of the first persons to announce dishonestly*. Note that in Assumption 3, agent i is one of the first persons to announce dishonestly, whereas agent j is one of the first persons except for agent i to announce dishonestly. In this case, by deferring the first time to announce dishonestly from time $t_i(s_i)$ to time $t_j(s_j)+\varepsilon$, agent i can avoid being one of the first persons to announce dishonestly and save her/his psychological cost at least by

$$\varepsilon L_i r \exp(-r t_j(s_j)),$$

where $\varepsilon > 0$ is chosen close to zero. Assumption 3 implies that even if each agent expects the other agents to announce *dishonesty* at all times, she/he feels guilty about her/his dishonest announcements; irrespective of whether the other agents' strategies are honest or not, she/he can save her/his psychological cost by announcing honestly for the first very short time interval $[0,\varepsilon)$. It is, however, implicit in Assumption 3 that the degree to which each agent i can save her/his psychological cost is very limited.[7]

[7] We must note that the notation s_{-i} in agent i's psychological cost $W_i(s_i, s_{-i})$ implies, not the profile of strategies that the other agents actually play, but the profile of strategies that agent i *expects* the other agents to play.

Example 1: Let us consider a special case in which each agent $i's$ psychological cost is defined as the expected value of her/his *experienced* psychological disutility, that is,

$$W_i(s) \equiv \int_{t=0}^{\infty} w_i(s;t)d[1-\exp(-rt)],$$

where agent i experiences her/his psychological disutility $w_i(s,t)$ when the announcement procedure terminates at time t. It was assumed that $w_i(s,t)$ is independent of $s(t')$ for all $t' > t$. In this case, Assumption 3 is replaced by the following inequalities; for every $i \in N$, every $j \in N \setminus \{i\}$, every $s \in S$, and every $t \geq 0$, if

$$t_j(s_j) \leq t, \text{ and}$$

$$t_i(s_i) \leq t_j(s_j) \leq t_h(s_h) \quad \text{for all} \quad h \in N \setminus \{i, j\},$$

then

$$\lim_{\varepsilon \downarrow 0} \frac{w_i(s,t) - w_i(s / s_{i,t_j(s_j)+\varepsilon}, t)}{\varepsilon} > L_i r.$$

Example 2: Following Example 1, let us specify $w_i(s,t)$ by

$$w_i(s,t) = \lambda \frac{\int_{\tau=0}^{t} \iota_i(s_i(\tau))dt}{t} + \varepsilon \qquad \text{if} \quad t_i(s_i) \leq t_h(s_h) \quad \text{for all} \quad h \in N,$$

and

$$w_i(s,t) = \lambda \frac{\int_{\tau=0}^{t} \iota_i(s_i(\tau))dt}{t} \qquad \text{otherwise,}$$

where $\lambda > 0$, $\varepsilon > 0$, and the function $\iota_i : M_i \to \{0,1\}$ is defined by

$$\iota_i(m_i^*) = 1, \text{ and}$$

$$\iota_i(m_i) = 0 \quad \text{for all} \quad m_i \in M_i \setminus \{m_i^*\}.$$

Note that $\frac{\int_{\tau=0}^{t} \iota_i(s_i(\tau))dt}{t}$ implies the proportion of the time length that agent i announces dishonestly when the announcement procedure terminates at time t. It is clear that Assumptions 2 and 3 hold in this example. This specification implies

$$\max_{(s,s')\in S^2} \left| W_i(s) - W(s') \right| \leq \lambda + \varepsilon .$$

Hence, by letting $\lambda > 0$ and $\varepsilon > 0$ get close to zero, we can make each agent's psychological cost as negligible as possible without harming Assumptions 2 and 3.

4. Mutual Dominance

We introduce a solution concept named *mutual dominance*, which is regarded as a generalization of mixed strategy Nash equilibrium, and also regarded as a minor modification of iterative dominance, as follows. We denote by $\tilde{S}_i \subset S_i$ a subset of strategies for agent i. Let us denote $\tilde{S} \equiv \underset{i \in N}{\times} \tilde{S}_i$ and $\tilde{S}_{-i} \equiv \underset{j \in N \setminus \{i\}}{\times} \tilde{S}_j$ for all $i \in N$. For convenience, let us confine our attention to any subset of strategy profiles $\tilde{S}$ satisfying a minimal requirement of compactness in the sense that either

$$\tilde{S}_i = \{s_i^*\} \text{ for all } i \in N,$$

or there exist $i \in N$ and $s_i \in \tilde{S}_i$ such that

$$t_i(s_i) \leq t_j(s_j) \text{ for all } j \in N \text{ and all } s_j \in \tilde{S}_j.$$

A subset of strategy profiles $\tilde{S} \subset S$ is said to be *mutually undominated* in the game $(\Gamma, (U_i)_{i \in N})$ if there exist no $i \in N$, $s_i \in \tilde{S}_i$, and $s_i' \in S_i$ such that

$$U_i(s_i', s_{-i}) > U_i(s_i, s_{-i}) \text{ for all } s_{-i} \in \tilde{S}_{-i}.$$

Mutual dominance implies that any strategy for each agent i in $\tilde{S}_i$ is undominated as long as the other agents follow any profile of their strategies in $\tilde{S}_{-i}$.

Mutual dominance is related to mixed strategy Nash equilibrium as follows. A *mixed* strategy for each agent i is defined as a simple lottery $\alpha_i : S \to [0,1]$ on S_i, where the support of α_i is countable, and $\sum_{s_i \in S_i} \alpha_i(s_i) = 1$. Let Λ_i denote the set of mixed strategies for agent i. Let $\Lambda \equiv \underset{i \in N}{\times} \Lambda_i$ denote the set of mixed strategy profiles.

A mixed strategy profile $\alpha \equiv (\alpha_i)_{i\in N} \in \Lambda$ is said to be a *Nash equilibrium* in the game $(\Gamma,(U_i)_{i\in N})$ if

$$U_i(\alpha) \geq U_i(\alpha_i', \alpha_{-i}) \text{ for all } i \in N \text{ and all } \alpha_i' \in \Lambda_i,$$

where we define $U_i(\alpha) \equiv \sum_{s\in S} U_i(s) \prod_{i\in N} \alpha_i(s)$. It is clear that *if a mixed strategy profile $\alpha \in \Lambda$ is a Nash equilibrium and $\tilde{S}_i$ is equivalent to the support of α_i for all $i \in N$, then $\tilde{S}$ is mutually undominated.* It is also clear that *if $\tilde{S}$ is mutually undominated and $\tilde{S}_i = \{s_i\}$ for all $i \in N$, then $s = (s_i)_{i\in N}$ must be a pure strategy Nash equilibrium.*

Mutual dominance is related to *iterative dominance* as follows. Let $S_i^0 = S_i$ for all $i \in N$. Recursively, for every positive integer $k \geq 1$, let $S_i^k \subset S_i$ denote the set of strategies s_i for agent $i \in N$ such that $s_i \in S_i^{k-1}$, and there exists no $s_i' \in S_i$ such that

$$U_i(s_i', s_{-i}) > U_i(s_i, s_{-i}) \text{ for all } s_{-i} \in S_{-i}^{k-1}.$$

We define the set of iteratively undominated strategies for agent i by

$$S_i^\infty \equiv \bigcap_{k=0}^{\infty} S_i^k.$$

It is clear that if $\tilde{S}_i = S_i^\infty$ for all $i \in N$, then $\tilde{S}$ is mutually undominated.

5. Main Theorem

The following theorem implies that *the honest strategy profile* s^* *is the only strategy profile that is consistent with mutual dominance, and therefore, it is the unique mixed strategy Nash equilibrium in the game* $(\Gamma, (U_i)_{i \in N})$.

Theorem 1: *A subset of strategy profiles* $\tilde{S} \subset S$ *is mutually undominated in the game* $(\Gamma, (U_i)_{i \in N})$ *if and only if*

$$\tilde{S}_i = \{s_i^*\} \text{ for all } i \in N .$$

Proof: Suppose $\tilde{S}_i = \{s_i^*\}$ for all $i \in N$. Then, it is clear from Assumptions 1 and 2 that $\tilde{S}$ is mutually undominated; inequalities (2) and (3) imply that for every $i \in N$ and every $s_i \in S_i \setminus \{s_i^*\}$,

$$V_i(s^*) - W_i(s^*) > V_i(s^* / s_i) - W_i(s^* / s_i), \text{ that is, } U_i(s^*) > U_i(s^* / s_i) .$$

Suppose that $\tilde{S}_i \neq \{s_i^*\}$ for some $i \in N$. Then, there exist $i \in N$ and $s_i \in \tilde{S}_i$ such that

$$t_i(s_i) < \infty, \text{ and}$$

$$t_i(s_i) \leq t_j(s_j) \text{ for all } j \in N \text{ and all } s_j \in \tilde{S}_j .$$

Let us fix $s_{-i} \in \tilde{S}_{-i}$ arbitrarily. Suppose that there exists $j \in N \setminus \{i\}$ such that

$$t_j(s_j) < \infty, \text{ and}$$

$$t_j(s_j) \leq t_h(s_h) \text{ for all } h \in N \setminus \{i, j\}.$$

Let us choose $\varepsilon > 0$ close to zero. Then, from Assumption 1 and the definition of L_i,

$$V_i(s) - V_i(s / s_{i,t_j(s_j)+\varepsilon})$$

$$= \int_{t=0}^{t_j(s_j)} \{v_i(g(s(t))) - v_i(g(s(t) / s_{i,t_j(s_j)+\varepsilon}(t)))\} d[1 - \exp(-rt)]$$

$$+ \int_{t=t_j(s_j)}^{t_j(s_j)+\varepsilon} \{v_i(g(s(t))) - v_i(g(s(t) / s_{i,t_j(s_j)+\varepsilon}(t)))\} d[1 - \exp(-rt)]$$

$$= \int_{t=0}^{t_j(s_j)} \{v_i(g(m^* / s_i(t))) - v_i(g(m^*))\} d[1 - \exp(-rt)]$$

$$+ \int_{t=t_j(s_j)}^{t_j(s_j)+\varepsilon} \{v_i(g(s(t))) - v_i(g(s(t) / m_i^*))\} d[1 - \exp(-rt)]$$

$$\leq L_i[\exp\{-rt_j(s_j)\} - \exp(-r\{t_j(s_j) + \varepsilon\})],$$

which is approximated by $\varepsilon L_i r \exp(-rt_j(s_j))$. Moreover, from inequalities (4) in Assumption 3,

$$W_i(s) - W_i(s / s_{i,t_j(s_j)+\varepsilon}) > \varepsilon L_i r \exp(-rt_j(s_j)).$$

From these observations, we have shown that

$$U_i(s) - U_i(s / s_{i,t_j(s_j)+\varepsilon})$$

$$= V_i(s) - V_i(s / s_{i,t_j(s_j)+\varepsilon}) - \{W_i(s) - W_i(s / s_{i,t_j(s_j)+\varepsilon})\}$$

$$< \varepsilon L_i r \exp(-rt_j(s_j)) - \varepsilon L_i r \exp(-rt_j(s_j)) = 0.$$

Next, suppose that

$$t_j(s_j) = \infty \text{ for all } j \in N / \{i\}\text{, that is, } s_{-i} = s_{-i}^*.$$

Then, from inequalities (2) and (3),

$$U_i(s) - U_i(s/s_i^*) = V_i(s) - V_i(s/s_i^*) - W_i(s) + W_i(s/s_i^*)$$

$$= V_i(s/s_i) - V_i(s^*) - W_i(s/s_i) + W_i(s^*) < 0 .$$

From the above arguments, we have proven that $\tilde{S}$ is not mutually undominated, which is a contradiction.

Q.E.D.

Each agent feels guilty about announcing dishonestly if she/he expects that the other agents have never announced dishonestly. Hence, in order to reduce her/his psychological cost caused by guilt-aversion, she/he may prefer postponing the time to announce dishonestly until any other agent announces dishonestly in her/his expectation. However, there may exist the difficulty that when postponing her/his announcement, she/he may be caught between the reduction in her/his psychological cost and the loss in her/his material payoff, both of which are commonly caused by the changes of her/his messages from the dishonest ones to the honest ones.

Assumption 3 does overcome this difficulty. Suppose that any agent can avoid being the first person to announce dishonestly by delaying her/his time for a very short interval. Note from Assumption 3 that the probability of the announcement procedure terminating during this short interval is kept small enough to make the expected value of the loss in her/his intrinsic utility less than the reduction in her/his psychological cost, both of which are commonly caused by her/his message changes during this short interval. Hence, the smallness of probabilities in this manner triggers the tail-chasing competition among the agents that edges their timing upward endlessly, eliminating unwanted strategies.

With respect to the functioning of the tail-chasing competition, our model is related to the basic concept of the Abreu-Matsushima mechanism (Abreu and Matsushima (1992a, 1994)). In the Abreu-Matsushima mechanism, each agent announces multiple messages[8] and is motivated to avoid being one of the first persons who announce messages that are inconsistent with their first messages, triggering a tail-chasing competition amongst the agents.

There exist many substantive points of difference between our model and the Abreu-Matsushima mechanism, and they are as follows. In order to trigger the tail-chasing competition, the Abreu-Matsushima mechanism uses any *contractual* device of side payments or similar to this, stipulating that any agent is fined by a small amount of money if and only if she/he is one of the first persons who announce messages that are inconsistent with their first messages. In contrast to the Abreu-Matsushima mechanism, our model never uses any such contractual device. The Abreu-Matsushima mechanism also needs additional devices that incentivize the agents to make honest first announcements, which are generally tailored for the details of the model specifications in complicated ways.[9] Because of the use of incentive devices in this manner, all the messages announced must be *verifiable* to the court in the Abreu-Matsushima mechanism. In contrast, our model does not require this verifiability except in the case of the agents' final announcements, because we do not use any further contractual device contingent on their announcements.

It was assumed in our model that each agent cannot observe the other agents'

---

[8] In the Abreu-Matsushima mechanism, the number of messages that each agent actually announces is fixed; this number is uncertain in our model since the announcement procedure is randomly terminated.

[9] Matsushima (2008a, 2008b) introduced psychological costs into the Abreu-Matsushima mechanism design in order to avoid this complexity.

announcements. If we permit each agent to observe them, then we must investigate a version of an infinitely repeated game and may struggle with the multiplicity of equilibria implied by the Folk Theorem or some similar principle.[10] In spite of this multiplicity, the above assumption is not very crucial for our main theorem. In fact, we can ease this multiplicity by setting the hazard rate sufficiently large. By doing this way, it is sure that the restriction implied by Assumption 3 becomes stronger. As Example 2 expresses, however, irrespective of how to set the hazard rate, Assumption 3 holds automatically whenever the function $W_i(s)$ of each agent $i's$ psychological payoff is discontinuous at the first time that she/he make a dishonest announcement earlier than any other agent.

[10] With minor modifications added, the Abreu-Matsushima mechanism functions even if the agents' announcements are observable; by regarding the last messages as the references and fining any final deviant, we can show that the permissive results hold irrespective of whether the agents' announcements are observable or not.

6. Incomplete Information

We can extend our model to the *incomplete information* case as follows. Before the announcement procedure is started, each agent $i \in N$ receives her/his *private signal* denoted by $\omega_i \in \Omega_i$, where Ω_i is the finite set of private signals for agent i. Let $\omega \equiv (\omega_i)_{i \in N}$ denote a private signal profile, and let $\Omega \equiv \underset{i \in N}{\times} \Omega_i$ denote the set of private signal profiles. Conditionally on agent $i's$ private signal $\omega_i \in \Omega_i$, the probability that the other agents receive any profile of their private signals $\omega_{-i} = (\omega_j)_{j \in N \backslash \{i\}} \in \Omega_{-i} \equiv \underset{j \in N \backslash \{i\}}{\times} \Omega_j$ is given by $p_i(\omega_{-i} | \omega_i) \in (0,1]$, where we assume $\sum_{\omega_{-i} \in \Omega_{-i}} p_i(\omega_{-i} | \omega_i) = 1$.

For every $i \in N$ and every $\omega_i \in \Omega_i$, let us fix a message $m_i^{*\omega_i} \in M_i$ as the honest message for agent i associated with her/his private signal $\omega_i \in \Omega_i$. Let $m^{*\omega} = (m_i^{*\omega_i})_{i \in N} \in M$ denote the honest message profile associated with the private signal profile $\omega \in \Omega$. Let us specify $s_i^{*\omega_i} \in S_i$ by

$$s_i^{*\omega_i}(t) = m_i^{*\omega_i} \quad \text{for all} \quad t \geq 0.$$

According to $s_i^{*\omega_i}$, agent i announces the honest message $m_i^{*\omega_i}$ associated with her/his private signal $\omega_i \in \Omega_i$ at all times. Let $s^{*\omega} = (s_i^{*\omega_i})_{i \in N} \in S$. For every $\omega_i \in \Omega_i$ and every $s_i \in S_i / \{s_i^{*\omega_i}\}$, we denote by $t_i^{\omega_i}(s_i) \in [0, \infty)$ the *first* time at which agent i announces any message that is different from the honest message $m_i^{*\omega_i}$ associated with her/his private signal ω_i, where

$$s_i(t_i^{\omega_i}(s_i)) \neq m_i^{*\omega_i} \text{, and}$$

$$s_i(t) = m_i^{*\omega_i} \quad \text{for all} \quad t < t_i^{\omega_i}(s_i) .$$

Let $t_i^{\omega_i}(s_i^{*\omega_i}) = \infty$. For every $t > 0$, every $\omega_i \in \Omega_i$, and every $s_i \in S_i / \{s_i^{*\omega_i}\}$, we define $s_{i,\omega_i,t} \in S_i$ by

$$s_{i,\omega_i,t}(\tilde{t}) = m_i^{*\omega_i} \quad \text{for all} \quad \tilde{t} \in [0,t) \text{, and}$$

$$s_{i,\omega_i,t}(\tilde{t}) = s_i(\tilde{t}) \quad \text{for all} \quad \tilde{t} \geq t .$$

According to $s_{i,\omega_i,t}$, agent i announces the honest message associated with her/his private signal ω_i before time t, whereas she/he follows s_i at or after time t.

We define the payoff function for agent i as $U_i : S \times \Omega \to R$, where

$$U_i(s,\omega) = V_i(s,\omega) - W_i(s,\omega) .$$

The first part $V_i(s,\omega)$ is called the *material* payoff associated with the private signal profile $\omega \in \Omega$, and the second part $W_i(s,\omega)$ is called the *psychological* cost associated with the private signal profile $\omega \in \Omega$. The material payoff $V_i(s,\omega)$ is regarded as the expected value of the intrinsic utility $v_i(a,\omega) \in R$ derived directly from the alternative choice when the private signal profile $\omega \in \Omega$ takes place, that is,

$$V_i(s,\omega) \equiv \int_{t=0}^{\infty} v_i(g(s(t)),\omega) d[1-\exp(-rt)] .$$

We introduce the following assumption on $v_i(a,\omega)$ that corresponds to Assumption 1 , which implies *Bayesian incentive compatibility in terms of intrinsic utilities derived directly from the alternative choices.*

Assumption 4: For every $i \in N$, every $\omega_i \in \Omega_i$, and every $m_i \in M_i$,

(5) $$\sum_{\omega_{-i}\in\Omega_{-i}} v_i(g(m^{*\omega}),\omega)p_i(\omega_{-i}\,|\,\omega_i) \geq \sum_{\omega_{-i}\in\Omega_{-i}} v_i(g(m^{*\omega}/m_i),\omega)p_i(\omega_{-i}\,|\,\omega_i)\,.$$

Assumption 4 implies that each agent can maximize the expected value of her/his intrinsic utility derived directly from the alternative choice conditional on her/his private signal by announcing honestly, provided the other agents announce honestly.

We introduce the following two assumptions on $W_i(s,\omega)$, that is, Assumptions 5 and 6, which correspond to Assumptions 2 and 3, respectively.

Assumption 5: For every $i \in N$ and every $\omega \in \Omega$,

(6) $$W_i(s^{*\omega},\omega) < W_i(s^{*\omega}/s_i,\omega) \quad \text{for all} \quad s_i \in S_i \setminus \{s_i^{*\omega_i}\}\,.$$

Assumption 6: For every $i \in N$, every $j \in N \setminus \{i\}$, every $\omega \in \Omega$, and every $s \in S$, if

$t_j^{\omega_j}(s_j) < \infty$, and

$t_i^{\omega_i}(s_i) \leq t_j^{\omega_j}(s_j) \leq t_h^{\omega_h}(s_h)$ for all $h \in N \setminus \{i, j\}$,

then

(7) $$\lim_{\varepsilon\downarrow 0} \frac{W_i(s,\omega) - W_i(s/s_{i,\omega_i,t_j^{\omega_j}(s_j)+\varepsilon},\omega)}{\varepsilon} > L_i r\exp(-rt_j^{\omega_j}(s_j))\,,$$

where we redefine $L_i \equiv \max\limits_{(a,a',\omega)\in A^2\times\Omega} |v_i(a,\omega) - v_i(a',\omega)|$.

Assumption 5 implies that whenever each agent expects the other agents to announce honestly at all times, then she/he can save her/his psychological cost by announcing honestly at all times. Assumption 6 implies that each agent can save her/his psychological cost by waiting for any other agent to announce dishonestly earlier than she/he.

We shall introduce an additional assumption on $W_i(s,\omega)$ as follows.

Assumption 7: For every $i \in N$, every $\omega \in \Omega$, every $s \in S$, and every $t > t_i^{\omega_i}(s_i)$,

(8) $$W_i(s,\omega) > W_i(s/s_{i,\omega_i,t},\omega).$$

Assumption 7 implies that each agent can save her/his psychological cost by postponing the first time to announce dishonestly, irrespective of whether the other agents announce honestly. It is implicit in Assumptions 5, 6, and 7 that the degree to which each agent i can save her/his psychological cost is very limited.

In the incomplete information case, *a strategy for agent* i is redefined as

$$\sigma_i : [0,\infty) \times \Omega_i \to M_i,$$

where $\sigma_i(t,\omega_i) \in M_i$ implies the message that agent i announces at time t when her/his private signal is given by $\omega_i \in \Omega_i$. We assume that $\sigma_i(t,\omega_i)$ is right-continuous with respect to $t \in [0,\infty)$. Let Σ_i denote the set of strategies for agent i. Let $\Sigma \equiv \underset{i \in N}{\times} \Sigma_i$ denote the set of strategy profiles. In the incomplete information case, the *honest strategy* for agent i is denoted by $\sigma_i^* \in \Sigma_i$, which is defined as

$$\sigma_i^*(\cdot,\omega_i) = s_i^{*\omega_i} \text{ for all } \omega_i \in \Omega_i.$$

The *expected* payoff for agent i when she/he receives the private signal $\omega_i \in \Omega_i$ is defined as

$$EU_i(\sigma,\omega_i) \equiv \sum_{\omega_{-i} \in \Omega_{-i}} U_i(s,\omega) p(\omega_{-i} \mid \omega_i).$$

We define a *Bayesian game* as $(\Gamma,(EU_i)_{i\in N})$. We denote by $\tilde{\Sigma}_i \subset \Sigma_i$ a subset of strategies for agent i. Let $\tilde{\Sigma} \equiv \underset{i\in N}{\times} \tilde{\Sigma}_i \subset \Sigma$. We shall confine our attention to any subset of strategy profiles $\tilde{\Sigma}$ satisfying that either

$$\tilde{\Sigma}_i = \{\sigma_i^*\} \text{ for all } i \in N,$$

or there exist $i \in N$, $\sigma_i \in \tilde{\Sigma}_i$, and $\omega_i \in \Omega_i$ such that

$$t_i^{\omega_i}{}_i(\sigma_i(\cdot,\omega_i)) \leq t_j^{\omega_j}(\sigma_j(\cdot,\omega_j)) \text{ for all } j \in N\text{, all } \sigma_j \in \tilde{\Sigma}_j\text{, and } \omega_j \in \Omega_j.$$

A subset of strategy profiles $\tilde{\Sigma} \subset \Sigma$ is said to be *Bayesian mutually undominated* in $(\Gamma,(EU_i)_{i\in N})$ if there exist no $i \in N$, $\sigma_i \in \tilde{\Sigma}_i$, $\sigma_i' \in \Sigma_i$, and $\omega_i \in \Omega_i$ such that

$$EU_i(\sigma/\sigma_i',\omega_i) > EU_i(\sigma,\omega_i) \text{ for all } \sigma_{-i} \in \tilde{\Sigma}_{-i}.$$

We can regard Bayesian mutual dominance as a generalized concept of Bayesian Nash equilibrium, and can relate Bayesian mutual dominance to iterative dominance in the incomplete information case, in the same way as in Section 4. The following theorem corresponds to Theorem 1; *the honest strategy profile σ^* is the only strategy profile that is consistent with Bayesian mutual dominance.*

Theorem 2: *A subset $\tilde{\Sigma} \subset \Sigma$ is Bayesian mutually undominated in the Bayesian game $(\Gamma,(EU_i)_{i\in N})$ if and only if*

$$\tilde{\Sigma}_i = \{\sigma_i^*\} \textit{ for all } i \in N.$$

Proof: Suppose $\tilde{\Sigma}_i = \{\sigma_i^*\}$ for all $i \in N$. Then, from Assumptions 4 and 5, it follows that for every $i \in N$, every $\omega_i \in \Omega_i$, and every $\sigma_i \in S_i \setminus \{\sigma_i^*\}$,

$$\sum_{\omega_{-i}\in\Omega_{-i}} V_i(s^{*\omega},\omega)p_i(\omega_{-i}\,|\,\omega_i) \geq \sum_{\omega_{-i}\in\Omega_{-i}} V_i(s^{*\omega}/\sigma_i(\cdot,\omega_i),\omega)p_i(\omega_{-i}\,|\,\omega_i)\text{, and}$$

$$\sum_{\omega_{-i}\in\Omega_{-i}} W_i(s^{*\omega},\omega)p_i(\omega_{-i}\,|\,\omega_i) < \sum_{\omega_{-i}\in\Omega_{-i}} W_i(s^{*\omega}/\sigma_i(\cdot,\omega_i),\omega)p_i(\omega_{-i}\,|\,\omega_i)\,,$$

which imply

$$EU_i(\sigma^*,\omega_i) - EU_i(\sigma^*/\sigma_i,\omega_i)$$

$$= \sum_{\omega_{-i}\in\Omega_{-i}} V_i(s^{*\omega},\omega)p_i(\omega_{-i}\,|\,\omega_i) - \sum_{\omega_{-i}\in\Omega_{-i}} W_i(s^{*\omega},\omega)p_i(\omega_{-i}\,|\,\omega_i)$$

$$- \sum_{\omega_{-i}\in\Omega_{-i}} V_i(s^{*\omega}/\sigma_i(\cdot,\omega_i),\omega)p_i(\omega_{-i}\,|\,\omega_i) + \sum_{\omega_{-i}\in\Omega_{-i}} W_i(s^{*\omega}/\sigma_i(\cdot,\omega_i),\omega)p_i(\omega_{-i}\,|\,\omega_i)$$

$$> 0\,.$$

Hence, $\tilde{\Sigma}$ is Bayesian mutually undominated.

Suppose that $\tilde{\Sigma}_i \neq \{\sigma_i^*\}$ for some $i \in N$. Then, there exist $i \in N$, $\sigma_i \in \tilde{\Sigma}_i$, and $\omega_i \in \Omega_i$ such that

$$t_i^{\omega_i}(\sigma_i(\cdot,\omega_i)) < \infty\text{, and}$$

$$t_i^{\omega_i}(\sigma_i(\cdot,\omega_i)) \leq t_j^{\omega_j}(\sigma_j(\cdot,\omega_j)) \text{ for all } j \in N\text{, all } \sigma_j \in \tilde{\Sigma}_j\text{, and all}$$

$$\omega_j \in \Omega_j\,.$$

Fix $\sigma_{-i} \in \tilde{\Sigma}_{-i}$ and $\omega_{-i} \in \Omega_{-i}$ arbitrarily. Let us choose $\varepsilon > 0$ close to zero. Suppose that there exists $j \in N/\{i\}$ such that

$$t_i^{\omega_i}(\sigma_i(\cdot,\omega_i)) = t_j^{\omega_j}(\sigma_j(\cdot,\omega_j))\,.$$

Then, from the definition of L_i, it follows

$$V_i(\sigma(\cdot,\omega),\omega) - V_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega)$$

$$= \int_{t=t_i(s_i)}^{t_i(s_i)+\varepsilon} \{v_i(g(\sigma(t,\omega)),\omega) - v_i(g(\sigma(t,\omega)/m_i^{*\omega_i}),\omega)\}d[1-\exp(-rt)]\,,$$

$$\leq L_i[\exp\{-rt_j^{\omega_j}(\sigma_j(\cdot,\omega_j))\} - \exp(-r\{t_j^{\omega_j}(\sigma_j(\cdot,\omega_j)) + \varepsilon\})],$$

which is approximated by $\varepsilon L_i r \exp(-rt_j^{\omega_j}(\sigma_j(\cdot,\omega_j)))$. Since $\varepsilon > 0$ is close to zero, it follows from Assumption 6 that

$$W_i(\sigma(\cdot,\omega),\omega) - W_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega).$$

$$> \varepsilon L_i r \exp(-rt_j^{\omega_j}(\sigma_j(\cdot,\omega_j))).$$

From these inequalities, we have proven that

$$U_i(\sigma(\cdot,\omega),\omega) - U_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega)$$

$$= V_i(\sigma(\cdot,\omega),\omega) - V_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega)$$

$$-W_i(\sigma(\cdot,\omega),\omega) + W_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega)$$

$$< \varepsilon L_i r \exp(-rt_j^{\omega_j}(\sigma_j(\cdot,\omega_j))) - \varepsilon L_i r \exp(-rt_j^{\omega_j}(\sigma_j(\cdot,\omega_j))) = 0.$$

Next, suppose that there exists $j \in N/\{i\}$ such that

$t_i^{\omega_i}(\sigma_i(\cdot,\omega_i)) < t_j^{\omega_j}(\sigma_j(\cdot,\omega_j))$, and

$t_j^{\omega_j}(\sigma_j(\cdot,\omega_j)) \leq t_{i'}^{\omega_{i'}}(\sigma_{i'}(\cdot,\omega_{i'}))$ for all $i' \in N \setminus \{i,j\}$.

Then, by letting $\varepsilon > 0$ be less than $t_j^{\omega_j}(\sigma_j(\cdot,\omega_j)) - t_i^{\omega_i}(\sigma_i(\cdot,\omega_i)) > 0$, it follows from Assumption 4 that

$$V_i(\sigma(\cdot,\omega),\omega) - V_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega)$$

$$= \int_{t=t_i(s_i)}^{t_i(s_i)+\varepsilon} \{v_i(g(m^{*\omega}/\sigma_i(t,\omega_i)),\omega) - v_i(g(m^{*\omega}),\omega)\} d[1-\exp(-rt)] \leq 0.$$

Assumption 7 implies

$$W_i(\sigma(\cdot,\omega),\omega) - W_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega) > 0.$$

From these inequalities, it follows that

$$U_i(\sigma(\cdot,\omega),\omega) - U_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega)$$

$$= V_i(\sigma(\cdot,\omega),\omega) - V_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega)$$

$$- W_i(\sigma(\cdot,\omega),\omega) + W_i(\sigma(\cdot,\omega)/\sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i),\omega) < 0 .$$

From the above observations, by letting $\varepsilon > 0$ get close to zero, we have proven that

$$EU_i(\sigma/\sigma_i',\omega_i) > EU_i(\sigma,\omega_i) \text{ for all } \sigma_{-i} \in \tilde{\Sigma}_{-i} ,$$

where $\sigma_i' \in \Sigma_i$ is specified by $\sigma_i'(\cdot,\omega_i) = \sigma_{i,\omega_i,t_i(s_i)+\varepsilon}(\cdot,\omega_i)$ for all $\omega_i \in \Omega_i$. This, however, contradicts the supposition of $\sigma_i \in \tilde{\Sigma}_i$.

Q.E.D.

Let us define a function $f : \Omega \to A$ by

$$f(\omega) = g(m^{*\omega}) \text{ for all } \omega \in \Omega ,$$

which corresponds to the concept of *social choice function* in the standard theory of implementation. Theorem 2 implies that *the social choice function* f *is implementable in Bayesian Nash equilibrium*, that is, σ^* is the unique Bayesian Nash equilibrium, and

$$g(s^*(\omega)) = f(\omega) \text{ for all } \omega \in \Omega .$$

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, 1992b.

Abreu, D. and H. Matsushima (1994): "Exact Implementation," *Journal of Economic Theory* 64, 1–19.

Abreu, D. and A. Sen (1990): "Virtual Implementation in Nash Equilibrium," *Econometrica* 59, 997–1021.

Akerlof, G. and W. Dickens (1982): "The Economic Consequences of Cognitive Dissonance," *American Economic Review* 72, 307–319.

Ash, S. (1955): Opinion and Social Pressure, *Scientific American* 193, 31–35.

Bébabou, R. (2007): "Groupthink: Collective Delusions in Organizations and Markets," mimeo.

Bernheim, D (1994)：A Theory of Conformity, *Journal of Political Economy* 102, 841–877.

Cialdini, R. (2001): *Influence: Science and Practice*. Boston: Allyn & Bacon.

Charness, G. and M. Dufwenberg (2006): "Promises and Partnerships," *Econometrica* 74, 1579–1601.

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

Hofling, C., E. Brontzman, S. Dalrymple, N. Graves, and C. Piece (1966): "An Experimental Study of Nurse-Physician Relationships," *Journal of Nervous and*

Mental Disease 143, 171–180.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior* 1, 60–79.

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

Gneezy, U. (2005): "Deception: The Role of Consequences," *American Economic Review* 95, 384–394.

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

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

Matsushima, H. (1988): "A New Approach to the Implementation Problem," *Journal of Economic Theory* 45, 128–144.

Matsushima, H. (2008a): "Role of Honesty in Full Implementation," *Journal of Economic Theory* 127, 353–359.

Matsushima, H. (2008b): "Behavioral Aspects of Implementation Theory," *Economics Letters* 100, 161–164.

Milgram, S. (1974): *Obedience to Authority: An Experimental View*. New York: Harper and Row.

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

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

MIT 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: Cambridge University Press.

Zimbardo, P., C. Haney, W. Banks, and D. Jaffe (1977): The Psychology of Imprisonment: Privation, Power, and Pathology, in *Contemporary Issues in Social Psychology*, ed. by J. Brigham and L. Wrightman. Cole Publishing Co.