CIRJE-F-147

# Stability and Implementation via Simple Mechanisms In the Complete Information Environments

Hitoshi Matsushima
The University of Tokyo

February 2002

# Stability and Implementation via Simple Mechanisms In the Complete Information Environments [+]

Hitoshi Matsushima[*]

Faculty of Economics, University of Tokyo

First Version: May 15, 2001
This version: February 18, 2002

## Abstract

This paper investigates implementation of social choice functions in the complete information environments. We construct particularly simple mechanisms named *local direct mechanisms*, which require each agent to make a single announcement about her own and neighbors' utility indices. We assume that each agent is *boundedly rational* in that she may announce any best reply, including disequilibrium messages, even if the others play a Nash equilibrium. We require that the honest message profile be *stable* in the global sense that it is reachable from every message profile and no other message profile is reachable from it. It is shown that with a minor restriction, every social choice function is virtually implementable. We provide naïve models of adaptive dynamics whose convergence characterizes the static definition of stability. We also investigate

---

[+] The earlier version of this paper was entitled "Stable Implementation."

[*] Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan. Fax: (03)3818-7082. e-mail: hitoshi@e.u-tokyo.ac.jp

several variants of the stability notion such as BR-stability, weak stability, and weak BR-stability.

**Keywords:** Complete Information, Local Direct Mechanisms, Stability, Small Fines, Virtual Implementation, Possibility Results, Bounded Rationality, Adaptive Dynamics.

## 1. Introduction

This paper investigates implementation of social choice functions in the complete information environments. We consider particularly simple mechanisms named *local direct mechanisms*, which require each agent to make a single announcement on her own and her two neighbors' utility indices. We allow agents to be levied only small amounts. The purpose of the paper is to address how implementation works when agents are *boundedly rational*, and show the possibility that a social choice function is virtually implementable via such simple mechanisms as local direct mechanisms.

We hypothesize that each agent is boundedly rational so that she may change her message into any message that is better than or indifferent to the message that she currently announces, given that she expects the others to continue announcing the same messages as their current messages. Hence, each agent may announce any best reply, including disequilibrium messages, even if the others play a Nash equilibrium. We require that the honest message profile be *stable* in the global sense that it is reachable from every message profile through finite steps, in each step of which, an agent *unilaterally* changes her message, whereas no other message profile is reachable from it. Hence, the honest message profile is the unique strict Nash equilibrium, and there exists no other absorbing limit cycle in terms of reachability. The main result of the paper is that with a minor restriction, every social choice function is virtually implementable in terms of stability *via* local direct mechanisms.

Several earlier works in the implementation literature have required the uniqueness of Nash equilibrium or its related solution concepts, and have provided their respective possibility theorems. See Matsushima (1988), Moore and Repullo (1988), Abreu and Sen (1990), Palfrey and Srivastava (1991), and Abreu and Matsushima (1992). See also the survey by Moore (1992). These works indicate that it might be inevitable to construct *complicated* mechanisms in that each agent must have redundant, *slack* messages that she never announces as long as agents play the equilibrium behaviors. The mechanisms constructed by these authors provide each agent with the incentive to deviate from any unwanted message profile by announcing such a slack message.

In contrast, local direct mechanisms are so simple as to have *no* slack messages. The size of the set of message profiles, however, is larger than the size of the set of preference profiles. This implies that there exist slack message *profiles* that agents never announce as long as they make the honest announcements. No messages are slack, but message profiles leading agents to announce different opinions are slack. The paper shows that the existence of such slack message profiles is enough to provide agents with the incentives to extricate themselves from the trap of the unwanted message profile into the honest message profile.

A local direct mechanism is specified as follows. For each agent, there exist three agents who announce their respective opinions about this agent's utility index. If at least two agents announce the same opinion, we shall regard it as *the public opinion*. If a message profile has the public opinion for every agent, the mechanism almost certainly enforces the allocation assigned by the social choice function to this profile. Otherwise, the mechanism certainly enforces the allocation that maximizes agent 1's utility on the supposition that she is correct, i.e., agent 1 becomes dictatorial.

A key heuristic of the paper is as follows. Suppose that a message profile does not have the profile of the public opinions. Then, each agent has a wide variety of messages satisfying that whichever message she announces agent 1 remains dictatorial and the resultant allocation never changes. This, together with the bounded rationality hypothesis, serves to provide each agent with the incentive to announce the honest message.

When the given message profile has the profile of the public opinions, we can use the idea of the *virtualness* addressed by Matsushima (1988) and Abreu and Sen (1990) to provide each agent with the incentive to make the honest announcement about her own utility index. We can specify the mechanism in ways that an agent is fined if and only if she is the single deviant from the profile of the public opinions and announces different opinions about her neighbors' utility indices. This, together with the virtualness, serves to provide each agent with the strict incentive to announce the honest message given that the others announce the honest messages. Since the use of fine serves mainly for this strictness, we can make the amount of fine as small as possible.

In order to clarify that this permissive result does not much depend on the specificity of the stability notion, we will investigate its robustness in several ways. Firstly, we will replace the bounded rationality hypothesis with a hypothesis that each agent who will change her message announces only best replies. Based on it we define the notion of *B-R stability*. B-R stability is more restrictive than stability in that it is impossible for a local direct mechanism to virtually implement a social choice function in terms of BR-stability, even though it is possible in terms of stability. We assume that there are four or more agents, and we introduce a *modified* local direct mechanism, in which each agent not only announces opinions on her own and two neighbors' utility indices but also announces either integer "0" or integer "1". The honest message for each agent is redefined as the message, according to which, she announces "0" as well as makes the honest announcements. Hence, messages with the announcement of "1" are slack. We show that with a minor restriction, every social choice function is virtually implementable in terms of BR-stability via modified local direct mechanisms.

Many previous works in the implementation literature have used integer mechanisms, in which each agent announces any nonnegative integer as well as opinions about agents' utility indices. The size of the set of message profiles is infinite, although the side of the set of preference profiles is finite. Many other previous works have used modulo mechanisms, in which each agent announces any integer in the set $\{1,...,n\}$ as well as opinions about agents' utility indices.[1] The size of the set of message profiles is finite, but it tends to infinity as the number of agents increases. Abreu and Matsushima (1992) have constructed their own mechanisms, in which each agent makes *multiple* announcements about all agents' utility indices. The size of the set of message profiles is finite,[2] but it tends to infinity as the upper bound of the amount of fine decreases. In contrast, in our modified local direct mechanisms, the size of the set of message profiles is finite, is constant with respect to the number of agents, and is constant with respect to the upper bound of the amount of fine.

We also show that if a social choice function is virtually implementable in terms of stability via local direct mechanisms, it is also virtually implementable in terms of BR-stability via *direct* mechanisms where each agent makes a single announcement about *all* agents' utility indices. Direct mechanisms have no slack messages, the size of the set of message profiles is finite, is constant with respect to the upper bound of the amount of fine, but is larger than that in modified local direct mechanisms.

In the definitions of stability and BR-stability, only a single agent is permitted to change her message at one time. We will permit two or more agents to simultaneously change their messages and introduce the notions of *weak stability* and *weak BR-stability*. We, rather trivially, show that every social choice function is virtually implementable in terms of weak stability or weak BR-stability via local direct mechanisms, even if all agents prefer the allocation that maximizes player 1's utility to the desired outcome.

In most parts of this paper, we investigate implementation on the basis of the static definition of stability, because of its tractability. In the last part of the paper we will characterize this stability notion from the viewpoints of *adaptive dynamics*. We consider the situation in which agents *asynchronically* play the component game infinitely many times. We introduce naïve models of adaptive dynamics, and show that there exists the unique convergent behavior if and only if the stable message profile exists in the

---

[1] Modulo mechanisms have unwanted mixed equilibria.

[2] The Abreu-Matsushima mechanism can eliminate all pure and mixed unwanted equilibria. The wanted equilibrium is the only iteratively undominated message profile.

component game, where the stable message profile will be this convergent point.

The crucial feature of our adaptive dynamics is that learning is *not* well sophisticated in that, irrespective of the past history of play, each agent always chooses any best reply with positive probability. Because of this, it is straightforward that every Nash equilibrium that is *not* strict can be eliminated. In this respect, the present paper has some distance from the models that were intensively studied in the evolution and learning literature[3] such as fictitious play, Bayesian learning and regret based dynamics addressed by Hart and Mas-Colell (2000). In this literature, learning is sometimes more involved so that an agent can learn not to play particular best replies. This may prevent the honest message profile from being reachable from other message profiles. Hence, we may say that the less boundedly rational agents are, the more complicated mechanisms are needed for implementation.

The elimination of unwanted absorbing limit cycles is not trivial at all, even if we use our naïve models. Hence, we have to make a careful check on how implementation works. Especially, when multiple agents are unlikely to adjust behaviors at one time, i.e., either stability or BR-stability must be used, and when all agents prefer agent 1's dictatorship to the socially desirable outcome, the social choice function is not necessarily virtually implementable. On the other hand, when at least one agent exists who prefers the socially desirable outcome to agent 1's dictatorship, every social choice function is virtually implementable. This point is the technical highlight of this paper.

The present paper could also be put in perspective of robustness of implementation to non-optimal behaviors, which were intensively studied after the survey by Moore (1992). Related works are Eliaz (2000), Maniquet (1998), Seften and Yavas (1996), Cabrales (1999), and others.

The organization of the paper is as follows. Section 2 provides the model. Section 3 considers stability and weak stability. Section 4 considers BR-stability and weak BR-stability. Section 5 considers adaptive dynamics. Section 6 concludes.

---

[3] See, for example, Fudenberg and Levine (1998).

## 2. The Model

Let $N = \{1,\ldots,n\}$ denote the finite set of agents where $n \geq 3$. Let $\Gamma$ denote the finite set of pure alternatives. Let $A$ denote the set of all lotteries over $\Gamma$. Agent $i's$ utility is indexed by a parameter $\omega_i$. Let $\Omega_i$ denote the finite set of her utility indices where $|\Omega_i| \geq 3$. Agent $i's$ preference is given by a utility function $u_i : A \times \Omega_i \to R$, and satisfies the expected utility hypothesis. We assume that no element of $\Omega_i$ induces complete indifference over all lotteries, and that any pair of distinct elements of $\Omega_i$ induces distinct orderings over $A$. Hence, for every $i \in N$, there exists a function $\alpha_i : \Omega_i \to A$ such that for every $\omega_i \in \Omega_i$, $a = \alpha_i(\omega_i)$ is the single maximizer of $u_i(a,\omega_i)$ among the set $\{\alpha_i(\omega_i') \in A : \omega_i' \in \Omega_i\}$, i.e., $u_i(\alpha_i(\omega_i),\omega_i) > u_i(\alpha_i(\omega_i'),\omega_i)$ for all $\omega_i' \neq \omega_i$. An example of $\alpha_i$ is as follows. For every $\omega_i \in \Omega_i$, and every $k \in \{1,...,|\Gamma|\}$, let $\gamma_i(\omega_i,k) \in \Gamma$ denote the pure alternative that player $i$ prefers in the k-th place among the set $\Gamma$, i.e.,

$$u_i(\gamma_i(\omega_i,1),\omega_i) \geq u_i(\gamma_i(\omega_i,2),\omega_i) \geq \cdots \geq u_i(\gamma_i(\omega_i,|\Gamma|),\omega_i).$$

Let $\alpha_i(\omega_i)(\gamma_i(\omega_i,k)) \equiv \dfrac{|\Gamma| - k + 1}{\sum_{l=1}^{|\Gamma|} l}$. Note the above inequalities hold because the probability $\alpha_i(\omega_i)(\gamma_i(\omega_i,k))$ is decreasing with respect to $k \in \{1,...,|\Gamma|\}$ and any pair of distinct elements of $\Omega_i$ induces distinct orderings over $A$.[4]

We assume that utilities are *quasi-linear*. We denote by $u_i(a,\omega_i) + t_i$ the utility for agent $i$ of a lottery $a \in A$ and a side payment $t_i \in R$ when her utility index is given by $\omega_i$.[5] A preference profile is described by $\omega = (\omega_1,\ldots,\omega_n)$. Let $\Omega \equiv \prod_{i \in N} \Omega_i$ denote the set of preference profiles.

A *social choice function* $x : \Omega \to A$ maps from preference profiles to lotteries. The lottery $x(\omega) \in A$ with no transfers is regarded as the most desirable allocation when the preference profile is given by $\omega \in \Omega$. We shall fix an arbitrary social choice function $x : \Omega \to A$ and an arbitrary positive real number $\varepsilon \in (0,1)$ that is close to zero. We specify

---

[4] See the Lemma in Abreu and Matsushima (1992).

[5] Quasi-linearity is assumed for simplicity. All we need is that there exists a private good and each agent's utility is denoted $u_i(a,t_i,\omega_i)$, which is increasing with respect to $t_i$.

another social choice function $\hat{x}$ by

$$\hat{x}(\omega) = (1-\varepsilon)x(\omega) + \frac{\varepsilon}{n}\sum_{i\in N}\alpha_i(\omega_i) \text{ for all } \omega \in \Omega .$$

Note that $\hat{x}$ is $\varepsilon$*-close to* $x$, that is, for every $\omega \in \Omega$, $x(\omega)\in A$ is $\varepsilon$*-close to* $\hat{x}(\omega)\in A$ in that the distance between them is at most $\varepsilon$ in the Euclidean metric. We call $\hat{x}$ the $\varepsilon$*-virtual social choice function* associated with $x$.

A *mechanism* is given by $G = (M_1,...,M_n; g,t)$ where $M_i$ is the finite set of messages for agent $i$, $M = \prod_{i\in N} M_i$, $g : M \to A$ is an outcome function, $t = (t_1,\dots,t_n)$, $t_i: M \to R$ is a side payment function to agent $i$, and $t$ is *budget-balancing* in that $\sum_{i\in N} t_i(m) = 0$ for all $m\in M$. When agents announce a message profile $m\in M$, the mechanism enforces the lottery $g(m)\in A$ and each agent $i\in N$ receives the side payment $t_i(m)\in R$. An arbitrary *message rule* is denoted by $\mu = (\mu_i)_{i\in N}$ where $\mu_i : \Omega \to M_i$ for all $i\in N$.[6]

Throughout the paper, we shall fix an arbitrary positive real number $\xi > 0$. We confine our attention to mechanisms $G$ satisfying that for every $i\in N$, and every $m\in M$,

$$t_i(m) \geq -\xi .$$

This implies that there exists an upper bound to the amount of fine such that each agent is able to pay up to this amount. A mechanism and a preference profile define a *game* $(G,\omega)$.

[6] The introduction of mixed message rules will not change any results of the paper.

## 3. Stable Implementation

### 3.1. Stability

A finite sequence $(i^k, m^k)_{k=1}^K \in (N \times M)^K$ is said to be *connected with* a message profile $m \in M$ if

$$m_i^k = m_i^{k-1} \text{ for all } k \in \{1,...,K\} \text{ and all } i \neq i^k,$$

where $m^0 = m$. A message profile $m' \in M$ is said to be *reachable from* another message profile $m \in M$ in $(G,\omega)$ if there exists $(i^k, m^k)_{k=1}^K$ connected with $m$ such that $m^K = m'$, and for every $k \in \{1,...,K\}$,

$$u_{i^k}(g(m^k), \omega_{i^k}) + t_{i^k}(m^k) \geq u_{i^k}(g(m^{k-1}), \omega_{i^k}) + t_{i^k}(m^{k-1}). \quad (1)$$

Hence, $m'$ is reachable from $m$ if and only if $m$ can be switched to $m'$ through finite steps, in each step of which, an agent unilaterally changes her message into any message that is better than or indifferent to that announced in the last step.

**Definition 1:** A message profile $m \in M$ is *stable* in $(G,\omega)$ if $m$ is reachable from every $m' \neq m$ in $(G,\omega)$ and no $m' \neq m$ is reachable from $m$ in $(G,\omega)$.

Note that $m$ is stable if and only if $m$ is the unique strict Nash equilibrium and there exists no other absorbing limit cycle in terms of reachability in that there exists no nonempty subset $\overline{M} \subset M$ such that $\overline{M} \neq \{m\}$, every pair of message profiles in $\overline{M}$ are mutually reachable, and there exists no $m' \notin \overline{M}$ that is reachable from some $m'' \in \overline{M}$.

### 3.2. Possibility Theorem

We specify a mechanism $G^* = (M, g^*, t^*)$ named a *local direct mechanism* as follows. Let $M_i = \Omega_{i-1} \times \Omega_i \times \Omega_{i+1}$ for all $i \in N$.[7] We denote $m_i = (m_{i,i-1}, m_{i,i}, m_{i,i+1}) \in M_i$, where $m_{i,j} \in \Omega_j$. Each agent $i \in N$ makes a single announcement about her own and neighbors' utility indices.[8] The *honest message rule* in $G^*$ is denoted by $\mu^* = (\mu_i^*)_{i \in N}$ where for every $i \in N$, and every $\omega \in \Omega$,

$$\mu_i^*(\omega) = (\omega_{i-1}, \omega_i, \omega_{i+1}).$$

Let $\tilde{M} \subset M$ denote the set of message profiles $m$ satisfying that there exists $\tilde{\omega}(m) \in \Omega$ such that for every $i \in N$,

$$m_{j,i} = \tilde{\omega}_i(m) \text{ for at least two agents } j \in \{i-1, i, i+1\}.$$

We shall call $\tilde{\omega}_i(m) \in \Omega_i$ and $\tilde{\omega}(m) \in \Omega$ *the public opinion for agent* $i$ and *the profile of the public opinions*, respectively. We specify $g^*$ by

---

[7] We denote $i-1=n$ if $i=1$, and $i+1=1$ if $i=n$.

[8] Matsushima (1988) investigated a variant of modulo mechanisms where each agent announces about her own and her neighbors' utility indices.

$$g^*(m) = (1-\varepsilon)x(\tilde{\omega}(m)) + \frac{\varepsilon}{n}\sum_{i\in N}\alpha_i(m_{i,i}) \text{ if } m \in \tilde{M},$$

and

$$g^*(m) = d(m_{1,1}) \text{ if } m \notin \tilde{M},$$

where $\alpha_i : \Omega_i \to A$ was the function introduced in Section 2, and $d : \Omega_1 \to A$ is a function satisfying that for every $\omega_1 \in \Omega_1$,

$$u_1(d(\omega_1),\omega_1) \geq u_1(g^*(m'),\omega_1) \text{ for all } m' \in M.$$

An example of $d(\omega_1) \in A$ is the lottery that maximizes agent $1's$ utility.

Suppose that there exists the profile of the public opinions $\tilde{\omega}(m) \in \Omega$, i.e., $m \in \tilde{M}$. Then, with probability $1-\varepsilon$, the mechanism enforces the lottery that the original social choice function $x$ assigns to $\tilde{\omega}(m)$, i.e., $x(\tilde{\omega}(m))$. With probability $\frac{\varepsilon}{n}$, each agent $i \in N$ becomes dictatorial and chooses the lottery $\alpha_i(m_{i,i}) \in A$. As Matsushima (1988) and Abreu and Sen (1990) have pointed out, the virtualness serves to provide each agent with the incentive to make the honest announcement about her own utility index. Next, suppose that there does not exist the profile of the public opinions, i.e., $m \notin \tilde{M}$. Then, agent 1 certainly becomes dictatorial and chooses the lottery $d(m_{1,1}) \in A$.

We specify $t^*$ by

$$t_i^*(m) = -\xi \text{ and } t_{i+2}^*(m) = \xi \text{ if } m \in \tilde{M},\ m_{i,j} \neq \tilde{\omega}_j(m) \text{ for some}$$
$$j \in \{i-1, i+1\}, \text{ and } m_{i'} = (\tilde{\omega}_{i'-1}(m), \tilde{\omega}_{i'}(m), \tilde{\omega}_{i'+1}(m)) \text{ for all } i' \neq i,$$

and

$$t_i^*(m) = 0 \text{ and } t_{i+2}^*(m) = 0 \text{ otherwise.}$$

Each agent $i \in N$ is fined if and only if there exists the profile of the public opinions $\tilde{\omega}(m)$, agent $i$ is the single deviant from $\tilde{\omega}(m)$, and she does not announce the public opinions about her neighbors' utility indices. If agent $i$ is fined, only agent $i+2$ is rewarded. The possibility of a player's being fined, together with the virtualness, serves to provide each agent with the strict incentive to announce the honest message when the others announce the honest messages.

Note

$$g^*(\mu^*(\omega)) = \hat{x}(\omega), \text{ and } t_i^*(\mu^*(\omega)) = 0 \text{ for all } i \in N.$$

When agents announce the honest message profile $\mu^*(\omega)$, the mechanism enforces the lottery that the $\varepsilon$-virtual social choice function $\hat{x}$ assigns to the true preference profile.

The following theorem provides a sufficient condition under which the mechanism $G^*$ and the honest message rule $\mu^*$ implement the $\varepsilon$-virtual social choice function $\hat{x}$ in terms of stability in that for every $\omega \in \Omega$, the honest message profile $\mu^*(\omega)$ is stable in $(G^*, \omega)$.

**Theorem 1:** *For every $\omega \in \Omega$, if there exists $j \neq 1$ such that*

$$u_j(\hat{x}(\omega),\omega_j) \geq u_j(d(\omega_1),\omega_j) - \xi, \tag{2}$$

*then, $\mu^*(\omega)$ is stable in $(G^*, \omega)$.*

Since Theorem 1 holds for every $\varepsilon > 0$, we can say that $x$ is virtually implementable in terms of stability via local direct mechanisms. That is, for every $\varepsilon > 0$, there exists an $\varepsilon$-virtual social choice function that is implemented by the local direct mechanisms $(G^*, \mu^*)$ in terms of stability, if for every $\omega \in \Omega$ there exists $j \neq 1$ such that

$$u_j(x(\omega), \omega_j) > u_j(d(\omega_1), \omega_j) - \xi .$$

Note that $x$ is virtually implementable in terms of stability via local direct mechanisms if

$$\xi > \max_{\omega \in \Omega} \min_{j \neq 1} \{u_j(d(\omega_1), \omega_j) - u_j(x(\omega), \omega_j)\} .$$

Hence, it follows that if $\xi > 0$ is large enough to satisfy that for every $a \in A$, and every $\omega \in \Omega$, there exists $j \neq 1$ such that $u_j(a, \omega_i) \geq u_j(d(\omega_1), \omega_j) - \xi$, then every social choice function is virtually implementable in terms of stability via local direct mechanisms. The following corollary is straightforward from Theorem 1.

**Corollary 2:** *For every $\omega \in \Omega$, if there exists $j \neq 1$ such that*

$$u_j(\hat{x}(\omega), \omega_j) \geq u_j(d(\omega_1), \omega_j), \tag{3}$$

*then, it holds, irrespective of $\xi > 0$, that $\mu^*(\omega)$ is stable in $(G^*, \omega)$.*

From Corollary 2 it follows that, irrespective of the upper bound of the amount of fine $\xi > 0$, a social choice function $x$ is virtually implementable in terms of stability via local direct mechanisms, if for every $\omega \in \Omega$, there exists $j \neq 1$ such that

$$u_j(x(\omega), \omega_j) \geq u_j(d(\omega_1), \omega_j). \tag{4}$$

A social choice function $x$ is said to be *efficient* if for every $\omega \in \Omega$, $x(\omega)$ is efficient in that there exists no $a \in A$ such that $u_i(a, \omega_i) \geq u_i(x(\omega), \omega_i)$ for all $i \in N$ and the strict inequality holds for some $i \in N$. Note that every efficient social choice function $x$ satisfies inequalities (4) for all $\omega \in \Omega$. Hence, it follows that, irrespective of the upper bound of the amount of fine $\xi > 0$, every efficient social choice function is virtually implementable in terms of stability via local direct mechanisms.

The complete proof of Theorem 1 will be shown in Appendix A. The next subsection provides an example for understanding the logical core of the proof.

## 3.3. Example I

This subsection assumes that $N = \{1,2,3\}$, and that the set of utility indices for each agent $i \in N$ consists of three distinct elements. Let $\Omega_i = \{\omega_i, \omega_i', \omega_i''\}$. Suppose that $\omega = (\omega_1, \omega_2, \omega_3)$ is the true preference profile, and agent $j = 2$ satisfies inequality (2). Let a message profile $m \in \tilde{M}$ be specified by $m_i = (\omega_{i-1}', \omega_i', \omega_{i+1}')$ for all $i \in N$. Note that the profile of the public opinions is given by $\tilde{\omega}(m) = \omega' = (\omega_1', \omega_2', \omega_3') \neq \omega$. Hence, the profile of the public opinions is different from the true preference profile. We will show that $\mu^*(\omega)$ is reachable from $m$ in $(G^*, \omega)$.

We specify a sequence $(i^k, m^k)_{k=1}^{8}$, connected with $m$, by

$i^1 = 1$, $m_1^1 = (\omega_3', \omega_1, \omega_2')$, $i^2 = 2$, $m_2^2 = (\omega_1', \omega_2, \omega_3')$,
$i^3 = 3$, $m_3^3 = (\omega_2', \omega_3, \omega_1')$, $i^4 = 1$, $m_1^4 = (\omega_3'', \omega_1, \omega_2'')$,
$i^5 = 2$, $m_2^5 = (\omega_1'', \omega_2, \omega_3')$, $i^6 = 1$, $m_1^6 = (\omega_3, \omega_1, \omega_2)$,
$i^7 = 2$, $m_2^7 = (\omega_1, \omega_2, \omega_3)$, $i^8 = 3$, and $m_3^8 = (\omega_2, \omega_3, \omega_1)$.

Note $m^8 = \mu^*(\omega)$. We can show that $m^3$ is reachable from $m^0$, because the virtualness provides each agent with the incentive to make the honest announcement about her own utility index. Note $i^1 = 1$, $m^1 \in \tilde{M}$, $\tilde{\omega}(m^1) = \omega'$,

$$u_{i^1}(g^*(m^1), \omega_{i^1}) + t_{i^1}^*(m^1) = u_1(\hat{x}(\omega') + \frac{\varepsilon}{3}\{\alpha_1(\omega_1) - \alpha_1(\omega_1')\}, \omega_1), \text{ and}$$

$$u_{i^1}(g^*(m^0), \omega_{i^1}) + t_{i^1}^*(m^0) = u_1(\hat{x}(\omega'), \omega_1).$$

Hence, it follows from the definition of $\alpha_1(\cdot)$ that inequality (1) holds for $k = 1$, and therefore, $m^1$ is reachable from $m^0 = m$. In the same way, the definition of $\alpha_2(\cdot)$ ($\alpha_3(\cdot)$) implies that $m^2$ is reachable from $m^1$ ($m^3$ is reachable from $m^2$, respectively).

We can show that $m^4$ is reachable from $m^3$, because agent 1 prefers the allocation that maximizes player $1's$ utility to the allocation chosen according to the profile of the public opinions. Note $i^4 = 1$, $m^4 \notin \tilde{M}$, $m_{1,1}^4 = \omega_1$,

$$u_{i^4}(g^*(m^4), \omega_{i^4}) + t_{i^4}^*(m^4) = u_1(d(\omega_1), \omega_1), \text{ and}$$

$$u_{i^4}(g^*(m^3), \omega_{i^4}) + t_{i^4}^*(m^3) = u_1(\hat{x}(\omega') + \frac{\varepsilon}{3}\sum_{i=1}^{3}\{\alpha_i(\omega_i) - \alpha_i(\omega_i')\}, \omega_1).$$

Hence, the definition of $d(\cdot)$ implies inequality (1) for $k = 4$, and therefore, $m^4$ is reachable from $m^3$.

We can show, as the key heuristic of the paper, that $m^6$ is reachable from $m^4$, because each agent has a wide variety of messages satisfying that whichever message she announces, the resultant allocation never changes. Note $i^5 = 2$, $m^5 \notin \tilde{M}$, $m_{1,1}^5 = \omega_1$,

$$u_{i^5}(g^*(m^5), \omega_{i^5}) + t_{i^5}^*(m^5) = u_2(d(\omega_1), \omega_2), \text{ and}$$

$$u_{i^5}(g^*(m^4), \omega_{i^5}) + t_{i^5}^*(m^4) = u_2(d(\omega_1), \omega_2).$$

Hence, inequality (1) holds for $k = 5$ with equality, and therefore, $m^5$ is reachable from $m^4$. In the same way, $m^6$ is reachable from $m^5$.

Note that $m^4$ and $m^5$ are weak Nash equilibria if for every $i \neq 1$, and every $\overline{\omega} \neq \omega$,

$$u_i(d(\omega_1), \omega_i) \geq u_i(\hat{x}(\overline{\omega}), \omega_i).$$

We can show that $m^7$ is reachable from $m^6$, because inequality (2) for $j = 2$ implies that agent $i^7 = 2$ prefers the allocation chosen according to the profile of the public opinions to the allocation that maximizes agent $1's$ utility. Since $i^7 = 2$, $m^7 \in \tilde{M}$, $\tilde{\omega}(m^7) = \omega$, agent 3 is the single deviant from $\tilde{\omega}(m^7)$, and $(m_{3,1}^7, m_{3,2}^7) \neq (\tilde{\omega}_1(m^7), \tilde{\omega}_2(m^7))$, it follows that

$$u_{i^7}(g^*(m^7), \omega_{i^7}) + t_{i^7}^*(m^7) = u_2(\hat{x}(\omega), \omega_2) + \xi, \text{ and}$$

$$u_{i^7}(g^*(m^6), \omega_{i^7}) + t_{i^7}^*(m^6) = u_2(d(\omega_1), \omega_1).$$

These inequalities, together with inequality (2) for $j = 2$, imply inequality (1) for $k = 7$, and therefore, $m^7$ is reachable from $m^6$.

In the mechanism $G^*$, the fine that agent $j - 2$ pays is always transferred to agent $j$. Agents $j$ and $j - 2$ have the same neighbor, i.e., agent $j - 1$, and announce opinions about this agent's utility index. This property is crucial in showing that the message profile can move from the set $M / \tilde{M}$ to the set $\tilde{M}$.

Note $i^8 = 3$, $m^8 \in \tilde{M}$, $m^8 = \mu^*(\omega)$,

$$u_{i^8}(g^*(m^8), \omega_{i^8}) + t^*_{i^8}(m^8) = u_3(\hat{x}(\omega), \omega_3) \text{, and}$$

$$u_{i^8}(g^*(m^7), \omega_{i^8}) + t^*_{i^8}(m^7) = u_3(\hat{x}(\omega), \omega_3) - \xi \text{ .}$$

These imply inequality (1) for $k = 8$, and therefore, $m^8$ is reachable from $m^7$.

From these observations, we have proved that $m^8 = \mu^*(\omega)$ is reachable from $m$.

## 3.4. Discussions

We will show below that when there exist four or more agent, the local direct mechanism $G^*$ *cannot* virtually implement a social choice function $x$ in terms of either Nash equilibrium or undominated Nash equilibrium. Fix $\omega \in \Omega$ arbitrarily. Assume $x(\omega) \neq d(\omega_1)$.

First, consider a message profile $m \in M$ satisfying that $m_{11} = \omega_1$, and for every $i \in N/\{1\}$, there exists $j \in \{i+2,...,i-2\}$ such that

$$m_{j-1,j} \neq m_{j,j} \neq m_{j+1,j} \neq m_{j-1,j}.$$

Note that $m$ is a Nash equilibrium in $(G^*, \omega)$, and $g^*(m) = d(\omega_1) \neq x(\omega)$. This implies that $G^*$ cannot virtually implement $x$ in terms of Nash equilibrium.

Next, for every $i \in N$, consider a message $m_i \in M_i$ satisfying that $m_{i,i} = \omega_i$. Choose $m_{-i} \in M_{-i}$ satisfying that

$$m_j = (\tilde{\omega}_{j-1}(m), \tilde{\omega}_{j-1}(m), \tilde{\omega}_{j+1}(m)) \text{ for all } j \neq i, \text{ and}$$

$$(m_{i,i-1}, m_{i,i+1}) = (\tilde{\omega}_{i-1}(m), \tilde{\omega}_{i+1}(m)).$$

Note that

$$g^*(m) = (1-\varepsilon)x(\omega') + \frac{\varepsilon}{n}\{\alpha_i(\omega_i) + \sum_{j \neq i} \alpha_j(\omega'_j)\},\ t_i^*(m) = 0,$$

and for every $m'_i \neq m_i$,

$$g^*(m/m'_i) = (1-\varepsilon)x(\omega') + \frac{\varepsilon}{n}\{\alpha_i(m'_{i,i}) + \sum_{j \neq i} \alpha_j(\omega'_j)\},$$

$$t_i^*(m/m'_i) = -\xi \text{ if } (m'_{i,i-1}, m'_{i,i+1}) \neq (\tilde{\omega}_{i-1}(m), \tilde{\omega}_{i+1}(m)),$$

and

$$t_i^*(m/m'_i) = 0 \text{ if } (m'_{i,i-1}, m'_{i,i+1}) = (\tilde{\omega}_{i-1}(m), \tilde{\omega}_{i+1}(m)).$$

Hence, the definition of $\alpha_i$ implies that $m_i$ is the strict best reply to $m_{-i}$, and therefore, $m_i$ is undominated in $(G^*, \omega)$. From the above arguments, we have shown that every message profile $m' \in M$ satisfying that $m_{i,i} = \omega_i$ and $m_{i-1,i} \neq m_{i,i} \neq m_{i+1,i} \neq m_{i-1,i}$ for all $i \in N$, is an undominated Nash equilibrium in $(G^*, \omega)$. Hence, it follows that $G^*$ cannot virtually implement $x$ in terms of undominated Nash equilibrium.

We will also show that when there exist five or more agents, $G^*$ can virtually implement $x$ in terms of stability, simply by using sequences $(i^k, m^k)_{k=1}^K$ satisfying that $m_{i^k}^k$ is *undominated* for every $k \in \{1,...,K\}$. Assume that $\varepsilon$ is so small relative to $\xi$ that for every $i \in N$, every $\omega_i \in \Omega_i$, and every $\omega'_i \in \Omega_i$,

$$\frac{\varepsilon}{n}\{u_i(\alpha_i(\omega_i), \omega_i) - u_i(\alpha_i(\omega'_i), \omega_i)\} < \xi.$$

Fix $i \in N$, $\omega_i \in \Omega_i$, and $m_i \in M_i$ arbitrarily. Choose $m_{-i} \in M_{-i}$ satisfying that

$$m \in \tilde{M},$$

$$m_j = (\tilde{\omega}_{j-1}(m), \tilde{\omega}_j(m), \tilde{\omega}_{j+1}(m)) \text{ for all } j \neq i-2,$$

$$m_{i-2,i-3} \neq \tilde{\omega}_{i-3}(m)\text{, and } (m_{i-2,i-2}, m_{i-2,i-1}) = (\tilde{\omega}_{i-2}(m), \tilde{\omega}_{i-1}(m))\,.$$

Since $n \geq 5$, it follows $i-3 \notin \{i-1, i, i+1\}$, and therefore,

$$m / m_i' \in \tilde{M} \text{ for all } m_i' \in M_i\,.$$

Hence,

$$g^*(m) = (1-\varepsilon)x(\tilde{\omega}(m)) + \frac{\varepsilon}{n}\sum_{j \in N}\alpha_j(\tilde{\omega}_j(m)),\ t_i^*(m) = \xi\,,$$

whereas for every $m_i' \neq m_i$,

$$g^*(m) = (1-\varepsilon)x(\tilde{\omega}(m)) + \frac{\varepsilon}{n}\{\alpha_i(m_{i,i}') + \sum_{j \neq i}\alpha_j(\tilde{\omega}_j(m))\}\text{, and } t_i^*(m) = 0\,.$$

These equalities, together with the definition of $\alpha_j$, imply that

$$u_i(g(m), \omega_{i^k}) + t_i(m) > u_i(g(m/m_i'), \omega_i) + t_{i^k}(m/m_i') \text{ for all } m_i' \neq m_i\,,$$

and therefore, $m_i$ is the strict best reply to $m_{-i}$. Hence, we have shown that every message is undominated, and therefore, $G^*$ can virtually implement $x$ by using sequences $(i^k, m^k)_{k=1}^K$ satisfying that $m_{i^k}^k$ is undominated for every $k \in \{1,...,K\}$.

## 3.5. Weak Stability

A message profile $m' \in M$ is said to be *weakly reachable from* another message profile $m \in M$ in $(G, \omega)$ if there exists a sequence of message profiles $(m^k)_{k=1}^K \in M^K$ such that $m^K = m'$, and for every $k \in \{1,...,K\}$, and every $i \in N$,

$$u_i(g(m^{k-1}/m_i^k), \omega_i) + t_i(m^{k-1}/m_i^k) \geq u_i(g(m^{k-1}), \omega_i) + t_i(m^{k-1})\,, \qquad (5)$$

where $m^0 = m$. Hence, $m'$ is weakly reachable from $m$ if and only if $m$ can be switched to $m'$ through finite steps, in each step of which, multiple agents simultaneously change their messages to messages better than or indifferent to the messages announced in the last step, provided that each agent expects the others to continue announcing the same messages. Note that if $m'$ is reachable from $m$, $m'$ is weakly reachable from $m$.

In each step, every agent who can change her message may misperceive what the others will announce. She expects the others to continue announcing the same messages, but there may exist other agents who will also change their messages at the same time. In contrast to weak reachability, reachability guarantees that an agent who can change her message will correctly perceive the others' announcements because there exist no other agents who can change their messages at the same time.

**Definition 2:** A message profile $m \in M$ is *weakly stable* in $(G, \omega)$ if $m$ is weakly reachable from every $m' \neq m$ in $(G, \omega)$ and no $m' \neq m$ is weakly reachable from $m$ in $(G, \omega)$.

Note that $m$ is weakly stable if and only if $m$ is the unique strict Nash equilibrium and there exists no other absorbing limit cycle in terms of weak reachability. Note that if $m$ is stable, it is weakly stable. The following theorem states that, without requiring the existence of $j \neq 1$ satisfying inequality (2), every social choice function is virtually implementable in terms of weak stability by using $(G^*, \mu^*)$.

**Theorem 3:** *For every* $\omega \in \Omega$, $\mu^*(\omega)$ *is weakly stable in* $(G^*, \omega)$.

Consider Example I and think about $m^5 \notin \widetilde{M}$. Each agent makes the honest announcement about her own utility index, and she and her two neighbors announce different opinions about her own utility index. Even if each agent $i \in N$ unilaterally changes her message into the honest message, the resultant allocation never changes. This implies that $\mu^*(\omega)$ is weakly reachable from $m^5$ through the single step that all agents simultaneously change their messages into the truthful messages. In this step we do not require the existence $j \neq 1$ satisfying inequality (2). For the detailed proof, see Appendix B.

# 4. BR-Stable Implementation

## 4.1. BR-Stability

A message profile $m' \in M$ is said to be *BR-reachable from* another message profile $m \in M$ in $(G,\omega)$ if there exists $(i^k, m^k)_{k=1}^K$ connected with $m$ such that $m^K = m'$, and, for every $k = 1,\ldots,K$, and every $m''_{i^k} \in M_{i^k}$,

$$u_{i^k}(g(m^k),\omega_{i^k}) + t_{i^k}(m^k) \geq u_{i^k}(g(m^k / m''_{i^k}),\omega_{i^k}) + t_{i^k}(m^k / m''_{i^k}) \qquad (6)$$

Note that $m'$ is BR-reachable from $m$ if and only if $m$ can be switched to $m'$ through finite steps, in each step of which, an agent unilaterally changes her message to *the best reply* to the profile in the last step. If $m'$ is BR-reachable from $m$, $m'$ is reachable from $m$.

**Definition 3:** A message profile $m \in M$ is *BR-stable* in $(G,\omega)$ if $m$ is BR-reachable from every $m' \neq m$ in $(G,\omega)$ and no $m' \neq m$ is BR-reachable from $m$ in $(G,\omega)$.

Note that $m$ is BR-stable if and only if $m$ is the unique strict Nash equilibrium and there exists no other absorbing limit cycle in terms of BR-reachability in that there exists no nonempty subset $\overline{M} \subset M$ such that $\overline{M} \neq \{m\}$, every pair of message profiles in $\overline{M}$ are mutually BR-reachable, and there exists no $m' \notin \overline{M}$ that is BR-reachable from some $m'' \in \overline{M}$. If $m$ is BR-stable, it is stable.

A stable message profile is not necessarily BR-stable. The following proposition provides a sufficient condition on $\omega \in \Omega$ under which the truthful message profile $\mu^*(\omega)$ is stable but not BR-stable in the game $(G^*,\omega)$.

**Proposition 4:** *Fix $\omega \in \Omega$ arbitrarily. Suppose that for every $i \in N$, and every $\omega' \in \Omega$,*

$$u_i(x(\omega'),\omega_i) > u_i(x(\omega),\omega_i) \text{ if } \omega'_j = \omega_j \text{ for all } j \notin \{i-1, i, i+1\} \text{ and}$$
$$(\omega'_{i-1},\omega'_i,\omega'_{i+1}) \neq (\omega_{i-1},\omega_i,\omega_{i+1}).$$

*Then, for every sufficiently small $\varepsilon > 0$, and every sufficiently small $\xi > 0$, the truthful message profile $\mu^*(\omega)$ is not BR-stable in $(G^*,\omega)$.*

Consider Example I and think about $m^7 \in \tilde{M}$. Let $\overline{m}_2 = (\omega'_1,\omega_2,\omega_3)$. Note that

$$g^*(m^7) = \hat{x}(\omega),\ t_2^*(m^7) = \xi,$$
$$g^*(m^7/\overline{m}_2) = \hat{x}(\omega'_1,\omega_2,\omega_3) + \frac{\varepsilon}{n}\{\alpha_1(\omega_1) - \alpha_1(\omega'_1)\}, \text{ and } t_2^*(m^7/\overline{m}_2) = 0.$$

Suppose that the supposition of Proposition 4 holds. Choose $\xi > 0$ so close to zero that

$$u_2(g^*(m^7/\overline{m}_2),\omega_2) + t_2^*(m^7/\overline{m}_2)$$
$$= u_2(\hat{x}(\omega'_1,\omega_2,\omega_3) + \frac{\varepsilon}{n}\{\alpha_1(\omega_1) - \alpha_1(\omega'_1)\},\omega_2)$$
$$> u_2(x(\omega),\omega_2) + \xi = u_2(g^*(m^7),\omega_2) + t_2^*(m^7).$$

Then, agent 2 prefers $\overline{m}_2$ to $m_2^7 = \mu_2^*(\omega)$ when the others announce $m_{-2}^6$, and therefore,

$m_2^7 = \mu_2^*(\omega)$ is not the best reply to $m^6$. Hence, inequalities (5) do not hold for $k = 7$, and therefore, $\mu^*(\omega)$ is not BR-reachable from $m$ by the sequence $(i^k, m^k)_{k=1}^8$. For the detailed proof, see Appendix B.

## 4.2. Possibility Theorem

We specify a mechanism $G^+ = (M, g^+, t^+)$ named *a modified local direct mechanism* as follows. Let

$$M_i = \Omega_{i-1} \times \Omega_i \times \Omega_{i+1} \times \{0,1\} \text{ for all } i \in N.$$

We denote $m_i = (m_{i,i-1}, m_{i,i}, m_{i,i+1}, m_{i,i+2}) \in M_i$ where

$$m_{i,j} \in \Omega_j \text{ for all } j \in \{i-1, i, i+1\}, \text{ and } m_{i,i+2} \in \{0,1\}.$$

Each agent announces not only an opinion about her and her two neighbors' utility indices but also either integer "0" or integer "1". The honest *message rule* in $G^+$ is denoted by $\mu^+ = (\mu_i^+)_{i \in N}$, where for every $i \in N$, and every $\omega \in \Omega$,

$$\mu_i^+(\omega) = (\omega_{i-1}, \omega_i, \omega_{i+1}, 0).$$

Each agent never announces integer "1" as long as she announces the honest message. We redefine $\tilde{M} \subset M$ by the set of message profiles $m$ satisfying that there exists the profile of the public opinions $\tilde{\omega}(m) \in \Omega$ such that for every $i \in N$,

$$m_{j,i} = \tilde{\omega}_i(m) \text{ for all } j \in \{i-1, i, i+1\} \text{ if } m_{i-2,i} = 1,$$

and

$$m_{j,i} = \tilde{\omega}_i(m) \text{ for at least two agents } j \in \{i-1, i, i+1\} \text{ if } m_{i-2,i} = 0.$$

The definition of the public opinions $\tilde{\omega}_i(m)$ for agent $i$ in $G^+$ differs from that in $G^*$ in that it must be announced by all three agents whenever agent $i-2$ announces integer "1".

Given this modification, we specify $g^+$ by

$$g^+(m) = (1-\varepsilon)x(\tilde{\omega}(m)) + \frac{\varepsilon}{n}\sum_{i \in N}\alpha_i(m_{i,i}) \text{ if } m \in \tilde{M},$$

$$g^+(m) = d(m_{1,1}) \text{ if } m \notin \tilde{M},$$

and specify $t^+$ by

$$t_i^+(m) = -\xi \quad \text{and} \quad t_{i+3}^+(m) = \xi \quad \text{if} \quad m \in \tilde{M},$$

$$(m_{i,i-1}, m_{i,i+1}, m_{i,i+2}) \neq (\tilde{\omega}_{i-1}(m), \tilde{\omega}_{i+1}(m), 0),$$

$$\text{and } m_{i'} = (\tilde{\omega}_{i'-1}(m), \tilde{\omega}_{i'}(m), \tilde{\omega}_{i'+1}(m), 0) \text{ for all } i' \neq i,$$

and

$$t_i^+(m) = 0 \text{ and } t_{i+3}^+(m) = 0 \text{ otherwise.}$$

Each agent $i \in N$ is fined if and only if there exists the profile of the public opinions $\tilde{\omega}(m)$, she is the single deviant from $\tilde{\omega}(m)$ and either announces integer "1" or does not announce the public opinions for her neighbors. If she is fined, only agent $i+3$ is rewarded. Note that

$$g^+(\mu^+(\omega)) = \hat{x}(\omega), \text{ and } t_i^+(\mu^+(\omega)) = 0 \text{ for all } i \in N.$$

Hence, the mechanism enforces the allocation assigned by $\hat{x}$ to the true preference profile when agents announce the honest message profile. The following theorem provides a sufficient condition under which $(G^+,\mu^+)$ implements $\hat{x}$ in terms of BR-stability.

**Theorem 5:** *Suppose that* $n \geq 4$. *For every* $\omega \in \Omega$, *if there exists* $j \neq 1$ *such that inequality (2) holds, then,* $\mu^+(\omega)$ *is BR-stable in* $(G^+,\omega)$.

Hence, given $n \geq 4$, a social choice function $x$ is virtually implementable in terms of BR-stability by $(G^+,\mu^+)$ under the same condition as in terms of stability via modified local direct mechanisms. The following corollary is straightforward from Theorem 5.

**Corollary 6:** *For every* $\omega \in \Omega$, *if* $n \geq 4$ *and there exists* $j \neq 1$ *satisfying inequality (3), then, it holds, irrespective of* $\xi > 0$, *that* $\mu^+(\omega)$ *is BR-stable in* $(G^+,\omega)$.

Hence, given $n \geq 4$, irrespective of the upper bound of the amount of fine $\xi > 0$, a social choice function $x$ is virtually implementable in terms of BR-stability via modified local direct mechanisms under the same condition as in terms of stability via local direct mechanisms.

The complete proof of Theorem 5 will be provided in Appendix C. The next subsection provides an example for understanding the logical core of the proof.

## 4.3. Example II

This subsection assumes that $N = \{1,2,3,4\}$, and that the set of utility indices for each agent $i \in N$ consists of three distinct elements. Let $\Omega_i = \{\omega_i, \omega_i', \omega_i''\}$. Suppose that $\omega = (\omega_1, \omega_2, \omega_3)$ is the true preference profile, and that agent $j = 2$ satisfies inequality (2). Let a message profile $m \in \tilde{M}$ be specified by

$$m_i = (\omega_{i-1}', \omega_i', \omega_{i+1}', 0) \text{ for all } i = 1,2,3.$$

Note that the profile of the public opinions is given by

$$\tilde{\omega}(m) = \omega' = (\omega_1', \omega_2', \omega_3') \neq \omega .$$

Hence, the profile of the public opinions is different from the true preference profile. We will show that the honest message profile $\mu^+(\omega)$ is BR-reachable from $m$ in $(G^+,\omega)$.

We specify $(i^k, m^k)_{k=1}^9$, connected with $m$, by

$$i^1 = 2,\ m_2^1 = (\omega_1', \omega_2, \omega_3', 0),\ i^2 = 1,\ m_1^2 = (\omega_4'', \omega_1, \omega_2'', 1),$$
$$i^3 = 4,\ m_4^3 = (\omega_3'', \omega_4, \omega_1'', 1),\ i^4 = 3,\ m_3^4 = (\omega_2', \omega_3, \omega_4', 1),$$
$$i^5 = 2,\ m_2^5 = (\omega_1', \omega_2, \omega_3', 1),\ i^6 = 4,\ m_4^6 = (\omega_3, \omega_4, \omega_1, 0),$$
$$i^7 = 1,\ m_1^7 = (\omega_4, \omega_1, \omega_2, 0),\ i^8 = 2,\ m_2^8 = (\omega_1, \omega_2, \omega_3, 0),$$
$$i^9 = 3, \text{ and } m_3^9 = (\omega_2, \omega_3, \omega_4, 0).$$

Note $m^9 = \mu^+(\omega)$. We can show that $m^1$ is BR-reachable from $m$, because the virtualness provides each agent with the incentive to make the honest announcement about her own utility index. Note that for every $m_2' \in M_2$,

$$m^1/m_2' \in \tilde{M}, \ \tilde{\omega}(m^1/m_2') = \omega',$$

$$u_2(g^+(m^1/m_2'), \omega_2) + t_2^+(m^1/m_2') \le u_2(\hat{x}(\omega') + \frac{\varepsilon}{4}\{\alpha_2(m_{2,2}) - \alpha_2(\omega_2')\}, \omega_2),$$

and

$$u_2(g^+(m^1), \omega_2) + t_2^+(m^1) = u_2(\hat{x}(\omega') + \frac{\varepsilon}{4}\{\alpha_2(\omega_2) - \alpha_2(\omega_2')\}, \omega_2).$$

Hence, it follows from $i^1 = 2$ and the definition of $\alpha_2(\cdot)$ that for every $m_2' \in M_2$,

$$u_2(g^+(m^1), \omega_2) + t_2^+(m^1) \ge u_2(g^+(m^1/m_2'), \omega_2) + t_2^+(m^1/m_2'),$$

i.e., inequalities (6) hold for $k = 1$, and therefore, $m^1$ is BR-reachable from $m$.

We can show that $m^2$ is BR-reachable from $m^1$, because agent 1 prefers the allocation that maximizes her utility to the allocation chosen according to the profile of the public opinions. Note $m^2 \notin \tilde{M}$, $m_{1,1}^2 = \omega_1$, and $t_1^+(m^2/m_1) = 0$ for all $m_1 \in M_1$. Hence, for every $m_1 \in M_1$,

$$u_1(g^+(m^2), \omega_1) + t_1^+(m^1) = u_1(d(\omega_1), \omega_1)$$

$$\ge u_1(g^+(m^2/m_1), \omega_1) = u_1(g^+(m^2/m_1), \omega_1) + t_1^+(m^1/m_1),$$

which, together with $i^2 = 1$, implies inequalities (6) for $k = 2$, and therefore, $m^2$ is BR-reachable from $m^1$.

We can show, as the key heuristic of the paper, that $m^7$ is BR-reachable from $m^2$, because whichever message an agent announces, the resultant allocation never changes. Note that for every $m_4 \in M_4$, $m^3/m_4 \notin \tilde{M}$, and therefore,

$$u_4(g^+(m^3/m_4), \omega_4) + t_4^*(m^3/m_4) = u_4(d(\omega_1), \omega_4).$$

This, together with $i^3 = 4$, implies inequalities (6) for $k = 3$ with strict equality, and therefore, $m^3$ is BR-reachable from $m^2$. In the same way, $m^4$, $m^5$, $m^6$, and $m^7$ are BR-reachable from $m^3$, $m^4$, $m^5$, and $m^6$, respectively.

For every $k \in \{3,...,7\}$, every message for player $i^k$ is the best reply to $m^k$. For every $k \in \{3,...,6\}$, every agent is indifferent among *all* her messages. Moreover, for every $k \in \{2,...,7\}$, the message profile $m^k$ is a Nash equilibrium in $(G^+, \omega)$.

We can show that $m^8$ is BR-reachable from $m^7$, because inequality (2) for $j = 2$ implies that agent $i^8 = 2$ prefers the allocation chosen according to the profile of the public opinions to the allocation that maximizes agent 1's utility. Note that $m^8 \in \tilde{M}$, $\tilde{\omega}(m^8) = \omega$, agent 3 is the single deviant from $\mu^+(\omega)$, and $(m_{3,2}^7, m_{3,4}^7, m_{3,1}^7) \ne (\omega_2, \omega_4, 0)$. Hence, it follows that $t_2^+(m^8) = -t_3^+(m^8) = \xi$, and therefore,

$$u_2(g^+(m^8), \omega_2) + t_2^+(m^8) = u_2(\hat{x}(\omega), \omega_2) + \xi.$$

Note that for every $m_2 \ne \mu_2^+(\omega)$,

$$u_2(g^+(m^8/m_2), \omega_2) + t_2^+(m^8/m_2) = u_2(d(\omega_1), \omega_2) \text{ if } m^8/m_2 \notin \tilde{M},$$

and

$$u_2(g^+(m^8/m_2), \omega_2) + t_2^+(m^8/m_2) = u_2(\hat{x}(\omega) + \frac{\varepsilon}{4}\{\alpha_2(m_{2,2}) - \alpha_2(\omega_2)\}, \omega_2)$$

if

$$m^8 / m_2 \in \tilde{M} \ .$$

These inequalities, together with $i^8 = 2$ and inequality (2) for $j = 2$, imply inequalities (6) for $k = 8$, and therefore, $m^8$ is BR-reachable from $m^7$.

In the modified local direct mechanism $G^+$ the fine paid by agent $j-3$ is always transferred to agent $j$. Agent $j$ announces her opinion about agent $j-1's$ utility index. Agent $j-3$ does not announce her opinion about agent $j-1's$ utility index, but the announcement of either integer "0" or integer "1" by agent $j-3$ influences how to determine the public opinion for agent $j-1$. This property is crucial in showing that the message profile can move from the set $M / \tilde{M}$ to the set $\tilde{M}$ through the agent $j's$ best-reply behavior.

Finally, note $i^9 = 3$, $m^9 \in \tilde{M}$, $m^9 = \mu^+(\omega)$, and therefore,

$$u_3(g^+(m^9), \omega_3) + t_3^+(m^9) = u_3(\hat{x}(\omega), \omega_3) \ ,$$

whereas, for every $m_3 \in M_3$,

$$u_3(g^+(m^9 / m_3), \omega_3) + t_3^+(m^9 / m_3) \leq u_3(\hat{x}(\omega), \omega_3) \ .$$

Hence, inequalities (6) hold for $k = 9$, i.e., $m^9 = \mu^+(\omega)$ is BR-reachable from $m^8$.

From these observations, we have proved that $\mu^+(\omega)$ is BR-reachable from $m$.

## 4.4. Direct Mechanisms

This subsection shows that not only modified local direct mechanisms but also direct mechanisms virtually implement $x$ in terms of BR-stability. Suppose $n \geq 4$, and $|\Omega_i| \geq 4$ for all $i \in N$. Specify a mechanism $G^{**} = (M, g^{**}, t^{**})$ as follows. Let $M_i = \Omega_1 \times \cdots \times \Omega_n$ for all $i \in N$. We denote $m_i = (m_{i,1}, ..., m_{i,n}) \in M_i$ where $m_{i,j} \in \Omega_j$ for all $j \in N$. Each agent announces opinions about *all* agents' utility indices. The honest message rule in $G^{**}$ is denoted by $\mu^{**} = (\mu_i^{**})_{i \in N}$ where $\mu_i^{*}(\omega) = \omega$ for all $\omega \in \Omega$. There are *no* slack messages in $G^{**}$, but $|\Omega_i^{**}| > |\Omega_i^{+}|$ for all $i \in N$, i.e., the size of the set of message profiles in $G^{**}$ is larger than in $G^{+}$.

We redefine $\tilde{M} \subset M$ by the set of message profiles $m$ satisfying that there exists the profile of the public opinions $\tilde{\omega}(m) \in \Omega$ such that for every $i \in N$,

$$m_{j,i} = \tilde{\omega}_i(m) \text{ for } n-1 \text{ or more agents.}$$[9]

We specify $g^{**}$ by

$$g^{**}(m) = (1-\varepsilon)x(\tilde{\omega}(m)) + \frac{\varepsilon}{n}\sum_{i \in N}\alpha_i(m_{i,i}) \text{ if } m \in \tilde{M},$$

$$g^{**}(m) = d(m_{1,1}) \text{ if } m \notin \tilde{M},$$

and, we specify $t^{**}$ by

$$t_i^{**}(m) = -\xi \text{ and } t_{i+3}^{**}(m) = \xi \text{ if } m \in \tilde{M},\ m_{i,-i} \neq \tilde{\omega}_{-i}(m), \text{ and } m_{i'} = \tilde{\omega}(m) \text{ for all } i' \neq i,$$

and

$$t_i^{**}(m) = 0 \text{ and } t_{i+3}^{**}(m) = 0 \text{ otherwise.}$$

Note that

$$g^{**}(\mu^{**}(\omega)) = \hat{x}(\omega), \text{ and } t_i^{**}(\mu^{**}(\omega)) = 0 \text{ for all } i \in N.$$

The following theorem provides a sufficient condition under which $(G^{**}, \mu^{**})$ implements $\hat{x}$ in terms of BR-stability.

**Theorem 7:** *Suppose that $n \geq 4$, and $|\Omega_i| \geq 4$ for all $i \in N$. For every $\omega \in \Omega$, if there exists $j \neq 1$ such that inequality (2) holds, then, $\mu^{**}(\omega)$ is BR-stable in $(G^{**}, \omega)$.*

**Proof:** See Appendix D.

Hence, given $n \geq 4$ and $|\Omega_i| \geq 4$ for all $i \in N$, a social choice function $x$ is virtually implementable in terms of BR-stability by $(G^{**}, \mu^{**})$ under the same condition as by $(G^{+}, \mu^{+})$. The following corollary is straightforward from Theorem 7.

---

[9] In $G^{**}$, the public opinions are not determined according to either the majority rule or the unanimity rule.

**Corollary 8:** *For every* $\omega \in \Omega$ *, if* $n \geq 4$ *,* $|\Omega_i| \geq 4$ *for all* $i \in N$ *, and there exists* $j \neq 1$ *satisfying inequality (3), then, it holds irrespective of* $\xi > 0$ *that* $\mu^{**}(\omega)$ *is BR-stable in* $(G^{**}, \omega)$ .

Hence, given $n \geq 4$ and $|\Omega_i| \geq 4$ for all $i \in N$ , irrespective of the upper bound of the amount of fine $\xi > 0$ , a social choice function $x$ is virtually implementable in terms of BR-stability by $(G^{**}, \mu^{**})$ under the same condition as $(G^{+}, \mu^{+})$ .

## 4.5. Discussions

In the same way as local direct mechanisms, it follows that when $n \geq 4$ , either modified local direct mechanisms or direct mechanisms cannot virtually implement a social choice function $x$ in terms of either Nash equilibrium or undominated Nash equilibrium. In the same way as local direct mechanisms, it follows that when $n \geq 5$ , both direct mechanisms and direct mechanisms can virtually implement $x$ in terms of BR-stability by using sequences $(i^k, m^k)_{k=1}^{K}$ satisfying that $m_{i^k}^k$ is undominated for every $k \in \{1,...,K\}$ .

Either integer mechanisms or modulo mechanisms cannot virtually implement a social choice function in terms of BR-stability. In integer mechanisms, the wanted message profile is never BR-reachable from every message profile triggering off the integer game. In modulo mechanisms, the set of all message profiles triggering off the modulo game is an absorbing limit cycle in terms of BR-stability. On the other hand, the Abreu-Matsushima mechanisms can virtually implement it in terms of BR-stability.

We can construct other mechanisms that virtually implement a social choice function in terms of BR-stability. Assume $n \geq 5$ . Consider a class of mechanisms in which each agent $i \in N$ announces about her own and her four neighbors' utility indices, i.e., about $(\omega_{i-2},...,\omega_{i+2})$ . We can construct a mechanism within this class that, together with the associated honest message rule, implements the virtual social choice function $\hat{x}$ in terms of BR-stability, provided that there exists $j \neq 1$ satisfying inequality (2). Hence, for every social choice function with the existence of $j \neq 1$ satisfying inequality (2), there exist mechanisms that virtually implement it in terms of BR-stability such that there exist no slack messages, the size of the set of message profiles is finite and constant with respect to the number of agents as well as the upper bound of fine.

## 4.6. Weak BR-Stability

A message profile $m' \in M$ is said to be *weakly BR-reachable from* another message profile $m \in M$ in $(G,\omega)$ if there exists a sequence of message profiles $(m^k)_{k=1}^K$ such that $m^K = m'$, and for every $k = 1,\dots,K$, and every $i \in N$, either $m_i^k = m_i^{k-1}$, or

$$u_i(g(m^k),\omega_i) + t_i(m^k) \geq u_i(g(m^k/m_i''),\omega_i) + t_i(m^k/m_i'') \text{ for all } m_i'' \in M_i,$$

where $m^0 = m$. Hence, $m'$ is weakly BR-reachable from $m$ if and only if $m$ can be switched to $m'$ through finite steps, in each step of which, multiple agents simultaneously change their messages to their best reply to the message profile in the last step.

**Definition 4:** A message profile $m \in M$ is *weakly BR-stable* in $(G,\omega)$ if $m$ is weakly BR-reachable from every $m' \neq m$ in $(G,\omega)$ and no $m' \neq m$ is not weakly BR-reachable from $m$ in $(G,\omega)$.

Note that $m$ is weakly BR-stable if and only if $m$ is the unique strict Nash equilibrium and there exists no other absorbing limit cycle in terms of weak BR-reachability. Note that if $m$ is BR-stable, it is weakly BR-stable. The following theorem states that without requiring the existence of $j \neq 1$ satisfying inequality (2), every social choice function $x$ is virtually implementable in terms of weak BR-stability by using $(G^+,\mu^+)$.

**Theorem 9:** *If $n \geq 4$, then, for every $\omega \in \Omega$, $\mu^+(\omega)$ is weakly BR-stable in $(G^+,\omega)$.*

Consider Example II and think about $m^5 \notin \tilde{M}$. Each agent $i \in N$ makes the honest announcement about her own utility index, and she and her two neighbors announce different opinions about her own utility index. She also announces $m_{i,i+2}^5 = 1$, and therefore, the public opinion for agent $i+2$ must be announced by all relevant agents. Even if each agent $i \in N$ unilaterally changes her message into any message, the resultant allocation never changes. Hence, $\mu_i^+(\omega)$ is the best reply to $m^5$, and therefore, $\mu^+(\omega)$ is weakly BR-reachable from $m^5$ through the single step that all agents simultaneously change their messages into the honest messages. In this step we do not require the existence of $j \neq 1$ satisfying inequality (2). We do not use this assumption in the steps that $m^5$ is BR-reachable from $m$ also. Hence, we have shown without requiring the existence of $j \neq 1$ satisfying inequality (2) that $\mu^*(\omega)$ is weakly BR-reachable from $m$. For the detailed proof, see Appendix B. In the same way, we can show that given that $n \geq 4$, for every $\omega \in \Omega$, $\mu^{**}(\omega)$ is weakly BR-stable in $(G^{**},\omega)$. We can show the same result for the mechanism in Subsection 4.5. Hence, given that $n \geq 4$, every social choice function can be virtually implemented by either $G^+$, $G^{**}$ or the mechanism in Subsection 4.5, together with their respective honest message rules.

## 5. Adaptive Dynamics

This section characterizes stability and its variants from the dynamical viewpoint. Fix a mechanism $G$ and a preference profile $\omega \in \Omega$ arbitrarily. Consider situations in which agents are faced with the same game $(G,\omega)$ infinitely many times. At the end of every period $t = 1,2,...$, all agents can observe their announced message profile $m(t) \in M$. Agents' behavior is described by a naïve model of *adaptive dynamics* defined by $(m(0), p)$, where $m(0) \in M$, $p = (p_i)_{i\in N}$, $p_i = (p_i(\cdot \mid m))_{m\in M}$, and $p_i(\cdot \mid m): M_i \to [0,1]$ is a conditional probability function on $M_i$. Two different scenarios, i.e., *alternating play* and *simultaneous play*, are provided.

### 5.1. Alternating Play

This subsection assumes that at the beginning of each period, at most one agent can change the announcement. For every $t \geq 1$, we define $i(t) \in N$ by

$$t = vn + i \text{ for some nonnegative integer } v \geq 0,$$

that is, $i(1) = 1$, $i(2) = 2, \ldots, i(n+1) = 1, \ldots$, and so on.[10] In period 1, all agents except agent $i(1) = 1$ announce the message profile $m(0) \in M$. In every period $t \geq 1$, agent $i(t)$ announces any message $m_{i(t)} \in M_{i(t)}$ with probability $p_{i(t)}(m_{i(t)} \mid m(t-1))$, whereas any other agent $i \neq i(t)$ announces the message announced in the last period, i.e., $m_i(t-1)$.

**Definition 5:** A message profile $m \in M$ is *the long-run behavior in* $G$ *with alternating play with respect to* $p = (p_i)_{i\in N}$, if for every $m(0) \in M$, and every $\lambda \in (0,1]$, there exists a positive integer $\hat{T}$ such that on the assumption of alternating play, at least with probability $1-\lambda$, the model $(m(0), p)$ induces agents to announce $m(t) = m$ in every period $t \geq \hat{T}$.

Note that $m$ is the long-run behavior if and only if, irrespective of which message profile agents announce in period 1, agents come to continue announcing $m$ in the long run. We introduce a condition on $(\omega, p)$ as follows.

**Condition 1:** For every $i \in N$, $m_i \in M_i$, and every $m' \in M$,

$p_i(m_i \mid m') > 0$ if and only if $u_i(g(m'/m_i),\omega_i) + t_i(m'/m_i) \geq u_i(g(m'),\omega_i) + t_i(m')$.

Condition 1 implies that in every period $t \geq 1$, agent $i(t)$ announces any message with a positive probability if and only if this message is better than or indifferent to

---

[10] The specificity of $(i(t))_{t=1}^{\infty}$ may not be necessary. All we need is that for every $i \in N$, there exists an infinite increasing sequence of periods $(t(i,s))_{s=1}^{\infty}$ such that $i(t(i,s)) = i$ for all $s = 1,2,...$, i.e., each agent $i \in N$ has infinitely many chances to change her message.

$m_{i(t)}(t-1)$. The following proposition states that Condition 1 is a sufficient condition under which the stable message profile is characterized by the long-run behavior in $G$ with alternating play.

**Proposition 10:** *Suppose that* $(\omega, p)$ *satisfies Condition 1. Then, a message profile* $m \in M$ *is stable in* $(G, \omega)$ *if and only if it is the long-run behavior in* $G$ *with alternating play with respect to* $p$ *.*

**Proof**: See Appendix E.

We introduce another condition on $(\omega, p)$ as follows.

**Condition 2:** For every $i \in N$, $m_i \in M_i$, and every $m' \in M$,

$$p_i(m_i \mid m') > 0 \text{ if and only if either } m_i = m'_i \text{ or } u_i(g(m'/m_i), \omega_i) + t_i(m'/m_i) \geq u_i(g(m'/m''_i), \omega_i) + t_i(m'/m''_i) \text{ for all } m''_i \in M_i,$$

Condition 2 implies that in every period $t \geq 1$, agent $i(t)$ announces any message with positive probability if and only if this message is either equal to $m_{i(t)}(t-1)$ or the best reply to $m(t-1)$. The following proposition states that Condition 2 is a sufficient condition under which the BR-stable message profile is characterized by the long-run behavior in $G$ with alternating play. We can prove the proposition in the same way as Proposition 10, by replacing stability and inequality (1) by BR-stability and inequalities (6), respectively.

**Proposition 11:** *Suppose that* $(\omega, p)$ *satisfies Condition 2. Then, a message profile* $m \in M$ *is BR-stable in* $(G, \omega)$ *if and only if it is the long-run behavior in* $G$ *with alternating play with respect to* $p$ *.*

## 5.2. Simultaneous Play

This subsection assumes that at the beginning of each period, two or more agents may simultaneously change their announcements with positive probability. In every period $t \geq 1$, each agent $i \in N$ announces any message $m_i \in M_i$ with probability $p_i(m_i \mid m(t-1))$.

**Definition 6:** A message profile $m \in M$ is *the long-run behavior in* $G$ *with simultaneous play with respect to* $p = (p_i)_{i \in N}$ if for every $m(0) \in M$, and every $\lambda \in (0,1]$, there exists a positive integer $\hat{T}$ such that on the assumption of simultaneous play, at least with probability $1-\lambda$, the model $(m(1), p)$ induces agents to announce $m(t) = m$ in every period $t \geq \hat{T}$.

The following proposition states that Condition 1 is a sufficient condition under which the weakly stable message profile is characterized by the long-run behavior in $G$ with simultaneous play.

**Proposition 12:** *Suppose that* $(\omega, p)$ *satisfies Condition 1. Then, a message profile* $m \in M$ *is weakly stable in* $(G,\omega)$ *if and only if it is the long-run behavior in* $G$ *with simultaneous play with respect to* $p$ *.*

**Proof:** See Appendix E.

The following proposition states that Condition 2 is a sufficient condition under which the weakly BR-stable message profile is characterized by the long-run behavior in $G$ with simultaneous play. We can prove the proposition in the same way as Proposition 12, by replacing weak stability and inequalities (5) by weak BR-stability and inequalities (8), respectively.

**Proposition 13:** *Suppose that* $(\omega, p)$ *satisfies Condition 2. Then, a message profile* $m \in M$ *is weakly BR-stable in* $(G,\omega)$ *if and only if it is the long-run behavior in* $G$ *with simultaneous play with respect to* $p$ *.*

## 5.3. Discussions

We have assumed that each agent always has static expectations on the other agents' announcements. We can provide the same results when we replace this assumption by a weaker assumption that each agent has static expectations *only if* the others have announced the same messages for a long time. As a generalization of $(m(0), p)$, we define a model of adaptive dynamics by $(m(0), q)$ where $q = (q_i)_{i \in N}$, $h$ denotes a partial history of message profiles, $q_i = (q_i(\cdot \mid h))_{m \in M}$, and $q_i(\cdot \mid h) : M_i \to [0,1]$ is a conditional probability function on $M_i$. Similarly we define the long-run behaviors with respect to $q$. We introduce two conditions on $(\omega, q)$ as follows.

**Condition 3:** There exist a positive integer $T > 0$ and a positive real number $\rho > 0$ such that for every $i \in N$, every $m_i \in M_i$, every $t \geq T$, and every $h = (m(\tau))_{\tau=1}^{t}$,

$$q_i(m_i \mid h) \geq \rho \text{ if } q_i(m_i \mid h) > 0,$$

$$q_i(m_i \mid h) > 0 \text{ if } m_i(t) = m_i,$$

and if there exists $m' \in M$ such that $m(\tau) = m'$ for all $\tau \in \{t-T+1,...,t\}$, then

$$q_i(m_i \mid h) > 0 \text{ if and only if } u_i(g(m'/m_i), \omega_i) + t_i(m'/m_i) \geq u_i(g(m'), \omega_i) + t_i(m').$$

**Condition 4:** There exist a positive integer $T > 0$ and a positive real number $\rho > 0$ such that for every $i \in N$, every $m_i \in M_i$, every $t \geq T$, and every $h = (m(\tau))_{\tau=1}^{t}$,

$$q_i(m_i \mid h) \geq \rho \text{ if } q_i(m_i \mid h) > 0,$$

$$q_i(m_i \mid h) > 0 \text{ if } m_i(t) = m_i,$$

and if there exists $m' \in M$ such that $m'_i \neq m_i$ and $m(\tau) = m'$ for all $\tau \in \{t-T+1,...,t\}$, then

$$q_i(m_i \mid h) > 0 \text{ if } u_i(g(m'/m_i),\omega_i) + t_i(m'/m_i) \geq u_i(g(m'/m_i''),\omega_i) + t_i(m'/m_i'') \text{ for all } m_i'' \in M_i,$$

and

$$q_i(m_i \mid h) = 0 \text{ if } u_i(g(m'/m_i),\omega_i) + t_i(m'/m_i) < u_i(g(m'),\omega_i) + t_i(m').$$

Condition 3 holds if Condition 1 holds. Condition 4 holds if either Condition 2 or Condition 3 holds. Condition 4 is the weakest among these four conditions.

In the same way as in the "only if" part of Proposition 10, it follows that under Condition 3, the stable message profile is the long-run behavior with alternating play with respect to $q$. In the same way as in the "only if" part of Proposition 11, it follows that under Condition 4, the BR-stable message profile is the long-run behavior with alternating play with respect to $q$. In the same way as in the "only if" part of Proposition 12, it follows that under Condition 3, the weakly stable message profile is the long-run behavior with simultaneous play with respect to $q$. In the same way as in the "only if" part of Proposition 13, it follows that under Condition 4, the weakly BR-stable message profile is the long-run behavior with simultaneous play with respect to $q$. Hence, it follows that under Condition 3, the stable message profile is the long-run behavior with respect to $q$. Under Condition 4, the BR-stable message profile is the long-run behavior with respect to $q$ irrespective of whether with alternating play or with simultaneous play. Both alternating play and simultaneous play may be even unnecessary for the long-run convergence to the stable, or BR-stable, message profile. All we have to require is that each player has infinitely many chances to change her message.

From the above arguments, we can obtain the following convergence results. With the existence of $j \neq 1$ satisfying inequality (2), the honest message profile $\mu^*(\omega)$ is the long-run behavior in the local direct mechanism $G^*$ under Condition 3 whenever each agent has infinitely many chances to change her message. Under Condition 3, $\mu^*(\omega)$ is the long-run behavior in $G^*$ with simultaneous play. With the existence of $j \neq 1$ satisfying inequality (2), the honest message profile $\mu^+(\omega)$ is the long-run behavior in the modified local direct mechanism $G^+$ under Condition 4 whenever each agent has infinitely many chances to change her message. Under Condition 4, $\mu^+(\omega)$ is the long-run behavior in $G^+$ with simultaneous play. The same results hold for the direct mechanism $G^{**}$ and the mechanism in Subsection 4.5.

In modulo mechanisms, under Condition 2, the wanted message profile is always the long-run behavior with simultaneous play because it is weakly BR-stable. However, in modulo mechanisms, under Condition 2, the wanted message profile is never the long-run behavior with alternating play because it is not BR-stable. In integer mechanisms, under Condition 2, the wanted message profile is always the long-run behavior with simultaneous play because it is weakly BR-stable.

This point is related to Cabrales (1999), which retrieved the honor of integer mechanisms by showing that the wanted outcomes can be achieved by boundedly rational agents. However, in integer mechanisms, under Condition 2, the wanted message profile is never the long-run behavior with alternating play because it is not BR-stable. On the other hand, in Abreu-Matsushima mechanisms, the wanted message profile is always the long-run behavior with respect to $q$ under Condition 4 irrespective of whether with alternating play or with simultaneous play, because it is BR-stable. Cabrales (1999) pointed out that Abreu-Matsushima mechanisms have an instability property in that if agents are allowed to announce even worse messages, there exists a trade-off between close implementability and stability of the wanted message profile. In the mechanisms studied in the present paper there exists no such trade-off.

The mechanisms studied in the paper have a nice feature with respect to the speed of convergence. In the local direct mechanism $G^*$, the honest message profile $\mu^*(\omega)$ is always reachable from every message profile through at most $5n+2$ steps. This number of steps does not depend on the upper bound of the amount of fine. The same property holds for BR-reachability in the modified local direct mechanism, in the direct mechanism, and also in the mechanism addressed in Subsection 4.5. In contrast, in Abreu-Matsushima mechanisms, the number of steps crucially depends on the upper bound of the amount of fine. Indeed, this number tends to infinity as the upper bound of the amount of fine approaches zero, which implies that it takes very long time to converge to the wanted outcome in Abreu-Matsushima mechanisms relative to the mechanisms in the paper.

## 6. Concluding Remarks

This paper has investigated implementation of social choice functions in terms of stability and its variants. We have shown that with a minor restriction every social choice function is virtually implementable in terms of stability via local direct mechanisms, which are of particularly simple form. We have shown that it is virtually implementable in terms of BR-stability via slightly more complex but still very simple mechanisms such as modified local direct mechanisms and direct mechanisms. Moreover, if multiple players are allowed to simultaneously change their announcements, every social choice function is virtually implementable with no restrictions.

The paper could be regarded as an early attempt to provide new ideas of designing mechanisms with bounded rationality. The results of the paper depend on the assumption that models of adaptive dynamics are naïve so that irrespective of the past history of play, each agent always chooses any best reply with positive probability. This naïve assumption might be inappropriate in some case when agents are more sophisticated so that they may learn not to announce a particular message even if it is a best reply. Hence, a further attempt might be needed to investigate other definitions of stability on the basis of more general and careful dynamic refinement of equilibrium.

It is important to investigate stable implementation in the incomplete information environments. The paper depends on the assumption that an agent's utility index is known to other agents. This assumption guarantees that a local direct mechanism and its variants have slack message profiles. The results rely on the existence of such slack message profiles. In contrast, in the incomplete information environments, direct mechanisms have no slack message profiles, because agents have no jointly possessed information. Hence, a question in the incomplete information environments would be what is the minimum requirement of adding slack messages to direct mechanisms.

## References


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

Abreu, D. and A. Sen (1991): “Virtual Implementation in Nash Equilibrium,” *Econometrica* 59, 997-1021.

Cabrales, A. (1999): “Adaptive Dynamics and the Implementation Problem with Complete Information,” *Journal of Economic Theory* 86, 159-184.

Eliaz, K. (2000): “Fault Tolerant Implementation,” mimeo, University of Tel-Aviv.

Fudenberg, D. and D. Levine (1998): *The Theory of Learning in Games*, Massachusetts Institute of Technology Press.

Hart, S. and A. Mas-Colell (2000): “A Simple Adaptive Procedure Leading to Correlated Equilibrium,” *Econometrica* 68, 1127-1150

Maniquet, F. (1998): “Implementation under Perfect Information in Economic Environments,” mimeo, University of Manur.

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

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

Matsushima, H. (2001): “Stable Implementation,” Discussion Paper CIRJE-F-128, Faculty of Economics, University of Tokyo.

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.

Moore, J. and R. Repullo (1988): “Subgame Perfect Implementation,” *Econometrica* 56, 1191-1220.

Palfrey, T. and S. Srivastava (1991): “Nash Implementation Using Undominated Strategies,” *Econometrica* 59, 479-501.

Sefton, M. and A. Yavas (1996): “Abreu-Matsushima Mechanisms: Experimental Evidence,” *Games and Economic Behavior* 16, 280-302.

# Appendix A: Proof of Theorem 1

Fix $\omega \in \Omega$ arbitrarily. Lemma A-1 is based on the fact that the virtualness serves to provide each agent with the incentive to make the honest announcement about her own utility index.

**Lemma A-1:** *For every* $m \in \tilde{M}$ *, if a message profile* $m' \in \tilde{M}$ *is given by*

$$m'_i = (\tilde{\omega}_{i-1}(m), \omega_i, \tilde{\omega}_{i+1}(m)) \text{ for all } i \in N,$$

*then,* $m' \in \tilde{M}$ *is reachable from* $m$ *in* $(G^*, \omega)$.

**Proof:** Note $\tilde{\omega}(m') = \tilde{\omega}(m)$. Consider $(i^k, m^k)_{k=1}^{2n}$ connected with $m$ satisfying that for every $k = 1,...,n$,

$$i^k = k,\ m^k_{i^k} = (\tilde{\omega}_{i^k-1}(m), m_{i^k,i^k}, \tilde{\omega}_{i^k+1}(m)),$$

and, for every $k = n+1,...,2n$,

$$i^k = k - n \text{ and } m^k_{i^k} = (\tilde{\omega}_{i^k-1}(m), \omega_{i^k}, \tilde{\omega}_{i^k+1}(m)).$$

Note $m^{2n} = m'$. Note that for every $k = 1,...,2n$,

$$t^*_{i^k}(m^k) \geq t^*_{i^k}(m^{k-1}),$$

for every $k = 1,...,n$,

$$g^*(m^k) = g^*(m) = (1-\varepsilon)x(\tilde{\omega}(m)) + \frac{\varepsilon}{n}\sum_{i \in N}\alpha_i(m_{i,i}),$$

and for every $k = n+1,...,2n$,

$$g^*(m^k) = (1-\varepsilon)x(\tilde{\omega}(m)) + \frac{\varepsilon}{n}\sum_{i \leq k}\alpha_i(\omega_i) + \frac{\varepsilon}{n}\sum_{i > k}\alpha_i(m_{i,i}).$$

Hence, it follows from the definition of $\alpha_i(\cdot)$ that for every $k = 1,...,2n$,

$$u_{i^k}(g^*(m^k), \omega_{i^k}) + t^*_{i^k}(m^k) \geq u_{i^k}(g^*(m^{k-1}), \omega_{i^k}) + t^*_{i^k}(m^{k-1}),$$

and therefore, $m'$ is reachable from $m$. **Q.E.D.**

Lemma A-2 is based on the fact that agent 1 becomes dictatorial when there does not exist the profile of the public opinions.

**Lemma A-2:** *For every* $m \in \tilde{M}$ *, if*

$$m_i = (\tilde{\omega}_{i-1}(m), \omega_i, \tilde{\omega}_{i+1}(m)) \text{ for all } i \in N, \text{ and}$$

$$\omega_i \neq \tilde{\omega}_i(m) \text{ for some } i \in N,$$

*then, there exists* $m' \notin \tilde{M}$ *that is reachable from* $m$ *in* $(G^*, \omega)$.

**Proof:** Suppose $\omega_2 \neq \tilde{\omega}_2(m)$. Specify $m'$ by

$$m'_{1,1} = \omega_1,\ m'_{1,2} \notin \{\tilde{\omega}_2(m), \omega_2\},\ m'_{1,n} = \tilde{\omega}_n(\omega), \text{ and } m'_i = m_i \text{ for all } i \neq 1.$$

Such an $m'$ exists because $|\Omega_2| \geq 3$. Since $m' \notin \tilde{M}$, it follows that

$$u_1(g^*(m'), \omega_1) + t^*_1(m') = u_1(d(\omega_1), \omega_1)$$

$$\geq u_1((1-\varepsilon)x(\tilde{\omega}(m))+\frac{\varepsilon}{n}\sum_{j\in N}\alpha_j(\omega_j),\omega_1)=u_1(g^*(m),\omega_1)+t_1^*(m),$$

and therefore, $m'$ is reachable from $m$. By replacing 2 and $n$ by $n$ and 2, respectively, we can obtain the same result in the case of $\omega_n \neq \tilde{\omega}_n(m)$.

Next, suppose $\omega_2 = \tilde{\omega}_2(m)$ and $\omega_n = \tilde{\omega}_n(m)$. Assume that agent 3 is not the single deviant from $\tilde{\omega}(m)$. Consider $(i^k, m^k)_{k=1}^2$ connected with $m$ satisfying that

$$i^1 = 3,\ m^1_{i^1,2} \neq \omega_2,\ (m^1_{i^1,i^1}, m^1_{i^1,i^1+1}) = (\omega_{i^1}, \tilde{\omega}_{i^1+1}(m)),$$

$$i^2 = 1,\ m^2_{i^2,2} \notin \{\omega_2, m^1_{i^1,2}\},\ \text{and}\ (m^2_{i^2,i^2-1}, m^2_{i^2,i^2}) = (\tilde{\omega}_{i^2-1}(m), \omega_{i^2}).$$

Such a sequence $(i^k, m^k)_{k=1}^2$ exists because $|\Omega_2| \geq 3$. Note $m^1 \in \tilde{M}$ because $\omega_2 = \tilde{\omega}_2(m)$, but note $m^2 \notin \tilde{M}$. Hence, it follows that

$$g^*(m^1) = g^*(m) = (1-\varepsilon)x(\tilde{\omega}(m)) + \frac{\varepsilon}{n}\sum_{j\in N}\alpha_j(\omega_j),\ t^*_{i^1}(m^1) = 0,$$

$$g^*(m^2) = d(\omega_1),\ \text{and}\ t^*_{i^2}(m^2) = 0.$$

Hence, for each $k \in \{1,2\}$,

$$u_{i^k}(g^*(m^k), \omega_{i^k}) + t^*_{i^k}(m^k) \geq u_{i^k}(g^*(m^{k-1}), \omega_{i^k}) + t^*_{i^k}(m^{k-1}),$$

and therefore, there exists $m' \notin \tilde{M}$ that is reachable from $m$. Next, assume that agent 3 is the single deviant from $\tilde{\omega}(m)$. Since agent $n-1$ is not the single deviant from $\tilde{\omega}(m)$, we can obtain the same result by replacing 3 and 2 by $n-1$ and $n$, respectively. **Q.E.D.**

Lemma A-3 is based on the key heuristic that for each agent, there exist a wide variety of messages such that whichever message she announces, the resultant allocation never changes.

**Lemma A-3:** For every $m \notin \tilde{M}$, *there exists* $m' \notin \tilde{M}$ *that is reachable from* $m$ *in* $(G^*, \omega)$ *such that for every* $i \in N$,

$$m'_{i,i} = \omega_i,\ \textit{and}\ m'_{i-1,i} \neq m'_{i,i} \neq m'_{i+1,i} \neq m'_{i-1,i}.$$

**Proof:** Since $m \notin \tilde{M}$, we can choose $\hat{i} \in N$ satisfying that

$$m_{\hat{i}-1,\hat{i}} \neq m_{\hat{i},\hat{i}} \neq m_{\hat{i}+1,\hat{i}} \neq m_{\hat{i}-1,\hat{i}}.$$

Consider $(i^k, m^k)_{k=1}^{2n}$ that is connected with $m$ satisfying that for every $k = 1,...,n$,

$$i^k = k,\ m^k_{i^k,\hat{i}} = m_{i^k,\hat{i}},$$

$$m^k_{i^k,i^k} = \omega_{i^k}\ \text{if}\ i^k \neq \hat{i},$$

$$m^k_{i^k,i^k+1} \notin \{\omega_{i^k+1}, m^{k-1}_{i^k+2,i^k+1}\}\ \text{if}\ i^k+1 \neq \hat{i},$$

$$m^k_{i^k,i^k-1} \notin \{\omega_{i^k-1}, m^{k-1}_{i^k+1,i^k-1}\}\ \text{if}\ i^k-1 \neq \hat{i},$$

for every $k = n+1,...,2n$,

$$i^k = k-n,\ m^k_{i^k,i^k} = \omega_{i^k},$$

$$m^k_{i^k,i^k+1} \notin \{\omega_{i^k+1}, m^{k-1}_{i^k+2,i^k+1}\}\ \text{if}\ i^k+1 = \hat{i},$$

$$m^k_{i^k,i^k-1} \notin \{\omega_{i^k-1}, m^{k-1}_{i^k+1,i^k-1}\} \text{ if } i^k - 1 = \hat{i}\ ,$$

and

$$m^k_{i^k,j} = m^n_{i^k,j} \text{ if } j \neq \hat{i}\ .$$

Note that for every $k \in \{1,...,2n\}$,

$$m^k \notin \tilde{M}\ ,$$

for every $k = 1,...,n+\hat{i}-1$,

$$g^*(m^k) = d(m_{1,1}) \text{ and } t^*_{i^k}(m^k) = 0\ ,$$

and, for every $k = n+\hat{i},...,2n$,

$$g^*(m^k) = d(\omega_1) \text{ and } t^*_{i^k}(m^k) = 0\ .$$

Hence, for every $k = 1,...,2n$,

$$u_{i^k}(g^*(m^k),\omega_{i^k}) + t^*_{i^k}(m^k) \geq u_{i^k}(g^*(m^{k-1}),\omega_{i^k}) + t^*_{i^k}(m^{k-1})\ ,$$

and therefore, $m^{2n}$ is reachable from $m$. Note that for every $i \in N$,

$$m^{2n}_{i,i} = \omega_i \text{ and } m^{2n}_{i-1,i} \neq m^{2n}_{i,i} \neq m^{2n}_{i+1,i} \neq m^{2n}_{i-1,i}\ .$$

**Q.E.D.**

Lemma A-4 is based on the fact that there exists an agent $j \neq 1$ satisfying inequality (2), together with the key heuristic of the paper.

**Lemma A-4:** *For every $m \notin \tilde{M}$, if for every $i \in N$,*

$$m_{i,i} = \omega_i \text{ and } m_{i-1,i} \neq m_{i,i} \neq m_{i+1,i} \neq m_{i-1,i}\ ,$$

*then, there exists $m' \in \tilde{M}$ that is reachable from $m$ in $(G^*,\omega)$ such that $\tilde{\omega}(m') = \omega$.*

**Proof:** Let $j \neq 1$ be the agent satisfying inequality (2). Consider $(i^k, m^k)^{n-1}_{k=1}$ that is connected with $m$ satisfying that

$$m^k_{i^k} = (\omega_{i^k-1}, \omega_{i^k}, \omega_{i^k+1}) \text{ for all } k = 1,...,n-1\ ,$$

$$i^k = k + j \text{ for all } k = 1,...,n-3\ ,$$

$$i^{n-2} = j-1, \text{ and } i^{n-1} = j\ .$$

Note that for every $k = 1,...,n-2$, $m^k \notin \tilde{M}$, $m^k_{1,1} = \omega_1$,

$$g^*(m^k) = d(\omega_1), \text{ and } t^*_i(m^k) = 0 \text{ for all } i \in N\ .$$

Note that $m^{n-1} \in \tilde{M}$, $\tilde{\omega}(m^{n-1}) = \omega$, and agent $j-2$ is the single agent $i$ satisfying that

$$m^{n-1}_i \neq (\tilde{\omega}_{i-1}(m^{n-1}), \tilde{\omega}_i(m^{n-1}), \tilde{\omega}_{i+1}(m^{n-1}))\ .$$

Note also that $m^{n-1}_{j-2,j-1} \neq \tilde{\omega}_{j-1}(m^{n-1})$. Hence, it follows that

$$g^*(m^{n-1}) = \hat{x}(\omega) \text{ and } t^*_{i^{n-1}}(m^{n-1}) = \xi\ .$$

These imply that for every $k = 1,...,n-1$,

$$u_{i^k}(g^*(m^k),\omega_{i^k}) + t^*_{i^k}(m^k) \geq u_{i^k}(g^*(m^{k-1}),\omega_{i^k}) + t^*_{i^k}(m^{k-1})\ ,$$

and therefore, $m^{n-1}$ is reachable from $m$. **Q.E.D.**

By using these lemmata, we can show that $\mu^*(\omega)$ is stable in $(G^*,\omega)$. For every $i \in N$, and every $m_i \neq \mu_i^*(\omega)$, if $m_{i,i} \neq \omega_i$, then

$$u_i(g^*(\mu^*(\omega)),\omega_i) + t_i^*(\mu^*(\omega)) = u_i(\hat{x}(\omega),\omega_i)$$

$$> u_i(\hat{x}(\omega) + \frac{\varepsilon}{n}\{\alpha_i(m_{i,i}) - \alpha_i(\omega_i)\},\omega_i)$$

$$\geq u_i(g^*(\mu^*(\omega)/m_i),\omega_i) + t_i^*(\mu^*(\omega)/m_i),$$

whereas, if $m_{i,i} = \omega_i$, then

$$u_i(g^*(\mu^*(\omega)),\omega_i) + t_i^*(\mu^*(\omega)) = u_i(\hat{x}(\omega),\omega_i) > u_i(\hat{x}(\omega),\omega_i) - \xi$$

$$= u_i(g^*(\mu^*(\omega)/m_i),\omega_i) + t_i^*(\mu^*(\omega)/m_i).$$

Hence, $\mu^*(\omega)$ is a strict Nash equilibrium in $(G^*,\omega)$. Lemma A-1 implies that if $m \in \tilde{M}$ and $\tilde{\omega}(m) = \omega$, then $\mu^*(\omega)$ is reachable from $m$. Lemmata A-3 and A-4 imply that if $m \notin \tilde{M}$, then there exists $m' \in \tilde{M}$ reachable from $m$ such that $\tilde{\omega}(m') = \omega$. Lemma A-2 implies that if $m \in \tilde{M}$ and $\tilde{\omega}(m) \neq \omega$, then there exists $m' \notin \tilde{M}$ reachable from $m$. Hence, $\mu^*(\omega)$ is reachable from every message profile, no other message profile is reachable from $\mu^*(\omega)$ because it is a strict Nash equilibrium, and therefore, $\mu^*(\omega)$ is stable in $(G^*,\omega)$.

# Appendix B

**Proof of Theorem 3:** Fix $\omega \in \Omega$ arbitrarily. Note that the proofs of Lemmata A-1, A-2, and A-3 do not depend on the existence of $j \neq 1$ satisfying inequality (2). In the same way as in the proof of Theorem 1, $\mu^*(\omega)$ is a strict Nash equilibrium in $(G^*, \omega)$. We will show that for every $m \notin \tilde{M}$, if $m_{i,i} = \omega_i$, and $m_{i-1,i} \neq m_{i,i} \neq m_{i+1,i} \neq m_{i-1,i}$ for all $i \in N$, then $\mu^*(\omega)$ is weakly reachable from $m$ in $(G^*, \omega)$. Note that for every $i \in N$, $m / \mu_i^*(\omega) \notin \tilde{M}$, and therefore,

$$u_i(g^*(m/\mu_i^*(\omega)), \omega_i) + t_i^*(m/\mu_i^*(\omega)) = u_i(d(\omega_1), \omega_i).$$

Hence, for every $i \in N$,

$$u_i(g^*(m/\mu_i^*(\omega)), \omega_i) + t_i^*(m/\mu_i^*(\omega)) = u_i(g^*(m), \omega_i) + t_i^*(m),$$

and therefore, $\mu^*(\omega)$ is weakly reachable from $m$ in $(G^*, \omega)$. This property, together with Lemmata A-1, A-2, and A-3, implies that for every $m \in M$, $\mu^*(\omega)$ is weakly reachable from $m$.

**Q.E.D.**

**Proof of Proposition 4:** Suppose that $\mu^*(\omega)$ is BR-stable in $(G^*, \omega)$. Then, there exist $m \notin \tilde{M}$, $i \in N$, and $(\omega'_{i-1}, \omega'_i, \omega'_{i+1}) \neq (\omega_{i-1}, \omega_i, \omega_{i+1})$ such that for every $j \notin \{i-1, i, i+1\}$,

$$m_{j',j} = \omega_j \text{ for at least two agents } j' \in \{j-1, j, j+1\},$$

for every $j \in \{i-1, i, i+1\}$,

$$m_{j',j} = \omega'_j \text{ for some } j' \in \{j-1, j, j+1\}/\{i\},$$

and for every $m''_i \in M_i$,

$$u_i(g^*(m/\mu_i^*(\omega)), \omega_i) + t_i^*(m/\mu_i^*(\omega)) \geq u_i(g^*(m/m''_i), \omega_i) + t_i^*(m/m''_i) \quad . \tag{B1}$$

Note from the Supposition of Proposition 4 that

$$u_i(g^*(m/\mu_i^*(\omega)), \omega_i) + t_i^*(m/\mu_i^*(\omega)) \leq u_i(\hat{x}(\omega) + \frac{\varepsilon}{n}\sum_{j\neq i}\{\alpha_j(m_{j,j}) - \alpha(\omega_j)\}, \omega_i) + \xi .$$

Let $m'_i = (\omega'_{i-1}, \omega'_i, \omega'_{i+1})$, and let $\overline{\omega} \in \Omega$ denote the preference profile satisfying that

$$\overline{\omega}_h = \omega'_h \text{ for all } h \in \{i-1, i, i+1\}, \text{ and } \overline{\omega}_h = \omega_h \text{ for all } h \notin \{i-1, i, i+1\}.$$

Note that $\tilde{\omega}(m/m'_i) = \overline{\omega}$, $m'_i = (\overline{\omega}_{i-1}, \overline{\omega}_i, \overline{\omega}_{i+1})$, and therefore,

$$u_i(g^*(m/m'_i), \omega_i) + t_i^*(m/m'_i) \geq u_i(\hat{x}(\overline{\omega}) + \frac{\varepsilon}{n}\sum_{j\neq i}\{\alpha_j(m_{j,j}) - \alpha(\omega'_j)\}, \omega_i).$$

Since $\varepsilon > 0$ and $\xi > 0$ are small, the supposition of Proposition 4 implies that

$$u_i(\hat{x}(\overline{\omega}) + \frac{\varepsilon}{n}\sum_{j\neq i}\{\alpha_j(m_{j,j}) - \alpha(\omega'_j)\}, \omega_i)$$

$$> u_i(\hat{x}(\omega) + \frac{\varepsilon}{n}\sum_{j\neq i}\{\alpha_j(m_{j,j}) - \alpha(\omega_j)\}, \omega_i) + \xi .$$

Hence, it follows from the above inequalities that

$$u_i(g^*(m/m_i'),\omega_i)+t_i^*(m/m_i') > u_i(g^*(m/\mu_i^*(\omega)),\omega_i)+t_i^*(m/\mu_i^*(\omega)).$$

This contradicts inequalities (B1). **Q.E.D.**

**Proof of Theorem 9:** Fix $\omega \in \Omega$ arbitrarily. The proof of Lemmata C-1, C-2, and C-3 do not depend on the existence of $j \neq 1$ satisfying inequality (2). In the same way as in the proof of Theorem 5, $\mu^+(\omega)$ is a strict Nash equilibrium in $(G^+,\omega)$. We will show that for every $m \notin \tilde{M}$, if for every $i \in N$, $m'_{i-1,i} \neq m'_{i,i} = \omega_i \neq m'_{i+1,i} \neq m'_{i-1,i}$ and $m'_{i-2,i} = 1$, then $\mu^+(\omega)$ is weakly BR-reachable from $m$ in $(G^+,\omega)$. Note that for every $i \in N$, $m/\mu_i^+(\omega) \notin \tilde{M}$, and therefore,

$$u_i(g^+(m/\mu_i^+(\omega)),\omega_i)+t_i^+(m/\mu_i^+(\omega)) = u_i(d(\omega_1),\omega_i).$$

Hence, for every $i \in N$,

$$u_i(g^+(m/\mu_i^+(\omega)),\omega_i)+t_i^+(m/\mu_i^+(\omega)) = u_i(g^+(m),\omega_i)+t_i^+(m),$$

and therefore, $\mu^+(\omega)$ is weakly BR-reachable from $m$ in $(G^+,\omega)$. This property, together with Lemmata C-1, C-2, and C-3, implies that for every $m \in M$, $\mu^+(\omega)$ is weakly BR-reachable from $m$ in $(G^*,\omega)$.
**Q.E.D.**

## Appendix C: Proof of Theorem 5

Fix $(\omega, m) \in \Omega \times M$ arbitrarily.

**Lemma C-1:** *For every* $m \in \tilde{M}$ *, if there exists a single agent* $i \in N$ *such that* $m_i \neq \mu_i^+(\omega)$*, then,* $\mu^+(\omega)$ *is BR-reachable from* $m$ *in* $(G^+, \omega)$*.*

**Proof:** Note $\tilde{\omega}(m) = \omega$, $m / \mu_i^+(\omega) = \mu^+(\omega)$, and therefore,

$$g^+(m / \mu_i^+(\omega)) = \hat{x}(\omega), \; t_i^+(m / \mu_i^+(\omega)) = 0,$$

and, for every $m_i' \in M_i$,

$$g^+(m / m_i') = \hat{x}(\omega) + \frac{\varepsilon}{n}\{\alpha_i(m_{i,i}') - \alpha_i(\omega_i)\}, \text{ and } t_i^+(m / m_i') \leq 0.$$

The definition of $\alpha_i(\cdot)$ implies that for every $m_i' \in M_i$,

$$u_i(g^+(m / \mu_i^+(\omega)), \omega_i) + t_i^+(m / \mu_i^+(\omega)) \geq u_i(g^+(m / m_i'), \omega_i) + t_i^+(m / m_i'),$$

and therefore, that $\mu^+(\omega) = m / \mu_i^+(\omega)$ is BR-reachable from $m$ . **Q.E.D.**

**Lemma C-2:** *For every* $m \in \tilde{M}$ *, either* $\mu^+(\omega)$ *is BR-reachable from* $m$ *in* $(G^+, \omega)$*, or there exists* $m' \notin \tilde{M}$ *that is BR-reachable from* $m$ *in* $(G^+, \omega)$*.*

**Proof:** Suppose that $\mu^+(\omega)$ is not BR-reachable from $m$ and there exists no $m' \notin \tilde{M}$ BR-reachable from $m$. Then, there exists $(i^k, m^k)_{k=1}^n$ connected with $m$ such that for every $k = 1,...,n$,

$$i^k = k, \; m^k \in \tilde{M}, \; (m_{i^k, i^k-1}^k, m_{i^k, i^k+1}^k, m_{i^k, i^k+2}^k) = (\tilde{\omega}_{i^k-1}(m), \tilde{\omega}_{i^k+1}(m), 0),$$

and, for every $m_{i^k}'' \in M_{i^k}$,

$$u_{i^k}(g^+(m^k), \omega_{i^k}) + t_{i^k}^+(m^k) \geq u_{i^k}(g^+(m^k / m_{i^k}''), \omega_{i^k}) + t_{i^k}^+(m^k / m_{i^k}'').$$

Note that $(m_{i,i-1}^n, m_{i,i+1}^n, m_{i,i+2}^n) = (\tilde{\omega}_{i-1}(m), \tilde{\omega}_{i+1}(m), 0)$ for all $i \in N$, and that $m^n$ is BR-reachable from $m$. Hence, without loss of generality, we can assume that $m$ satisfies that

$$(m_{i,i-1}, m_{i,i+1}, m_{i,i+2}) = (\tilde{\omega}_{i-1}(m), \tilde{\omega}_{i+1}(m), 0) \text{ for all } i \in N.$$

Consider $(i^k, m^k)_{k=1}^{n+2}$ that is connected with $m$ satisfying that for every $k = 1,...,n$,

$$i^k = k, \; m_{i^k}^k = (\tilde{\omega}_{i^k-1}(m), \omega_{i^k}, \tilde{\omega}_{i^k+1}(m), 0),$$

for $k = n+1$,

$$i^k = n, \; (m_{i^k, i^k-1}^k, m_{i^k, i^k}^k, m_{i^k, i^k+1}^k) = (\tilde{\omega}_{i^k-1}(m), \omega_{i^k}, \tilde{\omega}_{i^k+1}(m)),$$

$$m_{i^k,2}^k = 0 \text{ if } \tilde{\omega}_2(m) \neq \omega_2,$$

$$m_{i^k,2}^k = 1 \text{ if } \tilde{\omega}_2(m) = \omega_2,$$

for $k = n+2$,

$$i^k = 1, \; (m_{i^k, i^k-1}^k, m_{i^k, i^k}^k, m_{i^k, i^k+2}^k) = (\tilde{\omega}_{i^k-1}(m), \omega_{i^k}, 0), \text{ and } m_{i^k,2}^k \notin \{\tilde{\omega}_2(m), \omega_2\}.$$

Note $m^k \in \tilde{M}$ for all $k \neq n+2$, but $m^{n+2} \notin \tilde{M}$.

Fix $k = 1,...,n$ arbitrarily. Suppose inequalities (6) for every $k' \leq k-1$. Then, $m^{k-1}$ is BR-reachable from $m$, and therefore, it follows from the supposition at the beginning of this proof that there exists no $m' \notin \tilde{M}$ BR-reachable from $m^{k-1}$. For every $m''_{i^k} \in M_{i^k}$,

$$u_{i^k}(g^+(m^k), \omega_{i^k}) + t^+_{i^k}(m^k) \geq u_{i^k}(g^+(m^k / m''_{i^k}), \omega_{i^k}) + t^+_{i^k}(m^k / m''_{i^k}) \text{ if } m^k / m''_{i^k} \notin M .$$

Since $m^k \in \tilde{M}$, $m^k_{i^k, i^k} = \omega_{i^k}$, and $(m^k_{i,i-1}, m^k_{i,i+1}, m^k_{i,i+2}) = (\tilde{\omega}_{i-1}(m), \tilde{\omega}_{i+1}(m), 0)$ for all $i \in N$, it follows that for every $m''_{i^k} \in M_{i^k}$,

$$u_{i^k}(g^+(m^k), \omega_{i^k}) + t^+_{i^k}(m^k) \geq u_{i^k}(g^+(m^k / m''_{i^k}), \omega_{i^k}) + t^+_{i^k}(m^k / m''_{i^k}) \text{ if } m^k / m''_{i^k} \in \tilde{M} .$$

Since inequalities (6) automatically hold for $k = 1$, the above arguments imply inequalities (6) for every $k = 1,...,n$, and therefore, $m^n$ is BR-reachable from $m$.

Since $m^n$ is BR-reachable from $m$, it follows that $\mu^+(\omega)$ is not BR-reachable from $m^n$, and there exists no $m' \notin \tilde{M}$ BR-reachable from $m^n$, Hence, for every $m''_{i^n} \in M_{i^n}$,

$$u_{i^n}(g^+(m^n), \omega_{i^n}) + t^+_{i^n}(m^n) \geq u_{i^n}(g^+(m^n / m''_{i^n}), \omega_{i^n}) + t^+_{i^n}(m^n / m''_{i^n}) \text{ if } m^n / m''_{i^n} \notin M .$$

Since $m^n_i = (\tilde{\omega}_{i-1}(m), \omega_i, \tilde{\omega}_{i+1}(m), 0)$ for all $i \in N$, it follows from Lemma C-1 that there exist two agents $i \in N$ such that $m^n_{i,i} \neq \tilde{\omega}_i(m^n)$. This implies that $t^+_{i^{n+1}}(m^{n+1}) = t^+_{i^{n+1}}(m^n) = 0$, and therefore, for every $m''_{i^{n+1}} \in M_{i^{n+1}}$,

$$u_{i^{n+1}}(g^+(m^{n+1}), \omega_{i^{n+1}}) + t^+_{i^{n+1}}(m^{n+1}) \geq u_{i^{n+1}}(g^+(m^{n+1} / m''_{i^{n+1}}), \omega_{i^{n+1}}) + t^+_{i^{n+1}}(m^{n+1} / m''_{i^{n+1}}) \text{ if } m^{n+1} / m''_{i^{n+1}} \in \tilde{M} .$$

Hence, inequalities (6) hold for $k = n+1$, and therefore, $m^{n+1}$ is BR-reachable from $m$.

Inequalities (6) for $k = n+2$ are straightforward from the fact that $i^{n+2} = 1$, $m^{n+2} \notin \tilde{M}$, and $m^{n+2}_{1,1} = \omega_1$. Hence, $m^{n+2}$ is BR-reachable from $m$. But this contradicts $m^{n+2} \notin \tilde{M}$.

**Q.E.D.**

**Lemma C-3:** *For every $m \notin \tilde{M}$, there exists $m' \notin \tilde{M}$ that is BR-reachable from $m$ such that for every $i \in N$, $m'_{i-1,i} \neq m'_{i,i} = \omega_i \neq m'_{i+1,i} \neq m'_{i-1,i}$ and $m'_{i-2,i} = 1$.*

**Proof:** Since $m \notin \tilde{M}$, there exists $r \in N$ such that either

$$m_{r-1,r} \neq m_{r,r} \neq m_{r+1,r} \neq m_{r-1,r} \text{ and } m_{r-2,r} = 0 ,$$

or

$$m_{j,r} \neq m_{r,r} \text{ for some } j \in \{r-1, r+1\} \text{ and } m_{r-2,r} = 1 .$$

We can choose $(i(1),...,i(4)) \in \{r-2,...,r+1\}^4$ as follows. If $m_{r-2,r} = 0$, then,

$$i(1) = r-2 ,\ i(2) = r-1 ,\ i(3) = r+1 , \text{ and } i(4) = r .$$

If $m_{r-2,r} = 1$, then,

$i(1) \in \{r-1, r+1\}$, $i(2) \in \{r-1, r+1\} / \{i(1)\}$, $m_{i(2),r} \neq m_{r,r}$, $i(3) = r-2$, and $i(4) = r$.

We specify $(i^k, m^k)_{k=1}^{2n}$ connected with $m$ in the following way.

(1) For every $k \in \{1,...,2n\}$,

$$m^k_{i^k,i^k} = \omega_{i^k} \text{ and } m^k_{i^k,i^k+2} = 1.$$

(2) For every $k \in \{1,...,n\}$,

$$i^k = r + k + 1 \text{ if } k \in \{1,...,n-4\}$$

$$i^k = i(k - n + 4) \text{ if } k \in \{n-3,...,n\},$$

for every $i \in \{i^k - 1, i^k + 1\}$,

$$m^k_{i^k,i} \neq \omega_i \text{ and } m^k_{i^k,i} \neq m^{k-1}_{j,i} \text{ for all } j \in \{i-1, i, i+1\}/\{i^k\} \text{ if there exists}$$

$$\omega'_i \in \Omega_i \text{ such that } \omega'_i \neq \omega_i \text{ and } \omega'_i \neq m^{k-1}_{j,i} \text{ for all } j \in \{i-1, i, i+1\}/\{i^k\},$$

and

$$m^k_{i^k,i} = \omega_i \text{ if there exists no such } \omega'_i \in \Omega_i.$$

(3) For every $k \in \{n+1,...,2n\}$,

$$i^k = k - n,$$

and, for every $i \in \{i^k - 1, i^k + 1\}$,

$$m^k_{i^k,i} \neq \omega_i \text{ and } m^k_{i^k,i} \neq m^{k-1}_{j,i} \text{ for all } j \in \{i-1, i, i+1\}/\{i^k\}.$$

Note that for every $k \in \{1,...,2n\}$,

$$m^k / m''_{i^k} \notin \tilde{M} \text{ for all } m''_{i^k} \in M_{i^k}, \text{ and}$$

$$g^+(m^k) = d(\omega_1) \text{ if } i^k = 1.$$

Hence, inequalities (6) hold for every $k \in \{1,...,2n\}$, and therefore, $m^{2n}$ is BR-reachable from $m$. Note that for every $i \in N$, $m^{2n}_{i-1,i} \neq m^{2n}_{i,i} = \omega_i \neq m^{2n}_{i+1,i} \neq m^{2n}_{i-1,i}$ and $m^{2n}_{i-1,i} = 1$.
**Q.E.D.**

**Lemma C-4:** *For every $m \notin \tilde{M}$, if for every $i \in N$,*

$$m'_{i-1,i} \neq m'_{i,i} = \omega_i \neq m'_{i+1,i} \neq m'_{i-1,i} \text{ and } m'_{i-2,i} = 1,$$

*then, there exists $m' \in \tilde{M}$ that is BR-reachable from $m$ such that there exists a single agent $i \in N$ satisfying that $m'_i \neq \mu^+_i(\omega)$.*

**Proof:** Consider $(i^k, m^k)^{n-1}_{k=1}$ connected with $m$ satisfying that for every $k \in \{1,...,n-1\}$,

$$m^k_{i^k} = \mu^+_{i^k}(\omega),$$

$$i^k = j + k \text{ if } k \in \{1,...,n-4\},$$

$$i^{n-3} = j - 2,\ i^{n-2} = j - 1, \text{ and } i^{n-1} = j,$$

where inequality (2) holds for $j \neq 1$. Note that

$$m^k \notin \tilde{M} \text{ and } g^+(m^k) = d(\omega_1) \text{ for all } k \in \{1,...,n-2\},\ m^{n-1} \in \tilde{M},$$

$$m^{n-1}_i = \mu^+_i(\omega) \text{ for all } i \neq j-3,\ g^+(m^{n-1}) = \hat{x}(\omega), \text{ and } t^+_j(m^{n-1}) = \xi.$$

Note that for every $m''_j \neq \mu^+_j(\omega)$,

$$t^+_j(m^{n-1} / m''_j) = 0,$$

$$g^{+}(m^{n-1}/m''_{j})=\hat{x}(\omega)+\frac{\varepsilon}{n}\{\alpha_{j}(m''_{j,j})-\alpha_{j}(\omega_{j})\}\text{ if }m''_{j,j-1}=\omega_{j-1},$$

and

$$g^{+}(m^{n-1})=d(\omega_{1})\text{ if }m''_{j,j-1}\neq\omega_{j-1}.$$

These imply inequalities (6) for every $k\in\{1,...,n-1\}$. Hence, $m^{n-1}$ is BR-reachable from $m$. Since agent $j-3$ is the single deviant from $\mu^{+}(\omega)$, we have proved this lemma.

**Q.E.D.**

By using these lemmata, we will show that $\mu^{+}(\omega)$ is BR-stable in $(G^{+},\omega)$. Note that for every $i\in N$, and every $m_{i}\neq\mu_{i}^{+}(\omega)$, if $m_{i,i}\neq\omega_{i}$, then

$$u_{i}(g^{+}(\mu^{+}(\omega)),\omega_{i})+t_{i}^{+}(\mu^{+}(\omega))=u_{i}(\hat{x}(\omega),\omega_{i})$$

$$>u_{i}(\hat{x}(\omega)+\frac{\varepsilon}{n}\{\alpha_{i}(m_{i,i})-\alpha_{i}(\omega_{i})\},\omega_{i})$$

$$\geq u_{i}(g^{+}(\mu^{+}(\omega)/m_{i}),\omega_{i})+t_{i}^{+}(\mu^{+}(\omega)/m_{i}),$$

whereas, if $m_{i,i}=\omega_{i}$, then

$$u_{i}(g^{+}(\mu^{+}(\omega)),\omega_{i})+t_{i}^{+}(\mu^{+}(\omega))=u_{i}(\hat{x}(\omega),\omega_{i})>u_{i}(\hat{x}(\omega),\omega_{i})-\xi$$

$$=u_{i}(g^{+}(\mu^{+}(\omega)/m_{i}),\omega_{i})+t_{i}^{+}(\mu^{+}(\omega)/m_{i}).$$

These inequalities imply that $\mu^{+}(\omega)$ is a strict Nash equilibrium in $(G^{+},\omega)$. Lemma C-1 implies that for every $m\in\tilde{M}$, if there exists the single agent $i\in N$ such that $m_{i}\neq\mu_{i}^{+}(\omega)$, then $\mu^{+}(\omega)$ is BR-reachable from $m$. Lemmas C-3 and C-4 imply that for every $m\notin\tilde{M}$, there exists $m'\in\tilde{M}$ BR-reachable from $m$ such that there exists the single deviant from $\mu^{+}(\omega)$. Lemma C-2 implies that for every $m\in\tilde{M}$, either $\mu^{+}(\omega)$ is BR-reachable from $m$ or there exists $m'\notin\tilde{M}$ that is BR-reachable from $m$. These observations imply that $\mu^{+}(\omega)$ is BR-reachable from every message profile. Note that no $m\neq\mu^{+}(\omega)$ is reachable from $\mu^{+}(\omega)$ because $\mu^{+}(\omega)$ is a strict Nash equilibrium.

From the above arguments, we have proved that $\mu^{+}(\omega)$ is BR-stable in $(G^{+},\omega)$, and therefore, we have completed the proof of Theorem 5.

## Appendix D: Proof of Theorem 7

Fix $\omega \in \Omega$ and $m \in M$ arbitrarily.

**Lemma D-1:** *For every $m \in \tilde{M}$, if that there exists the single agent $i \in N$ such that $m_i \neq \mu_i^{**}(\omega)$, then, $\mu^{**}(\omega)$ is BR-reachable from $m$ in $(G^{**}, \omega)$.*

**Proof:** Note that $\tilde{\omega}(m) = \omega$ and $m / \mu_i^{**}(\omega) = \mu^{**}(\omega)$. Hence,

$$g^{**}(m / \mu_i^{**}(\omega)) = \hat{x}(\omega),\ t_i^{**}(m / \mu_i^{**}(\omega)) = 0,$$

and for every $m_i' \in M_i$,

$$g^{**}(m / m_i') = \hat{x}(\omega) + \frac{\varepsilon}{n}\{\alpha_i(m_{i,i}') - \alpha_i(\omega_i)\}, \text{ and } t_i^{**}(m / m_i') \leq 0.$$

The definition of $\alpha_i(\cdot)$ implies that for every $m_i' \in M_i$,

$$u_i(g^{**}(m / \mu_i^{**}(\omega)), \omega_i) + t_i^{**}(m / \mu_i^{**}(\omega)) \geq u_i(g^{**}(m / m_i'), \omega_i) + t_i^{**}(m / m_i'),$$

and therefore, that $\mu^{**}(\omega) = m / \mu_i^{**}(\omega)$ is BR-reachable from $m$. **Q.E.D.**

**Lemma D-2:** *For every $m \in \tilde{M}$, if $\mu^{**}(\omega)$ is not BR-reachable from $m$ in $(G^{**}, \omega)$, then, there exists $m' \notin \tilde{M}$ that is BR-reachable from $m$ in $(G^{**}, \omega)$ such that there exist $r \in N$ and $r' \in N$ such that for every $\omega_r' \in \Omega_r$,*

$m_{i,r} \neq \omega_r'$ *for at most* $n-2$ *agents* $i \in N$, *and*

$m_{i,r} \neq \omega_r'$ *for at most* $n-3$ *agents* $i \in N / \{r'\}$,

**Proof:** We specify $(i^k, m^k)_{k=1}^{2n}$ that is connected with $m$ as follows.

(1) For every $k = 1, ..., 2n$,

$$m^k \in \tilde{M},\ m_{i^k, -i^k}^k = \tilde{\omega}_{-i^k}(m),$$

and for every $m_{i^k}'' \in M_{i^k}$,

$$u_{i^k}(g^{**}(m^k), \omega_{i^k}) + t_{i^k}^{**}(m^k) \geq u_{i^k}(g^{**}(m^k / m_{i^k}''), \omega_{i^k}) + t_{i^k}^{**}(m^k / m_{i^k}'') \text{ if } m^k / m_{i^k}'' \in \tilde{M}.$$

(2) For every $k = 1, ..., n$,

$$i^k = n - k + 1,$$

and for every $k = n+1, ..., 2n$,

$$i^k = 2n - k + 1.$$

If for every $k = 1, ..., 2n$, and every $m_{i^k}'' \in M_{i^k}$ satisfying $m^k / m_{i^k}'' \notin \tilde{M}$,

$$u_{i^k}(g^{**}(m^k), \omega_{i^k}) + t_{i^k}^{**}(m^k) \geq u_{i^k}(g^{**}(m^k / m_{i^k}''), \omega_{i^k}) + t_{i^k}^{**}(m^k / m_{i^k}''),$$

then inequalities (6) hold for every $k = 1, ..., 2n$, and therefore, $m^{2n} = \mu^{**}(\omega)$ is BR-reachable from $m$, which is, however, a contradiction. Hence, there exist $k' \in \{1, ..., 2n\}$ and $m_{i^k}'' \in M_{i^k}$ such that $m^k / m_{i^k}'' \notin \tilde{M}$, inequalities (6) hold for every $k < k'$, and

$$u_{i^k}(g^{**}(m^k),\omega_{i^k})+t_{i^k}^{**}(m^k)<u_{i^k}(g^{**}(m^k/m''_{i^k}),\omega_{i^k})+t_{i^k}^{**}(m^k/m''_{i^k}).$$

Suppose that $k'\notin\{n,2n\}$, i.e., $i^{k'}\neq 1$. Then, there exist $r\in N$, $i\in N/\{i^{k'}\}$, and $i'\in N/\{i,i^{k'}\}$ such that $m_{i,r}^{k'}\neq m_{i',r}^{k'}$. Note that for every $m''_{i^{k'}}\in M_{i^{k'}}$,

$$g^{**}(m^{k'}/m''_{i^{k'}})=d(m_{1,1}^{k'})\text{ and }t_{i^{k'}}^{**}(m^{k'}/m''_{i^{k'}})=0\text{ if }m^{k'}/m''_{i^{k'}}\notin\tilde{M}.$$

Hence, without loss of generality we assume that $m_{i^{k'},r}^{k'}\notin\{m_{i,r}^{k'},m_{i',r}^{k'}\}$. For every $\omega'_r\in\Omega_r$,

$$m_{r'',r}\neq\omega'_r\text{ for at most }n-2\text{ agents }r''\in N,$$

and for every $r'\in N/\{i,i',i^{k'}\}$,

$$m_{r'',r}\neq\omega'_r\text{ for at most }n-3\text{ agents }r''\in N/\{r'\}.$$

This implies that the lemma holds true.

Suppose that $k'\in\{n,2n\}$, i.e., $i^{k'}=1$. Note that

$$m_{i,j}^{k'}=\omega_j\text{ for all }i\in N/\{1\}\text{ and all }j\in N/\{i\}.$$

Hence, $m_{i,1}^{k'}=\omega_1$ for all $i\in N/\{1\}$, and therefore, there exist $r\in N/\{1\}$ and $i\in N/\{1\}$ such that $m_{i,r}^{k'}\neq\omega_r$. Note that for every $m''_{i^{k'}}\in M_{i^{k'}}$,

$$g^{**}(m^{k'}/m''_{i^{k'}})=d(m_{1,1}^{k'})\text{ and }t_{i^{k'}}^{**}(m^{k'}/m''_{i^{k'}})=0\text{ if }m^{k'}/m''_{i^{k'}}\notin\tilde{M}.$$

Hence, without loss of generality, we assume $m_{i^{k'},r}^{k'}\notin\{m_{i,r}^{k'},\omega_r\}$ and $m_{1,1}^{k'}=\omega_1$. For every $\omega'_r\in\Omega_r$,

$$m_{r'',r}\neq\omega'_r\text{ for at most }n-2\text{ agents }r''\in N,$$

and for every $r'\in N/\{i',i^{k'}\}$,

$$m_{r'',r}\neq\omega'_r\text{ for at most }n-3\text{ agents }r''\in N/\{r'\}.$$

This implies that the lemma holds true. **Q.E.D.**

**Lemma D-3:** *For every $m\notin\tilde{M}$, if there exist $r\in N$ and $r'\in N$ such that for every $\omega'_r\in\Omega_r$,*

$$m_{i,r}\neq\omega'_r\text{ for at most }n-2\text{ agents }i\in N\text{, and}$$

$$m_{i,r}\neq\omega'_r\text{ for at most }n-3\text{ agents }i\in N/\{r'\},$$

*then, there exists $m'\notin\tilde{M}$ that is BR-reachable from $m$ such that for every $r\in N$,*

$$m_{r,r}=\omega_r,$$

*for every $r'\in N$,*

$$m_{r',r}\neq\omega_r,$$

*and for every $\omega'_r\in\Omega_r$,*

$$m_{i,r}\neq\omega'_r\text{ for at most }n-3\text{ agents }i\in N/\{r'\}.$$

**Proof:** Without loss of generality, we can assume $r'\neq r$, because there are at least two such agents $r'$ when $n\geq 4$. Let $N(\omega_i,m)\subset N$ denote the set of agents $i'\in N$ satisfying that $m_{i',i}=\omega_i$. We specify $(i^k,m^k)_{k=1}^{2n}$ that is connected with $m$ as follows.

(1) For every $k\in\{1,...,2n\}$,

$m^k_{i^k,i^k} = \omega_{i^k}$ .

(2) For every $k \in \{1,...,n\}$,

$i^k \neq i^{k'}$ for all $k' \in \{1,...,n\}/\{k\}$,

$i^1 = r'$, $i^n = r$,

and for every $i \in \{i^k - 1, i^k + 1\}$,

either $\left|N(m^k_{i^k,i}, m^k)\right| = 1$, or $\left|N(m^k_{i^k,i}, m^k)\right| \leq \left|N(\omega'_i, m^k)\right|$ for all $\omega'_i \in \Omega_i$ .

(3) For every $k \in \{n+1,...,2n\}$,

$i^k = k - n$ .

and for every $i \in \{i^k - 1, i^k + 1\}$,

$m^k_{i^k} \neq \omega_i$ ,

and

either $\left|N(m^k_{i^k,i}, m^k)\right| = 1$, or $\left|N(m^k_{i^k,i}, m^k)\right| \leq \left|N(\omega'_i, m^k)\right|$

for all $\omega'_i \in \Omega_i / \{\omega_i\}$ .

Since $n \geq 4$ and $\left|\Omega_i\right| \geq 4$ for all $i \in N$, it follows that for every $k \in \{1,...,2n\}$,

$m^k / m''_{i^k} \notin \tilde{M}$ for all $m''_{i^k} \in M_{i^k}$, and

$g^+(m^k) = d(\omega_1)$ if $i^k = 1$.

This implies that inequalities (6) hold for every $k \in \{1,...,2n\}$, and therefore, $m^{2n}$ is BR-reachable from $m$. Note from the specification of $(i^k, m^k)^{2n}_{k=1}$ that for every $r \in N$,

$m^{2n}_{r,r} = \omega_r$ ,

for every $r' \in N$,

$m^{2n}_{r',r} \neq \omega_r$ ,

and for every $\omega'_r \in \Omega_r$,

$m^{2n}_{i,r} \neq \omega'_r$ for at most $n-3$ agents $i \in N / \{r'\}$.

**Q.E.D.**

**Lemma D-4:** *For every $m \notin \tilde{M}$, if for every $r \in N$,*

$m_{r,r} = \omega_r$ ,

*for every $r' \in N$,*

$m_{r',r} \neq \omega_r$ ,

*and for every $\omega'_r \in \Omega_r$,*

$m_{i,r} \neq \omega'_r$ *for at most $n-3$ agents $i \in N / \{r'\}$,*

*then there exists $m' \in \tilde{M}$ that is BR-reachable from $m$ such that there exists the single agent $i \in N$ satisfying that $m'_i \neq \mu^{**}_i(\omega)$.*

**Proof:** Choose $j \neq 1$ satisfying that inequality (2) hold. Consider $(i^k, m^k)^{n-1}_{k=1}$ that is connected with $m$ satisfying that for every $k \in \{1,...,n-1\}$,

$m^k_{i^k} = \mu^{**}_{i^k}(\omega)$,

$i^k = j + k$ if $k \in \{1,...,n-4\}$,

$i^{n-3} = j-2$, $i^{n-2} = j-1$, and $i^{n-1} = j$.

Note that for every $k \in \{1,...,n-2\}$,

$$g^{**}(m^k) = d(\omega_1) \text{, and } m^k / m'_k \notin \tilde{M} \text{ for all } m'_k \in M_k ,$$

and therefore, inequalities (6) hold. Note that

$m^{n-1} \in \tilde{M}$,

$m^{n-1}_i = \mu^{**}_i(\omega)$ for all $i \neq j-3$,

$g^{**}(m^{n-1}) = \hat{x}(\omega)$, and $t^{**}_j(m^{n-1}) = \xi$.

Note also that for every $m''_j \neq \mu^{**}_j(\omega)$,

$$t^{**}_j(m^{n-1} / m''_j) = 0,$$

$$g^{**}(m^{n-1} / m''_j) = \hat{x}(\omega) + \frac{\varepsilon}{n}\{\alpha_j(m''_{j,j}) - \alpha_j(\omega_j)\} \text{ if } m''_{j,j-1} = \omega_{j-1},$$

and

$$g^{**}(m^{n-1}) = d(\omega_1) \text{ if } m''_{j,j-1} \neq \omega_{j-1}.$$

These inequalities imply that inequalities (6) hold for $k = n-1$, and therefore, $m^{n-1}$ is BR-reachable from $m$. Since agent $j-3$ is the single deviant from $\mu^{**}(\omega)$, we have proved this lemma.
**Q.E.D.**

**Lemma D-5:** *For every $m \notin \tilde{M}$, if there exists no $m' \in \tilde{M}$ that is BR-reachable from $m$, then there exists $m' \notin \tilde{M}$ that is BR-reachable from $m$ in $(G^{**}, \omega)$ such that there exist $r \in N$ and $r' \in N$ such that for every $\omega'_r \in \Omega_r$,*

*$m_{i,r} \neq \omega'_r$ for at most $n-2$ agents $i \in N$, and*

*$m_{i,r} \neq \omega'_r$ for at most $n-3$ agents $i \in N / \{r'\}$,*

**Proof:** Note that for every $i \neq 1$, and every $m''_i \in M_i$,

$$u_i(g^{**}(m/m''_i), \omega_i) + t^{**}_i(m/m''_i) > u_i(g^{**}(m), \omega_i) + t^{**}_i(m) \text{ if } m/m''_i \in \tilde{M}.$$

Note also that for every $m''_i \in M_i$,

$$(g^{**}(m/m''_i), t^{**}_i(m/m''_i)) = (d(m_{1,1}), 0) \text{ if } m/m''_i \notin \tilde{M}.$$

Hence, there exists the best-reply $m''_i \in M_i$ to $m$ satisfying that for every $j \in N$,

$$\text{either } |N(m''_{i,j}, m^k)| = 1 \text{, or } |N(m''_{i,j}, m^k)| \leq |N(\omega'_j, m^k)| \text{ for all } \omega'_j \in \Omega_i .$$

This implies that there exists $r \in N$ and $r' \in N$ such that for every $\omega'_r \in \Omega_r$,

$m_{i,r} \neq \omega'_r$ for at most $n-2$ agents $i \in N$, and

$m_{i,r} \neq \omega'_r$ for at most $n-3$ agents $i \in N / \{r'\}$.

**Q.E.D.**

By using these lemmata, we will show that $\mu^{**}(\omega)$ is BR-stable in $(G^{**},\omega)$. Note that for every $i \in N$, and every $m_i \neq \mu_i^{**}(\omega)$, if $m_{i,i} \neq \omega_i$, then

$$u_i(g^{**}(\mu^{**}(\omega)),\omega_i) + t_i^{**}(\mu^{**}(\omega)) = u_i(\hat{x}(\omega),\omega_i)$$

$$> u_i(\hat{x}(\omega) + \frac{\varepsilon}{n}\{\alpha_i(m_{i,i}) - \alpha_i(\omega_i)\},\omega_i)$$

$$\geq u_i(g^{**}(\mu^{**}(\omega)/m_i),\omega_i) + t_i^{**}(\mu^{**}(\omega)/m_i),$$

whereas, if $m_{i,i} = \omega_i$, then

$$u_i(g^{**}(\mu^{**}(\omega)),\omega_i) + t_i^{**}(\mu^{**}(\omega)) = u_i(\hat{x}(\omega),\omega_i) > u_i(\hat{x}(\omega),\omega_i) - \xi$$

$$= u_i(g^{**}(\mu^{**}(\omega)/m_i),\omega_i) + t_i^{**}(\mu^{**}(\omega)/m_i).$$

These inequalities imply that $\mu^{**}(\omega)$ is a strict Nash equilibrium in $(G^{**},\omega)$.

Lemma D-1 implies that for every $m \in \tilde{M}$, if there exists the single deviant from $\mu^{**}(\omega)$, then $\mu^{**}(\omega)$ is BR-reachable from $m$. Lemmata D-2, D-3 and D-4 imply that for every $m \in \tilde{M} / \{\mu^{**}(m)\}$, there exists $m' \in \tilde{M}$ that is BR-reachable from $m$ such that there exists the single deviant from $\mu^{**}(\omega)$. Lemmata D-5, D-3, and D-4 imply that for every $m \notin \tilde{M}$, either $\mu^{**}(\omega)$ is BR-reachable from $m$, or there exists $m' \in \tilde{M}$ that is BR-reachable from $m$ such that there exists the single deviant from $\mu^{**}(\omega)$. These observations imply that $\mu^{**}(\omega)$ is BR-reachable from every message profile.

From the above arguments, we have proved that $\mu^{**}(\omega)$ is BR-stable in $(G^{**},\omega)$, and therefore, we have completed the proof of Theorem 7.

## Appendix E

**Proof of Proposition 10:** We can choose a positive real number $b>0$ satisfying that for every $i\in N$, every $m_i\in M_i$, and every $m'\in M$,

$$p_i(m_i\mid m')\geq b \text{ if } p_i(m_i\mid m')>0.$$

Suppose that $m\in M$ is stable in $(G,\omega)$. Choose a positive integer $K$ sufficiently large. For every $m'\in M$, and every $t\geq 1$, there exists $\Gamma(m,t)=(i^k,m^k)_{k=1}^K$ that is connected with $m'$ such that $m^K=m$, and for every $k\in\{1,...,K\}$, inequality (1) holds and $i^k=i(t+k)$. Fix $t\geq 1$ and $(m(\tau))_{\tau=1}^t\in M^t$ arbitrarily, and denote $\Gamma(m(t),t)=(i^k,m^k)_{k=1}^K$. Given the history $(m(\tau))_{\tau=1}^t$ up to period $t$, it holds at least with probability $b^K>0$ that the realized history $(m(\tau))_{\tau=1}^{t+K}$ up to period $t+K$ satisfies that

$$m(t+k)=m^k \text{ for all } k\in\{1,...,K\},$$

and therefore,

$$m(t+K)=m.$$

This implies that agents almost certainly come to continue announcing $m$ in the long run, and therefore, $m$ is the long-run behavior in $G$ with alternating play with respect to $p$.

Suppose that $m$ is the long-run behavior in $G$ with alternating play with respect to $p$. Note from the definition of the long-run behavior that once agents announce $m$, they continue announcing it forever. This implies that for every $i\in N$, there exists no other message that is better than or indifferent from $m_i$, provided that the other agents announce $m_{-i}$. Hence, $m$ is a strict Nash equilibrium, and therefore, there exists no other message profile that is reachable from $m$. Note also from the definition of the long-run behavior that for every $m(1)\in M$, there exists a positive integer $K>0$ and a history $(m(t))_{t=1}^{K+1}$ up to period $K+1$ such that $m(K+1)=m$, and for every $t\in\{2,...,K+1\}$,

$$p_{i(t)}(m_{i(t)}(t)\mid m(t-1))>0, \text{ and}$$

$$m_i(t)=m_i(t-1) \text{ for all } i\neq i(t).$$

This implies that there exists $(i^k,m^k)_{k=1}^K$ connected with $m(1)$ such that $m^K=m$, and for every $k\in\{1,...,K\}$, inequality (1) holds, where $m^0=m(1)$. Hence, $m$ is reachable from every message profile, and therefore, we have proved that $m$ is stable in $(G,\omega)$.
**Q.E.D.**

**Proof of Proposition 12:** We can choose a positive real number $b>0$ satisfying that for every $i\in N$, every $m_i\in M_i$, and every $m'\in M$,

$$p_i(m_i\mid m')\geq b \text{ if } p_i(m_i\mid m')>0.$$

Suppose that $m\in M$ is weakly stable in $(G,\omega)$. Choose a positive integer $K$ sufficiently large. For every $m'\in M$, there exists $\Gamma(m)=(m^k)_{k=1}^K$ such that $m^K=m$, and for every $k\in\{1,...,K\}$, and every $i\in N$, inequality (5) holds, where $m^0=m'$. Fix

$t \geq 1$ and $(m(\tau))_{\tau=1}^{t} \in M^t$ arbitrarily, and denote $\Gamma(m(t)) = (m^k)_{k=1}^{K}$. Given the history $(m(\tau))_{\tau=1}^{t}$ up to period $t$, it holds at least with probability $b^{nK} > 0$ that the realized history $(m(\tau))_{\tau=1}^{t+K}$ up to period $t+K$ satisfies that

$$m(t+k) = m^k \text{ for all } k \in \{1,...,K\},$$

and therefore,

$$m(t+K) = m.$$

Hence, agents almost certainly come to continue announcing $\mu(\omega)$ in the long run, and therefore, $m$ is the long-run behavior in $G$ with simultaneous play with respect to $p$.

Suppose that $m$ is the long-run behavior in $G$ with simultaneous play with respect to $p$. In the same way as in the proof of Proposition 10, there exists no other message profile that is weakly reachable from $m$. Note that for every $m(1) \in M$, there exists a positive integer $K > 0$ and a history $(m(t))_{t=1}^{K+1}$ up to period $K+1$ such that $m(K+1) = m$, and for every $t \in \{2,...,K+1\}$,

$$p_i(m_i(t) \mid m(t-1)) > 0 \text{ for all } i \in N.$$

Hence, it follows that there exists a finite sequence $(m^k)_{k=1}^{K}$ such that $m^K = m$, and for every $k \in \{1,...,K\}$, and every $i \in N$, inequality (5) holds, where $m^0 = m(1)$. Hence, $m$ is weakly reachable from every message profile, and therefore, we have proved that $m$ is weakly stable in $(G,\omega)$.

**Q.E.D.**