UTMD Working Paper
The University of Tokyo
Market Design Center

UTMD-036

Honesty and Epistemological Implementation of Social Choice Functions with Asymmetric Information

Hitoshi Matsushima
The University of Tokyo

First Version: July 20, 2021
This Version: November 20, 2022

Honesty and Epistemological Implementation of Social Choice Functions with Asymmetric Information[1]

Hitoshi Matsushima[2]

Department of Economics, University of Tokyo,

Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

First Version: July 20, 2021

This Version: November 20, 2022

Abstract

We investigate the implementation of social choice functions with asymmetric information concerning the state from an epistemological perspective. Although agents are either selfish or honest, they do not expect other participants to be honest. However, an honest agent may exist not among participants but in their higher-order beliefs. We assume that "all agents are selfish" never happens to be common knowledge. We show a positive result in general asymmetric information environments, demonstrating that with a minor restriction on signal correlation called information diversity, any incentive-compatible social choice function, whether ethical or nonethical, is uniquely implementable in the Bayesian Nash equilibrium.

Keywords: unique implementation, asymmetric information, honesty, no common knowledge on selfishness, information diversity
JEL Classification Numbers: C72, D71, D78, H41

---

[1] An earlier version of this study was presented as Section 5 of the discussion paper (CARF-F-518, University of Tokyo, 2021) entitled "Epistemological Implementation of Social Choice Functions" by Hitoshi Matsushima (https://www.carf.e.u-tokyo.ac.jp/admin/wp-content/uploads/2021/07/F518.pdf).

[2] Corresponding author; E-mail: hitoshi@e.u-tokyo.ac.jp

1. Introduction

In this study, we examine the implementation problem of social choice functions (SCFs) with asymmetric information. A central planner attempts to implement a desirable allocation implied by an SCF, contingent on the state. The central planner does not know the state, unlike multiple agents (participants) who are partly and privately informed of it. Although they receive only inadequate information about the state, combining their information can provide the central planner with a complete picture of the state. The central planner seeks to hear from these agents about their private signals, but they might lie and manipulate the information to render the central planner's decisions more beneficial for them. Hence, the central planner attempts to design a decentralized mechanism that comprises message spaces, an allocation rule, and a payment rule and incentivizes these agents to announce sincerely. This study clarifies the conditions under which the central planner can implement the SCF despite agents' potential manipulation.

The SCF must be incentive-compatible, that is, require truthful revelation, to be a Bayesian Nash equilibrium (BNE) in the direct revelation game associated with the SCF. However, this is not enough because, depending on the specifications of the underlying information structures, unwanted equilibria that fail to achieve the SCF values may exist. We clarify a range of asymmetric information environments in which any incentive-compatible SCF is uniquely implementable in a BNE; that is, a mechanism that has a unique BNE exists and this unique BNE correctly achieves the SCF values. Moreover, we demonstrate that this range is quite wide once we consider the epistemological possibility that an agent is not necessarily selfish.

We assume that each agent is either selfish or honest. A selfish agent is only concerned about their material utility, that is, the utility derived directly from the central planner's decisions. By contrast, an honest agent is only concerned about the intrinsic preference for honesty regarding their attitude in situations in which they announce messages about the state. However, we do not assume that an honest agent exists as a participant in the central planner's problem. Instead, we consider the epistemological possibility that an honest agent exists not in the mechanism but in the participants' higher-order beliefs, regarding which, the other agent types are between honest and

selfish. Hence, we assume incomplete information concerning the agents' epistemological types, as well as incomplete (asymmetric) information concerning the state.

We also assume only a slight possibility of an honest agent in higher-order beliefs, that is, we assume "no common knowledge of selfishness" (NCKS) in the sense that "all agents are selfish" never happens to be common knowledge. Agents do not necessarily expect the possibility of the existence of an honest participant; they may have mutual knowledge that all agents are selfish (i.e., all agents know that all agents are selfish). We show that despite these weaknesses in honesty requirements, the central planner can elicit correct information from all agents if the epistemology satisfies NCKS. With this finding, we can demonstrate a positive result such that in a wide range of asymmetric information environments, any incentive-compatible SCF, whether ethical or nonethical, is uniquely implementable in a BNE.

The early literature on implementation theory assumes that "all agents are selfish" is common knowledge. Under this assumption, ethical SCFs, that is, SCFs that consider social factors such as ethics and fairness unrelated to participants' selfish motives, are excluded from consideration (Arrow, 1951; Hurwicz, 1972; Gibbard, 1973; Satterthwaite, 1975; Maskin, 1977/1999; Abreu and Matsushima, 1992a; 1992b).[3] [4] Recent progress in this literature indicates that by excluding such common knowledge of selfishness from considerations, the scope of implementable SCFs can be greatly expanded. Matsushima (2008a) considers the possibility that agents are not purely selfish but honest as well, and then show a positive result, illustrating that with complete (i.e., symmetric) information concerning the state, any SCF, whether ethical or nonethical, is uniquely implementable if such honest agents exist in the mechanism.

Starting with Matsushima (2008a), many subsequent studies including Matsushima (2008b), Dutta and Sen (2012), Matsushima (2013), Kartik, Tercieux, and Holden (2014), Saporiti (2014), Ortner (2015), Mukherjee, Muto, and Ramaekers (2017), Yadav (2016), Lombardi and Yoshihara (2018), Dogan (2017), and Savva

[3] Arrow (1951) considers the social choice theory as a problem of preferences aggregation, which inevitably excludes some aspects of ethics and fairness concerns.
[4] An exception is Matsushima (2021), who assumes that the state is ex-post verifiable with a positive probability and showed that ethical SCFs are implementable.

(2018) show positive results. As a particularly important contribution, Matsushima (2022a) introduces the epistemological type space and demonstrate that under complete information environments concerning the state, any SCF is uniquely implementable in a BNE with three or more agents if NCKS holds.

The main result of this study is the generalization of Matsushima (2022a) from complete information to asymmetric information. Instead of assuming information symmetry (i.e., complete information), we assume information diversity (ID), implying that no private signal changes the prior in the same direction as any other private signal does. We then show that with ID and NCKS, any incentive-compatible SCF, whether ethical or nonethical, is uniquely implementable in a BNE.

ID is a weak restriction on private signal correlation: it is much weaker than any informational restriction discussed in the implementation literature, such as Bayesian monotonicity (Jackson, 1991), measurability (Abreu and Matsushima, 1992b), non-consistent deception (Matsushima, 1993), and various dimensionality requirements. Unlike these informational restrictions, ID does not require any relationship between the state and the agents' material utilities. Therefore, ID can render any ethical SCF that meets incentive compatibility uniquely implementable.

Matsushima (2008b) is the first and only one to investigate asymmetric information environments with honest agents. Matsushima (2008b) demonstrates that even under asymmetric information, any incentive-compatible SCF is uniquely implementable in a BNE if all agents consider honesty. By contrast, we do not assume the possibility that such an honest agent exists as a participant in the central planner's problem.

Similar to Matsushima (2022a), in this study, we require each agent to announce probability distributions over private signals (not pointwise private signals) and utilize a quadratic scoring rule (Brier, 1950) that aligns agents' payoffs with the distance between their messages. The quadratic scoring rule plays a significant role in incentivizing selfish agents to announce sincerely in information elicitations. However, our generalization from complete information to asymmetric information is not straightforward, and we need an additional mechanism design device. In a complete information environment, we can utilize a simple form of a quadratic scoring rule with single announcements. The usefulness of this simple form depends substantially on the

complete information setting. To prove the positive result in general asymmetric information environments, we need a more elaborate design method with multiple messages that have a nested structure of the quadratic scoring rule and its variants. This method can induce agents to gradually reveal their private signals through multiple announcements. We need multiple message announcements because, unlike complete information, no one else knows the same thing, which dissuades each agent from being honest on the first attempt. By inducing agents to speak a lot, we can gradually expand the selfish types that have an incentive to be honest (at least virtually).

This study's scientific contribution lies in proving that all incentive compatible SCFs, whether ethical or nonethical, are uniquely implementable in a BNE even under asymmetric information if we eliminate the common knowledge of selfishness. By adopting the new design method with multiple announcements, the central planner succeeds in extracting correct information about ethics and fairness that is scattered and buried among agents who are not interested in it due to personal motives for allocations.

Abeler, Nosenzo, and Raymond (2019) empirically and experimentally show that subjects who trade-off material interest against honesty forego a lot of potential benefits from lying, such as those derived from adversarial motives. Their report supports the validity of this study's assumptions.

The remainder of this paper is organized as follows. Section 2 presents the implementation problem, ID, NCKS, and the main theorem (Theorem 1). Section 3 considers the information elicitation problem as a special case, demonstrates the nested quadratic scoring rule design, and illustrates the proof of Theorem 1 for this case (Proposition 1). Section 4 presents the proof of Theorem 1. Section 5 discusses weak honesty. Section 6 concludes.

2. The Model

We investigate a situation in which a central planner attempts to achieve a desirable allocation contingent on the state as follows. Let $N \equiv \{1,...,n\}$ denote a finite set of all agents, where $n \geq 2$. Let A denote the non-empty and finite set of all allocations. Let Ω denote a non-empty and finite set of states. The SCF is defined as

$f:\Omega\to\Delta(A)$.[5] For every $\omega\in\Omega$, $f(\omega)\in\Delta(A)$ implies a desirable allocation distribution at state ω.[6] The central planner does not know the state.

We consider asymmetric information environments concerning the state where each agent is only partly and privately informed of the state. Let

$$\Omega=\underset{i\in N}{\times}\Omega_i,\ \ \omega_i\in\Omega_i\text{, and }\ \omega=(\omega_i)_{i\in N}\in\underset{i\in N}{\times}\Omega_i.$$

Each agent $i\in N$ is informed of the $i-th$ component ω_i as their private signal. For each $j\neq i$, let $p_{i,j}(\cdot\,|\,\omega_i):\Omega_j\to[0,1]$ denote the probability distribution on Ω_j conditional on $\omega_i\in\Omega_i$. We assume that a common prior $p:\Omega\to[0,1]$ exists from which $p_{i,j}(\cdot\,|\,\omega_i)$ is derived. Let $p_i:\Omega_i\to[0,1]$ denote the prior distribution over Ω_i, where $p_i(\omega_i)\equiv\sum_{\omega_{-i}\in\Omega_{-i}}p(\omega)$ for all $\omega_i\in\Omega_i$.

ID: For every $i\in N$, $j\in N\setminus\{i\}$, $\omega_i\in\Omega_i$, and $\omega_i'\neq\omega_i$, no $\beta\geq 0$ exists such that

$$p_{i,j}(\cdot\,|\,\omega_i)-p_j(\cdot)=\beta\left\{p_{i,j}(\cdot\,|\,\omega_i')-p_j(\cdot)\right\}.$$

ID implies that no private signal changes the prior in the same direction as any other private signal does. ID excludes the case in which private signals are independent of each other. Hence, ID implies that $p_{i,j}(\cdot\,|\,\omega_i)\neq p_j(\cdot)$ and $p_{i,j}(\cdot\,|\,\omega_i)\neq p_{i,j}(\cdot\,|\,\omega_i')$ for all $j\neq i$. However, ID permits each private signal ω_i to have a mixture on Ω_i and $\alpha_i\in\Delta(\Omega_i)\setminus\{\omega_i\}$, which brings the same posterior as ω_i does; that is, $p_{i,j}(\cdot\,|\,\alpha_i)=p_{i,j}(\cdot\,|\,\omega_i)$, where we denote $p_{i,j}(\cdot\,|\,\alpha_i)\equiv\sum_{\omega_i\in\Omega_i}p_{i,j}(\cdot\,|\,\omega_i)\alpha_i(\omega_i)$. Hence, ID can be considered a very weak restriction on private-signal correlations.

We assume that each agent is either selfish or honest. No agent knows whether the other agents are selfish or honest. Following Matsushima (2022a), to describe agents'

---

[5] $\Delta(Z)$ denotes the space of probability measures on the Borel field of a measurable space Z. If Z is finite and $\rho\in\Delta(Z)$ satisfies $\rho(z)=1$ for some $z\in Z$, we simply write $\rho=z$.

[6] We consider both deterministic and stochastic SCFs.

higher-order beliefs about their selfishness and honesty, we define an epistemological type space separately from the private signal space:

$$\Gamma \equiv (T_i, \pi_i, \theta_i)_{i \in N},$$

where $t_i \in T_i$ is the agent $i's$ epistemological type, $\pi_i : T_i \to \Delta(T_{-i})$, and $\theta_i : T_i \to \{0,1\}$.[7] The agent i is selfish (honest) if $\theta_i(t_i) = 0$ ($\theta_i(t_i) = 1$). The agent i expects that the epistemological types of other agents are distributed according to the probability measure $\pi_i(t_i) = \pi_i(\cdot \mid t_i) \in \Delta(T_{-i})$. We assume that a common prior $\pi \in \Delta(T)$ exists from which $(\pi_i)_{i \in N}$ is derived, where $T \equiv \times_{i \in N} T_i$. We assume that T_i is non-empty and finite for each $i \in N$, and that ω and t are independently drawn.

We call a subset of epistemological type profiles $E \subset T$ an event. We write $\pi_i(E \mid t_i) \equiv \pi_i(E_{-i}(t_i) \mid t_i)$ for convenience, where we denote $E_{-i}(t_i) \equiv \{t_{-i} \in T_{-i} \mid (t_i, t_{-i}) \in E\}$. Let $E^* \subset T$ denote the event in which all agents are selfish, that is,

$$E^* \equiv \{t \in T \mid \forall i \in N : \theta_i(t_i) = 0\}.$$

For each agent $i \in N$, we define the set of all selfish types as

$$E_i^* \equiv \{t_i \in T_i \mid \theta_i(t_i) = 0\}.$$

Consider an arbitrary event $E \subset T$. Let

$$V_i^1(E) = \{t_i \in T_i \mid \pi_i(E \mid t_i) = 1\},$$

which denotes the set of agent $i's$ types who know the occurrence of E. Let

$$V_i^2(E) = \{t_i \in T_i \mid \pi_i(\underset{j \in N}{\times} V_j^1(E) \mid t_i) = 1\},$$

which denotes the set of agent $i's$ types who know the occurrence of $\underset{j \in N}{\times} V_j^1(E)$, that is, are aware that all agents know the occurrence of E. Recursively, for each positive integer $h \geq 2$, let

$$V_i^h(E) = \{t_i \in T_i \mid \pi_i(\underset{j \in N}{\times} V_j^{h-1}(E) \mid t_i) = 1\},$$

---

[7] We denote $Z \equiv \underset{i \in N}{\times} Z_i$, $Z_{-i} \equiv \underset{j \neq i}{\times} Z_j$, $z = (z_i)_{i \in N} \in Z$, and $z_{-i} = (z_j)_{j \neq i} \in Z_{-i}$.

which denotes the set of agent $i's$ types who know the occurrence of $\underset{j\in N}{\times} V_j^{h-1}(E)$, that is, are aware that all agents know the occurrence of $\underset{j\in N}{\times} V_j^{h-2}(E)$. We then define

$$V_i^{\infty}(E) \equiv \bigcap_{h=1}^{\infty} V_i^h(E).$$

An event $E \subset T$ is said to be *common knowledge* in an epistemological type profile $t \in T$ if $t \in \underset{i\in N}{\times} V_i^{\infty}(E)$. Note that if E is common knowledge at $t \in T$, then

$$\pi_i\left(\underset{j\in N}{\times} V_j^{\infty}(E) \middle| t_i \right) = 1 \text{ for all } i \in N.$$

We explain later that whether the event of "all agents are selfish" (i.e., $E = E^*$) is common knowledge has a decisive impact on the implementability of SCFs.

NCKS: "All agents are selfish" never happens to be common knowledge:

$$\underset{i\in N}{\times} V_i^{\infty}(E^*) = \phi.$$

Fix an arbitrary positive but small real number $\varepsilon > 0$ as agents' limited liability. We define a mechanism as $G \equiv (M, g, x)$, where $M = \underset{i\in N}{\times} M_i$, M_i denotes agent $i's$ message space, $g : M \to \Delta(A)$ denotes an allocation rule, $x = (x_i)_{i\in N}$ a payment rule, and $x_i : M \to [-\varepsilon, \varepsilon]$ the payment rule for the agent i (with limited liability ε). Each agent i simultaneously announces a message $m_i \in M_i$. The central planner then determines the allocation according to $g(m) \in \Delta(A)$ and pays a monetary amount $x_i(m) \in R$ to each agent i.

Each agent $i's$ material payoff at state $\omega \in \Omega$ is given by $v_i(a, \omega) + r_i$, provided the central planner determines the allocation $a \in A$ and offers the monetary amount $r_i \in R$ to the agent i. If agents announce $m \in M$, then the resultant expected material payoff is given by $v_i(f(m), \omega) + x_i(m)$, where we denote $v_i(\alpha, \omega) \equiv \sum_{a\in A} v_i(a, \omega)\alpha(a)$.

We consider a mechanism (M,g,x) in which a positive integer L exists such that

$$M_i = \underset{l=1}{\overset{L}{\times}} M_i^l \quad \text{for each} \quad i \in N .$$

From a semantic perspective, we focus on mechanisms (M,g,x) such that

$$\Omega_i \subset M_i^l \subset \Delta(\Omega_i) \quad \text{for all} \quad l \in \{1,...,L\} .$$

Each agent i announces L sub-messages at once. At each $l-th$ sub-message, the agent i announces a probability distribution over the private signal $m_i^l \in \Delta(\Omega_i)$. At each of the agent $i's$ sub-messages, we assume the agent i announces more truthfully, as their announcement at this sub-message grants a greater probability to the true private signal. An agent can announce different distributions across sub-messages.

Denote $m_i^l = (m_i^l(\omega_i))_{\omega_i \in \Omega_i} \in \Delta(\Omega_i)$. At each $l-th$ sub-message, the agent i announces that each private signal $\omega_i \in \Omega_i$ occurs with the probability of $m_i^l(\omega_i) \in [0,1]$. We simply write $m_i^l = \omega_i$ if $m_i^l(\omega_i) = 1$. We also denote $m_i(\omega_i) = (m_i^l(\omega_i))_{l=1}^L \in [0,1]^L$. We consider the agent i with a private signal ω_i acting more honestly when they announce m_i rather than $\tilde{m}_i$, if the vector m_i assigns a higher probability to the true private signal than the vector $\tilde{m}_i$ in each component, that is,

$$m_i(\omega_i) \neq \tilde{m}_i(\omega_i) \quad \text{and} \quad m_i(\omega_i) \geq \tilde{m}_i(\omega_i) .^{8}$$

We define a strategy for the agent i as

$$s_i : \Omega_i \times T_i \to M_i ,$$

according to which, the agent i with private signal ω_i and epistemological type t_i announces $m_i = s_i(\omega_i, t_i) \in M_i$. Denote $s_i = (s_i^l)_{l=1}^L$, $s_i^l : \Omega_i \times T_i \to M_i^l$, and $s_i(\omega_i, t_i) = (s_i^l(\omega_i, t_i))_{l=1}^L$, where $s_i^l(\omega_i, t_i)$ represents the agent $i's$ $l-th$ sub-message. We also denote $s_i(\omega_i, t_i)(\omega_i') = (s_i^l(\omega_i, t_i)(\omega_i'))_{l=1}^L \in [0,1]^L$, where the agent i

---

[8] For two vectors $z = (z^l)_{l=1}^L$ and $z = (z^l)_{l=1}^L$, we write $z \geq \tilde{z}$ if and only if $z^l \geq \tilde{z}^l$ for all $l \in \{1,...,L\}$.

with a private signal ω_i and epistemological type t_i announces as their $l-th$ sub-message that each private signal $\omega_i' \in \Omega_i$ occurs with the probability of $s_i^l(\omega_i, t_i)(\omega_i') \in [0,1]$. We simply write $s_i^l(\omega_i, t_i) = \omega_i'$ if $s_i^l(\omega_i, t_i)(\omega_i') = 1$.

Each agent $i \in N$ is either selfish ($\theta_i(t_i) = 0$) or honest ($\theta_i(t_i) = 1$). An honest agent ($\theta_i(t_i) = 1$) always announces truthfully, whereas a selfish agent maximizes their expected material payoff:

$$[\theta_i(t_i) = 0]$$
$$\Rightarrow [s_i(\omega_i, t_i) \in \arg\max_{m_i \in M_i} E[v_i(g(m), \omega) + x_i(m) \mid \omega_i, t_i, s_{-i}]],$$

where we assume that the other agents announce according to $s_{-i} = (s_j)_{j \neq i}$.[9] A strategy profile s is said to be a BNE in the mechanism G if for every $i \in N$, $\omega_i \in \Omega_i$, and $t_i \in T_i$,

$$s_i^l(\omega_i, t_i) = \omega_i \qquad \text{if } \theta_i(t_i) = 1,$$

and

$$E[v_i(g(s_i(\omega_i, t_i), m_{-i}), \omega) + x_i(s_i(\omega_i, t_i), m_{-i}) \mid \omega_i, t_i, s_{-i}]$$
$$\geq E[v_i(g(m_i, m_{-i}), \omega) + x_i(m_i, m_{-i}) \mid \omega_i, t_i, s_{-i}] \text{ for all } m_i \in M_i$$
$$\text{if } \theta_i(t_i) = 0.$$

A mechanism G is said to *uniquely implement* an SCF f if a unique BNE s exists, and s satisfies

$$g(s(\omega, t)) = f(\omega) \text{ for all } \omega \in \Omega \text{ and } t \in T,$$

where we denote $s(\omega, t) \equiv (s_i(\omega_i, t_i))_{i \in N}$. An SCF is said to be *uniquely implementable* if a mechanism that uniquely implements it exists. An SCF f is said to be *incentive-compatible* if for every $i \in N$ and $\omega_i \in \Omega_i$,

$$E[v_i(f(\omega)) \mid \omega_i] \geq E[v_i(f(\omega_i', \omega_{-i})) \mid \omega_i] \text{ for all } \omega_i' \in \Omega_i.$$

---

[9] $E[\cdot \mid \xi]$ denotes the expectation operator conditional on ξ.

Theorem 1: *An incentive-compatible SCF* f *is uniquely implementable if NCKS and ID hold.*

3. Special Case: Information Elicitation

To understand how to prove Theorem 1, investigating the information elicitation problem as a special case, where each agent's material payoff is irrelevant to the allocation, is helpful, that is,

$$v_i(a,\omega)=0 \quad \text{for all} \quad i\in N,\ a\in A, \text{ and } \omega\in\Omega.$$

Note that any SCF satisfies incentive compatibility in this information elicitation problem. However, we have a serious multiplicity of unwanted equilibria. As each agent's material payoff is independent of the state, we cannot elicit correct information from an agent by relying solely on the agent's selfish motives. Therefore, honesty is expected to resolve the incentive issue in terms of uniqueness. Moreover, in the information elicitation problem, each agent's material payoff is independent of the allocation as well. Hence, an SCF can be interpreted as primarily related to the welfare of citizens in a society other than these agents.

Proposition 1: *In the information elicitation problem, any SCF* f *is uniquely implementable if ID and NCKS hold.*

3.1. Mechanism Design

To prove Proposition 1, we design a mechanism $G=(M,g,x)$ as follows. Fix an arbitrary positive integer H, which is set sufficiently large. Let

$$L=(n-1)H.$$

We denote (j,h) and $M_{i,j}^{h}$ for l and M_i^l, respectively, where we denote

$$l=(n-1)(h-1)+j \qquad \text{if} \quad j<i,$$

and

$$l=(n-1)(h-1)+j-1 \qquad \text{if} \quad j>i.$$

Hence, we write

$$M_i = \underset{h=1}{\overset{H}{\times}} M_i^h \quad \text{and} \quad M_i^h = \underset{j \neq i}{\times} M_{i,j}^h .$$

While each agent i simultaneously announces multiple (i.e., $L = (n-1)H$) sub-messages, we call $m_i^h = (m_{i,j}^h)_{j \neq i} \in M_i^h$ the h-th sub-message and $m_{i,j}^h \in M_{i,j}^h$ its sub-sub-message for convenience. That is, each agent i announces H messages, each of which consists of $n-1$ sub-sub-messages. We then specify

$$M_{i,j}^h = \{\alpha_i \in \Delta(\Omega_i) \mid \exists(\omega_i, \lambda) \in \Omega_i \times [0,1] : \alpha_i = \lambda\omega_i + (1-\lambda)p_i\} .$$

Note that $\Omega_i \subset M_{i,j}^h \subset \Delta(\Omega_i)$ and for each $m_{i,j}^h \in M_{i,j}^h \setminus \{p_i\}$, $(\omega_i, \lambda) \in \Omega_i \times (0,1]$ uniquely exists such that $m_{i,j}^h = \lambda\omega_i + (1-\lambda)p_i$. Hence, we can define $I_i : \Delta(\Omega_i) \to \Delta(\Omega_i)$ as follows:

$$I_i(\alpha_i) = \omega_i \qquad \text{if } \alpha_i = \lambda\omega_i + (1-\lambda)p_i \text{ for some } \lambda \in (0,1],$$

and

$$I_i(\alpha_i) = p_i \qquad \text{if no such } \omega_i \text{ exists.}$$

Note that $I_i(\alpha_i) = \omega_i$ implies that the announcement of α_i (virtually) reveals a private signal ω_i, whereas $I_i(\alpha_i) = p_i$ implies that the announcement of α_i reveals nothing. For every $\alpha_i \in M_{i,j}^h$, we have

$$[\, I_i(\alpha_i) = p_i \,] \Leftrightarrow [\, \alpha_i = p_i \,],$$

and that whenever $m_{i,j}^h \neq p_i$ occurs, the announcement of $m_{i,j}^h$ reveals some private signal, whether true or not.

For each $h \in \{1, ..., H\}$, we define $I_i^h : M_i^h \to \Delta(\Omega_i)$ as follows:

$$I_i^h(m_i^h) = \omega_i \qquad \text{if } I_i(m_{i,j}^h) = \omega_i \text{ for some } j \neq i \text{ and}$$

$$I_i(m_{i,j}^h) \in \{\omega_i, p_i\} \text{ for all } j \neq i,$$

and

$$I_i^h(m_i^h) = p_i \qquad \text{if no such } \omega_i \text{ exists.}$$

Note that $I_i^h(m_i^h) = \omega_i$ implies that an h-th sub-sub-message of the agent i exists that reveals ω_i and no other h-th sub-sub-message of the agent i exists that reveals a

different private signal. In this case, the agent i is considered to (virtually) reveal ω_i in their h-th sub-message. Otherwise, the agent i is considered to reveal nothing in their h-th sub-message (i.e., $I_i^h(m_i^h) = p_i$).

Fix an arbitrary allocation $a^* \in A$ as the default. We specify the allocation rule g as follows: for every $m \in M$,

$$g(m) = f(\omega) \qquad \text{if } I_i^H(s_i^H(\omega_i, t_i)) = \omega_i \text{ for all } i \in N,$$

and

$$g(m) = a^* \qquad \text{if no such } \omega \text{ exists.}$$

The central planner selects the allocation according to $f(\omega)$ if every agent i reveals ω_i in the H-th (final) sub-message. Otherwise, they select the default a^*.

We define $\gamma_i : \Delta(\Omega_i)^2 \to R$ as a quadratic scoring rule:

$$\gamma_i(\alpha_i, \alpha_i') = -\sum_{\omega_i \in \Omega_i} \{\alpha_i(\omega_i) - \alpha_i'(\omega_i)\}^2 .$$

Note that $\alpha_i = \alpha_i'$ uniquely maximizes $\gamma_i(\alpha_i, \alpha_i')$. We further define $\gamma_{i,j} : \Delta(\Omega_i) \times \Delta(\Omega_j) \to R$ as a variant of the following quadratic scoring rule:

$$\gamma_{i,j}(\alpha_i, \alpha_j) = \gamma_j(p_{i,j}(\cdot \mid \alpha_i), \alpha_j).$$

Note that α_i uniquely maximizes $\gamma_{i,j}(\alpha_i, \alpha_j)$ whenever $p_{i,j}(\cdot \mid \alpha_i) = \alpha_j(\cdot)$. We specify the payment rule x: for every $m \in M$ and $i \in N$,

$$x_i(m) = \frac{\varepsilon}{(n-1)H} \sum_{j \neq i} \left[\gamma_i(m_{i,j}^1, p_i) + \sum_{h=2}^{H} \gamma_{i,j}(m_{i,j}^h, I_j^{h-1}(m_j^{h-1})) \right].$$

Clearly, the specified x satisfies the limited liability (i.e., $-\varepsilon \le x_i(m) \le \varepsilon$ for all $i \in N$ and $m \in M$).

3.2. Proof of Proposition 1

We show, as follows, that the mechanism (M, g, x) designed in Subsection 3.1 uniquely implements the SCF f in the information elicitation problem. By definition, any honest agent announces their private signal truthfully at every sub-message. Given

the quadratic scoring rules $\gamma_i(m^1_{i,j}, p_i)$, each selfish agent i announces $m^1_{i,j} = p_i$ for all $j \neq i$, that is, reveals nothing at the first sub-message. However, at each of the subsequent sub-messages, any selfish agent who expects another agent to (at least virtually) reveal truthfully with a positive probability at the previous sub-message is willing to (virtually) reveal their private signal truthfully. This property is derived from the application of the variants of the quadratic scoring rule $\gamma_{i,j}(m^h_{i,j}, I^{h-1}_j(m^{h-1}_j))$, along with ID. Given the nested structure of the quadratic scoring rule and its variant, the range of selfish agents who reveal truthfully expands as the sub-message becomes later. Moreover, given NCKS, at the final sub-message, no agent exists who reveals nothing, and every agent is willing to reveal their private signal truthfully.

Example: Let $n = 2$, $\Omega_1 = \Omega_2 = \{1,2\}$, $T_1 = \{1,2\}$, $T_2 = \{1,2,3\}$, $\theta_i(1) = \theta_i(2) = 0$ for each $i \in \{1,2\}$, and $\theta_2(3) = 1$. The common priors, $p(\omega)$ and $\theta(t)$, are presented in Figures 1 and 2, respectively. Note $p_1(1) = 3/10$, $p_2(\omega_2) = 2/5$, $V^1_1(E^*) = \{2\}$, $V^1_2(E^*) = \{1,2,3\}$, $V^2_1(E^*) = \{2\}$, $V^2_2(E^*) = \phi$, $V^3_1(E^*) = \phi$, and $V^3_2(E^*) = \phi$.

Figure 1: $p(\omega_1, \omega_2)$

	1	1
1	1/10	1/5
2	3/10	2/5

Figure 2: $\theta(t_1, t_2)$

	1	2	3
1	0	1/4	1/4
2	1/4	1/4	0

Let $H = 4$. As type $t_2 = 3$ is honest, we have

$$m^h_{2,1} = \omega_2 \text{ for all } h \in \{1,...,4\} \text{ if } t_2 = 3.$$

As any other type is selfish, according to the maximization of $\gamma_i(m^1_{i,j}, p_i)$, the first sub-message must satisfy

$$m^1_{i,j} = p_i \quad \text{if} \quad t_i \neq 3 .$$

Consider the second sub-message announcement by type t_i, where we assume that this type is selfish (i.e., $t_i \neq 3$). If $t_i \in V^1_i(E^*)$, then the agent knows that the other agent reveals nothing at the first sub-message announcement. Hence, according to the expected value maximization of $\gamma_{i,j}(m^2_{i,j}, I^1_j(m^1_j)) = \gamma_{i,j}(m^2_{i,j}, p_j)$, their second message must satisfy

$$m^2_{i,j} = p_i \quad \text{if} \quad t_i \in V^1_i(E^*) .$$

If $t_i \notin V^1_i(E^*)$, then they know that with a positive probability, the other agent is honest and reveals their private signal truthfully at the first sub-message announcement. Hence, according to the maximization of the expected value of $\gamma_{i,j}(m^2_{i,j}, I^1_j(m^1_j))$, their second message must satisfy

$$m^2_{i,j} = \lambda\omega_i + (1-\lambda)p_i \quad \text{if} \quad t_i \notin V^1_i(E^*) ,$$

where λ is the conditional probability that the other agent is honest. Hence, type $t_i \notin V^1_i(E^*)$ (virtually) reveals their private signal

$$I^2_i(m^2_i) = \omega_i \qquad \text{if} \quad t_i \notin V^1_i(E^*), \text{ that is, } \quad i = 1 \quad \text{and} \quad t_i = 1 ,$$

whereas

$$I^2_i(m^2_i) = p_i \qquad \text{if} \quad t_i \in V^1_i(E^*) .$$

Consider the third sub-message announcement by type t_i. Similar to the above, we have

$$I^3_i(m^3_i) = \omega_i \qquad \text{if} \quad t_i \notin V^2_i(E^*), \text{ that is, } \quad i = 1 \quad \text{and} \quad t_i = 1 ,$$

whereas

$$I^3_i(m^3_i) = p_i \qquad \text{if} \quad t_i \in V^2_i(E^*), \text{ that is, } \quad i = 1 \quad \text{and} \quad t_i = 2 .$$

Similarly, we have

$$I^4_i(m^4_i) = \omega_i \qquad \text{if} \quad t_i \notin V^3_i(E^*) ,$$

whereas

$$I_i^4(m_i^4) = p_i \qquad \text{if } t_i \in V_i^3(E^*).$$

As $V_1^3(E^*) = V_2^3(E^*) = \phi$, any selfish type, whether agent 1 or agent 2 (at least virtually), reveals their private signal truthfully at the fourth sub-message announcement.

To be precise, for each $h \in \{1,...,H\}$, we define

$$T_i^h(\omega_i, s_i) \equiv \{t_i \in T_i \mid I_i(s_i^h(\omega_i, t_i)) = p_i\},$$

and

$$T_i^h(\omega_i, \omega_i', s_i) \equiv \{t_i \in T_i \mid I_i(s_i^h(\omega_i, t_i)) = \omega_i'\}.$$

Note that $T_i^h(\omega_i, s_i)$ is the set of the agent $i's$ types that reveal nothing at the $h-th$ sub-message when their private signal is ω_i, whereas $T_i^h(\omega_i, \omega_i', s_i)$ is the set of the agent $i's$ types that reveal ω_i' at the $h-th$ sub-message when their private signal is ω_i. According to the iterative eliminations of dominated strategies from the 1-st sub-message to the H-th sub-message, we prove that a unique BNE s exists and s satisfies

$$T_i^H(\omega_i, \omega_i, s_i) = T_i \quad \text{for all } i \in N \text{ and } \omega_i \in \Omega_i,$$

that is, the H-th sub-messages of all agents succeed in revealing their private signals truthfully.

Consider the 1-st sub-message. Due to the specification of x, any selfish agent i maximizes the sum of quadratic scoring rules $\sum_{j \neq i} \gamma_i(m_{i,j}^1, p_i)$. This maximization uniquely determines

$$m_{i,j}^1 = s_{i,j}^1(\omega_i, t_i) = p_i \quad \text{for all } j \neq i.$$

Hence, we have

$$I_i^1(s_i^1(\omega_i, t_i)) = p_i,$$

that is, any selfish agent reveals nothing at the 1-st sub-message. Clearly, any honest agent truthfully reveals their private signal, that is, $I_i^1(s_i^1(\omega_i, t_i)) = \omega_i$. Hence, no agent reveals incorrectly at the 1-st sub-message. Accordingly, we have

$$T_i^1(\omega_i, s_i) \subset E_i^*,$$

$$T_i^1(\omega_i, \omega_i, s_i) = T_i \setminus T_i^1(\omega_i, s_i),$$

and

$$T_i^1(\omega_i, \omega_i', s_i) = \phi \quad \text{for all} \quad \omega_i' \neq \omega_i.$$

As $T_i^1(\omega_i, s_i)$ is independent of ω_i, we can write

$$T_i^1(s_i) \equiv T_i^1(\omega_i, s_i).$$

Consider the 2-nd sub-message. Any selfish agent i maximizes the expected value of $\gamma_{i,j}(m_{i,j}^2, I_j^1(m_j^1))$. If the agent i is selfish and expects an agent $j \neq i$ to belong to $T_j^1(s_j)$ with certainty, they maximize the value of the quadratic scoring rule $\gamma_{i,j}(m_{i,j}^2, p_j)$. This maximization uniquely determines

$$m_{i,j}^2 = s_{i,j}^2(\omega_i, t_i) = p_i.$$

Hence, if the agent i is selfish and expects any other agent $j \neq i$ to belong to $T_j^1(s_j)$ with certainty, then we have

$$I_i^2(s_i^2(\omega_i, t_i)) = p_i.$$

By contrast, if the agent i is selfish and expects an agent $j \neq i$ to belong to $T_j \setminus T_j^1(s_j)$ with a positive probability, a $\lambda > 0$ exists such that the agent i maximizes the expected value of $\gamma_{i,j}(m_{i,j}^2, \lambda p_{i,j}(\cdot \mid \omega_i) + (1-\lambda) p_j)$. Given ID, this maximization uniquely determines

$$m_{i,j}^2 = s_i^2(\omega_i, t_i) = \lambda \omega_i + (1-\lambda) p_i.$$

Hence, this selfish agent reveals their private signal correctly; we have

$$I_i^2(s_i^2(\omega_i, t_i)) = \omega_i.$$

Any honest agent reveals their private signal correctly and no agent reveals it incorrectly at the 2-nd sub-message. Accordingly, we have

$$T_i^2(\omega_i, s_i) \subset V_i^1(E^*),$$

$$T_i^2(\omega_i, \omega_i, s_i) = T_i \setminus T_i^2(\omega_i, s_i),$$

and

$$T_i^2(\omega_i, \omega_i', s_i) = \phi \quad \text{for all} \quad \omega_i' \neq \omega_i .$$

As $T_i^2(\omega_i, s_i)$ is independent of ω_i, we can write $T_i^2(s_i) \equiv T_i^2(\omega_i, s_i)$.

Consider an arbitrary $h \in \{3, ..., H\}$ and the h-th sub-message. Suppose that for every $i \in N$, $\omega_i \in \Omega_i$, and $h' \in \{1, ..., h-1\}$, $T_i^{h'}(\omega_i, s_i)$ is independent of ω_i, that is, we can write $T_i^{h'}(s_i) \equiv T_i^{h'}(\omega_i, s_i)$. Further, suppose that for every $i \in N$, $\omega_i \in \Omega_i$, and $h' \in \{1, ..., h-1\}$,

$$T_i^{h'}(\omega_i, \omega_i, s_i) = T_i \setminus T_i^{h'}(s_i),$$

$$T_i^{h'}(\omega_i, \omega_i', s_i) = \phi \quad \text{for all} \quad \omega_i' \neq \omega_i ,$$

and

$$T_i^{h'}(s_i) \subset V_i^{h'-1}(E^*).$$

Similar to the argument for the 2-nd sub-message, if the agent i is selfish and expects all agents $j \neq i$ to belong to $T_j^{h-1}(s_j)$ with certainty, then we have $I_i^h(s_i^h(\omega_i, t_i)) = p_i$. If the agent i is selfish and expects some agent $j \neq i$ to belong to $T_j \setminus T_j^{h-1}(s_j)$ with a positive probability, then we have $I_i^h(s_i^h(\omega_i, t_i)) = \omega_i$. If the agent i is honest, then $I_i^h(s_i^h(\omega_i, t_i)) = \omega_i$. Accordingly, we have

$$T_i^h(\omega_i, s_i) \subset V_i^{h-1}(E^*),$$

$$T_i^h(\omega_i, \omega_i, s_i) = T_i \setminus T_i^h(\omega_i, s_i),$$

$$T_i^h(\omega_i, \omega_i', s_i) = \phi \quad \text{for all} \quad \omega_i' \neq \omega_i ,$$

and $T_i^h(\omega_i, s_i)$ is independent of ω_i, that is, we can write $T_i^h(s_i) \equiv T_i^h(\omega_i, s_i)$.

From the above observations, we have

$$T_i^H(s_i) \subset V_i^{H-1}(E^*),$$

$$T_i^H(\omega_i, \omega_i, s_i) = T_i \setminus T_i^H(s_i),$$

and

$$T_i^H(\omega_i, \omega_i', s_i) = \phi \quad \text{for all} \quad \omega_i' \neq \omega_i .$$

As the epistemological type space is finite, we can derive the common knowledge event through finite iteration steps; that is, for each event $E \subset T$, a positive integer K exists such that

$$V_i^{\infty}(E) = V_i^{k}(E) \quad \text{for all} \quad k \geq K \quad \text{and} \quad i \in N .$$

Hence, from ID and NCKS, a K exists such that $V_i^{k}(E^*) = \phi$ for all $k \geq K$. By selecting $H > K$, we obtain $T_i^{H}(s_i) \subset V_i^{H-1}(E^*) = \phi$, which implies

$$T_i^{H}(\omega_i, \omega_i, s_i) = T_i ,$$

that is,

$$T_i^{H}(\omega_i, \omega_i, s_i) = T_i \quad \text{for all} \quad i \in N , \quad \omega_i \in \Omega_i , \text{ and } \quad t_i \in T_i .$$

Hence, Proposition 1 is proven.

4. General Case

4.1. Mechanism Design

To prove Theorem 1 generally, we design the following mechanism $G = (M, g, x)$ as an extension of the mechanism designed in Subsection 3.1. We fix arbitrary positive integers H and K, which are set sufficiently large. Let

$$L = (n-1)H + K ,$$

and specify

$$M_i = \underset{h=1}{\overset{H+K}{\times}} M_i^h .$$

For each $h \in \{1, ..., H\}$, let

$$M_i^h = \underset{j \neq i}{\times} M_{i,j}^h .$$

We specify $M_{i,j}^h$ similarly as in Subsection 3.1:

$$M_{i,j}^h = \{\alpha_i \in \Delta(\Omega_i) \mid \exists (\omega_i, \lambda) \in \Omega_i \times [0,1] : \alpha_i = \lambda \omega_i + (1-\lambda) p_i\} .$$

Each agent i announces the h-th sub-message $m_{i,j}^h$ for each $h \in \{1,...,H\}$ similar to the mechanism designed in the information elicitation problem, which comprises $n-1$ sub-sub-messages, that is, $m_i^h = (m_{i,j}^h)_{j \neq i}$.

In addition to these H sub-messages, each agent announces K more sub-messages as follows. For each $k \in \{H+1,...,H+K\}$, each agent i announces an element of Ω_i as the k-th sub-message, where

$$M_i^k = \Omega_i .$$

Hence, for each of the first H sub-messages, the agent i makes $n-1$ sub-sub-message announcements (i.e., $m_{i,j}^h \in \Delta(\Omega_i)$ for $j \neq i$), whereas, for each of the last K sub-messages, they make a single announcement (i.e., $m_i^k \in \Omega_i$). The first H sub-messages play a central role in making the (H+1)-th sub-messages truthful, and therefore, suitable for the reference in judging whether the last K-1 sub-messages are correct. The last K-1 sub-messages play a central role in determining the allocation and side payment decisions. Note that in contrast to the mechanism designed for the information elicitation problem, the H-th sub-messages have no direct effect on the determination of allocation and side payment decisions.

We specify the allocation rule g as follows:

$$g(m) = \frac{\sum_{k=H+2}^{H+K} f(m^k)}{K-1} \quad \text{for all} \quad m \in M .$$

Note that $g(m)$ is independent of the first H+1 sub-messages. The central planner randomly selects $k \in \{H+2,...,H+K\}$ from the last K-1 sub-message profiles and then determines the allocation according to $f(m^k) \in \Delta(A)$.

We now specify the payment rule x. We define $\hat{w}_i : M \to [-2,0]$ as follows:

$$\hat{w}_i(m) = -1 \qquad \text{if } k \in \{H+2,...,H+K\} \text{ exists such that } m_i^k \neq m_i^{H+1}, \text{ and } m_j^{k'} = m_j^{H+1} \text{ for all } k' \in \{H+2,...,k-1\} \text{ and } j \in N,$$

and

$$\hat{w}_i(m) = 0 \qquad \text{if no such } k \in \{H+2,...,H+K\} \text{ exists.}$$

Note that $\hat{w}_i(m)$ indicates whether the agent i is the first deviant from their (H+1)-th sub-message (the reference).

Let $\hat{r}_i(m_i) \in \{0,...,K-1\}$ denote the number of integers, $k \in \{H+2,...,H+K\}$ such that $m_i^k \neq m_i^{H+1}$, that is, the number of the agent $i's$ misannouncements during the last K-1 sub-message announcements.

We specify the payment rule x_i for the agent i as a combination of the payment rule specified in the information elicitation problem, the above-specified functions $\hat{w}_i$ and $\hat{r}_i$, and a quadratic scoring rule given by $\gamma_{i,j}(m_{i,j}^{H+1}, I_j^H(m_j^H))$ as follows: for every $m \in M$ and $i \in N$,

$$x_i(m) = \frac{\varepsilon}{3+\xi}\left[\frac{1}{(n-1)H}\sum_{j\neq i}\left\{\gamma_i(m_{i,j}^1, p_i) + \sum_{h=2}^{H}\gamma_{i,j}(m_{i,j}^h, I_j^{h-1}(m_j^{h-1}))\right\}\right.$$

$$\left. +\xi\sum_{j\neq i}\gamma_{i,j}(m_{i,j}^{H+1}, I_j^H(m_j^H)) + \hat{w}_i(m) - \frac{\hat{r}_i(m_i)}{K-1}\right],$$

where $\xi > 0$ is an arbitrarily positive real number, which is sufficiently large. Note that the specified x satisfies the limited liability. We select a sufficiently large H and K to satisfy that for every $i \in N$ and $\omega \in \Omega$,

(1) $$K > \frac{3+\xi}{\varepsilon}\max_{(a,a')\in A^2}\{v_i(a,\omega) - v_i(a',\omega)\} + 1.$$

The aforementioned payment rule comprises three parts. The first part, which is given by

$$\frac{1}{(n-1)H}\sum_{j\neq i}\left\{\gamma_i(m_{i,j}^1, p_i) + \sum_{h=2}^{H}\gamma_{i,j}(m_{i,j}^h, I_j^{h-1}(m_j^{h-1}))\right\},$$

corresponds to the payment rule designed for the information elicitation problem. As in Subsection 3.2, we can demonstrate that all agents reveal their private signals truthfully at the H-th sub-message.

The second part is the variant of the quadratic scoring rule that is given by

$$\xi\sum_{j\neq i}\gamma_{i,j}(m_{i,j}^{H+1}, I_j^H(m_j^H)),$$

which, along with the truthful revelations at the H-th sub-message, succeeds in incentivizing selfish agents to make their H+1-th sub-message truthful and is regarded as the reference.

The third part, which is given by

$$\hat{w}_i(m) - \frac{\hat{r}_i(m_i)}{K-1},$$

corresponds to the Abreu–Matsushima mechanism design, which is the general standard method in unique implementation for eliminating unwanted equilibria that are inconsistent with the references (Abreu and Matsushima, 1992a; 1992b).

4.2. Proof of Theorem 1

We show that the mechanism G designed in Subsection 4.1 uniquely implements the SCF f. The proof of Theorem 1 is divided into two parts: "information elicitation" and "implementation with provability."

Part 1 (Information Elicitation): As H is sufficiently large, we can show in the same manner as in the information elicitation problem (in Subsection 3.2) that any BNE s satisfies

$$I_i^H(s_i^H(\omega_i, t_i)) = \omega_i \quad \text{for all} \quad i \in N, \quad \omega_i \in \Omega_i, \text{ and } t_i \in T_i.$$

Hence, any agent truthfully reveals their private signal at the H-th sub-message announcement. As in Subsection 3.2, we also have a uniqueness in BNE for the first H sub-message announcements.

Part 2 (Implementation with Provability): Consider a strategy profile s whose first H sub-message announcements satisfy the BNE property (i.e., Part 1). We define the sincere strategy for the agent i, denoted by $\hat{s}_i = (\hat{s}_i^k)_{k=1}^{H+K}$, as

$$\hat{s}^h = s^h \quad \text{for all} \quad h \in \{1, ..., H\},$$

and

$$\hat{s}_i^k(\omega_i, t_i) = \omega_i \quad \text{for all} \quad k \in \{H+1, ..., H+K\}.$$

Clearly, $\hat{s}$ induces the value of the SCF f. Part 2 shows that if s is a BNE, then $s = \hat{s}$ must hold.

Note that as

$$I_j^H(m_j^H) = \omega_j,$$

$$p_{i,j}(\cdot \mid \omega_i) \neq p_{i,j}(\cdot \mid \omega_i') \text{ for all } \omega_i' \neq \omega_i,$$

and ξ is sufficiently large, the nature of the variant of the quadratic scoring rule $\gamma_{i,j}(m_{i,j}^{H+1}, I_j^H(m_j^H))$ implies that each agent i is willing to announce $m_{i,j}^{H+1} = \omega_i$ uniquely, irrespective of the selection of $(m_i^k)_{k=H+2}^{H+K}$ (we must note that $m_{i,j}^{H+1}$ is only relevant to $\gamma_{i,j}(m_{i,j}^{H+1}, I_j^H(m_j^H))$ in the mechanism). Hence, we have

$$s^{H+1} = \hat{s}^{H+1}.$$

If an agent i announces a sub-message that differs from their (H+1)-th sub-message as the first deviation among all agents starting from the (H+2)-th sub-messages, this agent is fined the monetary amount $\frac{\varepsilon}{3+\xi}$. As K is sufficiently large, that is, we have the inequality of (1), the impact of the selection of each sub-message on the determination of the allocation is small compared with the monetary amount $\frac{\varepsilon}{3+\xi}$. Following Abreu and Matsushima (1992a, 1992b), this drives agents into tail-chasing competition through which each agent avoids becoming the first deviant. Given that all agents reveal truthfully at their (H+1)-th sub-message, this competition drives them to announce the state truthfully from the (H+2)-th sub-message to the (H+K)-th sub-message.

To be precise, consider an arbitrary $k \in \{H+2, ..., H+K\}$ and suppose that

$$s^{k'} = \hat{s}^{k'} \text{ for all } k' < k.$$

If $m_j^k \neq \omega_j$ for some $j \neq i$, the agent i strictly prefers announcing truthfully at the k-th sub-message because the agent i can avoid being the first deviant (here, inequality (1) induces an incentive for this avoidance). Even if $m_j^k = \omega_j$ for all $j \neq i$, the agent i still strictly prefers announcing truthfully at the k-th sub-message because

they do not want to increase $r_i(m_i)$ and the SCF is incentive-compatible. Accordingly, through the iterative elimination of dominated strategies, we can inductively prove that

$$s^k = \hat{s}^k \quad \text{for all} \quad k \in \{H+2, ..., H+K\}.$$

Hence, no BNE other than $\hat{s}$ exists. As $\hat{s}$ is a BNE and achieves the value of f, the proof of Theorem 1 is completed.

5. Weak Honesty

Throughout this study, we assume strict honesty, in the sense that honest agents never consider their material payoffs. A weaker version of honesty permits an agent to decide how to make announcements by weighing their preferences for honesty and material interest. Fortunately, we can illustrate the same positive result as in Theorem 1 even if agents are either selfish or weakly honest (but not strictly honest). In our design of quadratic scoring rules and its variant, a weakly honest agent is willing to announce more honestly than a selfish agent. Hence, given the continuum of message spaces, even weakly honest agents will virtually reveal their private signals at the first sub-message announcement.

The previous work by Matsushima (2022a) demonstrates a definition of weak honesty, where only a tiny cost of adopting dishonest attitudes is considered. Matsushima (2022a) then assumes that agents are either selfish or weakly honest and proves that under complete information environments with three or more agents, any SCF is uniquely implementable under NCKS. Although not specifically proven in this study, the generalizability of the study of weak honesty by Matsushima (2022a) for complete information to asymmetric information is almost self-evident. That is, even if we replace strict honesty with such weak honesty, we can prove the positive result that any SCF, whether ethical or nonethical, is uniquely implementable if NCKS and ID hold.

6. Conclusion

This study investigates the unique implementation problem of SCFs in asymmetric information environments, in which we assume that agents are either selfish or honest. Following Matsushima (2022), we introduce the epistemological framework and assume NCKS, that is, an honest agent exists in agents' higher-order beliefs. We show a positive result that any incentive-compatible SCF, whether ethical or nonethical, is uniquely implementable in BNE under the minor restriction on private signal correlation, termed ID. The generalization from complete information to asymmetric information is essential when considering the implementability of equitable social choice. This is because important information is dissipatively distributed and only privately known by people other than the person concerned who are not tied to their self-interests. This study presents a unique method to successfully extract information from such people who are privately informed but selfish people.

We assume that an agent who is not selfish is honest. However, if we permit the possibility that an agent is neither selfish nor honest but adversarial (anti-social or spiteful), the situation may change. For example, the SCF can fail to be implementable even if an honest participant exists, and all participants are either selfish or honest. If agents expect adversarial agents to exist in their higher-order beliefs, any selfish participant may be willing to lie.

Abeler, Nosenzo, and Raymond (2019) show that subjects who trade-off between material interest and honesty forego a large fraction of potential benefits from lying. Surely, they support the validity of this study; although, the support is not sufficient. The expected presence of even a few adversarial people in agents' epistemology may cause all the differences, depending on the shape of the epistemological network structure. Under these circumstances, future research on implementation theory must further develop the epistemological framework by considering various non-selfish motives besides honesty.[10]

[10] See Matsushima (2022b), which considers adversarial types, as well as selfish and honest types.

Acknowledgements

This study was supported by a grant-in-aid for scientific research (KAKENHI 20H00070) from the Japan Society for the Promotion of Science (JSPS) and the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT). There are no conflict of interest to declare.

References

Abeler, J., D. Nosenzo, and C. Raymond (2019): Preference for Truth-Telling, *Econometrica*, 87 (4), 1115–1153.

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

Abreu, D., and H. Matsushima (1992b): Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information, mimeo. https://www.econexp.org/hitoshi/AMincomplete.pdf

Arrow, K. (1951): *Social Choice and Individual Values*, New Haven: Yale University Press.

Brier, G. (1950): Verification of Forecasts Expressed in Terms of Probability, *Monthly Weather Review*, 78, 1–3.

Dogan, B. (2017): Eliciting the Socially Optimal Allocation from Responsible Agents, *Journal of Mathematical Economics*, 73, 103–110.

Dutta, B., and A. Sen (2012): Nash Implementation with Partially Honest Individuals, *Games and Economic Behavior*, 74 (1), 154–169.

Gibbard, A. (1973): Manipulation of Voting Schemes: A General Result, *Econometrica*, 41 (4), 587–601.

Hurwicz, L. (1972): On Informationally Decentralized Systems, in Decision and Organization, ed. by C.B. McGuire and R. Radner. Amsterdam: North-Holland.

Jackson, M. (1991): Bayesian Implementation, *Econometrica*, 59 (2), 461–477.

Kartik, N., O. Tercieux, and R. Holden (2014): Simple Mechanisms and Preferences for Honesty, *Games and Economic Behavior*, 83, 284–290.

Lombardi, M., and N. Yoshihara (2018): Treading a Fine Line: (Im)Possibilities for Nash Implementation with Partially-Honest Individuals, *Games and Economic Behavior*, 111, 203–216.

Maskin, E. (1977/1999): Nash Equilibrium and Welfare Optimality, *Review of Economic Studies*, 66, 23–38.

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

Matsushima, H. (2008a): Behavioral Aspects of Implementation Theory, *Economics*

Letters, 100 (1), 161–164.

Matsushima, H. (2008b): Role of Honesty in Full Implementation, *Journal of Economic Theory*, 139, 353–359.

Matsushima, H. (2013): Process Manipulation in Unique Implementation, *Social Choice and Welfare*, 41 (4), 883–893.

Matsushima, H. (2021): Partial Ex-Post Verifiability and Unique Implementation of Social Choice Functions, *Social Choice and Welfare*, 56, 549–567.

Matsushima, H. (2022a): Epistemological Implementation of Social Choice Functions, *Games and Economic Behavior*, 136, 389–402.

Matsushima, H. (2022b): Conformity and Mechanism Design, in preparation.

Mukherjee, S., N. Muto, and E. Ramaekers (2017): Implementation in Undominated Strategies with Partially Honest Agents, *Games and Economic Behavior*, 104, 613–631.

Ortner, J. (2015): Direct Implementation with Minimally Honest Individuals, *Games and Economic Behavior*, 90, 1–16.

Saporiti, A. (2014): Securely Implementable Social Choice Rules with Partially Honest Agents, *Journal of Economic Theory*, 154, 216–228.

Satterthwaite, M. (1975): Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions, *Journal of Economic Theory,* 10 (2), 187–217.

Savva, F. (2018): Strong Implementation with Partially Honest Individuals, *Journal of Mathematical Economics*, 78, 27–34.

Yadav, S. (2016): Selecting Winners with Partially Honest Jurors, *Mathematical Social Sciences*, 83, 35–43.