Double implementation with partially honest agents*

Makoto Hagiwara†

August 9, 2017

Abstract

Theoretical studies usually assume that all agents only care about the outcome obtained in the mechanism. In the standard setting, Maskin monotonicity is necessary, and along with no veto power is sufficient for double implementation in Nash equilibria and undominated Nash equilibria with at least three agents. However, there are unanimous social choice correspondences failing the two conditions so that the SCCs cannot be doubly implemented. In this paper, we assume that there are some partially honest agents in the sense of Dutta and Sen[5]. As our main result, we show that if at least two agents are partially honest, unanimity is sufficient for double implementation with at least three agents. From this result, we derive a number of positive corollaries in some problems.

JEL Classification: C72, D71, D78
Key words: Partial honesty, Double implementation, Unanimity, Social choice correspondence, Social responsibility

1 Introduction

The theory of mechanism design aims to identify a mechanism achieving a social goal across a domain of agents' preferences. Theoretical studies usually assume that all agents only care about the outcome obtained in the mechanism. On the other hand, experimental studies observed that some agents have intrinsic preferences for honesty. For example, Gneezy [7] and Hurkens and Kartik [10] reported that agents are one of two kinds: either an agent will never lie, or an agent will lie whenever he prefers the outcome obtained by lying over the outcome obtained by telling the truth. Following such experimental observations, a bunch of studies discuss the issue of implementation when agents have intrinsic preferences for honesty.

Dutta and Sen [5] construct a mechanism in which each agent reports a preference profile and an outcome. Under their mechanism, Dutta and Sen[5] assume that some

*I thank Bhaskar Dutta, Hirofumi Yamamura, and Takehiko Yamato for their invaluable advice and suggestions. I am grateful to two anonimous reviewers of an earlier version of the paper and Keisuke Bando as well as to participants at the 2016 Fall National Conference of The Operations Research Society of Japan (Yamagata University, 2016). This work was partially supported by JSPS KAKENHI Grant Number 17J01520.

†Department of Industrial Engineering and Economics, School of Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552, Japan; E-mail: *hagiwara.m.af@m.titech.ac.jp*

agent has a small intrinsic preference for telling the true preference profile whom they call a *partially honest* agent. An agent is partially honest if he prefers reporting the true preference profile whenever a lie does not allow him to obtain a more preferred outcome; otherwise, he prefers to announce a message inducing a more preferred outcome. They prove that if at least one agent is partially honest, then every social choice correspondence (SCC) satisfying no veto power can be implemented in Nash equilibria with at least three agents by their mechanism.[1] Also, Kimya [17] establishes that if there are at least three agents and all agents are partially honest, then every SCC satisfying unanimity can be implemented in Nash equilibria with at least three agents by Dutta and Sen's mechanism. He mentions that his result is still valid if at least two agents are partially honest.

For Dutta and Sen's mechanism, however, the set of undominated Nash equilibrium outcomes may be strictly smaller than the set of Nash equilibrium outcomes.[2] Thus, Dutta and Sen's mechanism may not implement an SCC with partially honest agents if agents use undominated strategies.

Is it sufficient to design a mechanism that implements an SCC in just undominated Nash equilibria with partially honest agents? Our answer is negative because laboratory evidence casts doubt on the assumption that agents adopt undominanted strategies. In pivotal mechanism experiments in which for each agent, telling her true value is a dominant strategy, Attiyeh et al. [1] and Kawagoe and Mori [16] observed that more than half of subjects adopted weakly dominated strategies. Moreover, in second price auction experiments in which for each agent, bidding her true value is a dominant strategy, Kagel et al. [14], Kagel and Levin [13], and Harstad [9] observed that most bids did not reveal true values. It was not obvious whether or not each agent adopted undominated strategies. Thus, it is desirable to construct mechanisms that are applicable not only when agents use undominanted strategies but also when they do not. On the other hand, as Cason et al. [2] point out, the high rate of the observed undominanted strategy outcomes were Nash equilibria in their experiments. Therefore, although subjects frequently played Nash equilibria, there was no guarantee that they did not use weakly dominated strategies. Then, we are concerned with the design of a mechanism that doubly implements an SCC in Nash equilibria and undominated Nash equilibria with partially honest agents.

Previous studies show that if all agents only care about the outcome obtained in the mechanism, Maskin monotonicity is necessary, and along with no veto power is sufficient for double implementation in Nash equilibria and undominated Nash equilibria with at least three agents (Jackson et al.[12], Tatamitani[26], and Yamato [27][28]). However, there are unanimous social choice correspondences failing the two conditions so that the SCCs cannot be doubly implemented. In this paper, we assume that there are some partially honest agents in the sense of Dutta and Sen[5].[3] We show that if at least one agent is partially honest, no veto power is sufficient for double implementation with at

[1] Lombardi and Yoshihara [19] provide a characterization of implementation in Nash equilibria with at least three agents if at least one agent is partially honest.

[2] Yamato[28] provides an example that in a mechanism used by Maskin[22], the set of undominated Nash equilibrium outcomes may be strictly smaller than the set of Nash equilibrium outcomes. Also, we can easily show that, in Dutta and Sen's mechanism, the set of undominated Nash equilibrium outcomes may be strictly smaller than the set of Nash equilibrium outcomes even if two agents are partially honest. See Yamato[28] and Example 4 of this paper.

[3] There are other definitions of preferences for honesty. For instance, see Corchón and Herrero [3], Lombardi and Yoshihara [20], Matsushima [21], and Mukherjee et al. [23].

least three agents. Therefore, we no longer need Maskin monotonicity as a necessary condition of double implementability. Moreover, we show that if at least two agents are partially honest, unanimity is sufficient for double implementation with at least three agents. Hence, more social choice correspondences can be doubly implemented if at least two agents are partially honest since unanimity is weaker than no veto power.

To provide the practical value of our main results, we examine the double implementability in problems of allocating an infinitely divisible resource, coalitional games, general problems of one-to-one matching, and voting games. Since all SCCs considered in Section 4 violate Maskin monotonicity, the SCCs cannot be doubly implemented in the standard setting. On the other hand, the SCCs can be doubly implemented with partially honest agents.

We have considered a truth-telling messages with regard to a preference profile following Dutta and Sen [5] and used a complicated mechanism. On the other hand, motivated by some studies (Gneezy[7], Hurkens and Kartik[10], Doğan[4], and Matsushima[21]), Hagiwara et al.[8] consider a truth-telling messages with regard to an outcome. Specifically, they assume that some agent has a small intrinsic preference for reporting a socially desirable outcome whom they call a *socially responsible* agent. They design a simple and natural mechanism for implementation in Nash equilibria with socially responsible agents which they call the *outcome mechanism*. We show that if all agents are socially responsible, then the outcome mechanism can doubly implement any single-valued SCC satisfying unanimity with at least three agents. Therefore, in problems of Section 4, if an SCC is single-valued, then the simple and natural mechanism can doubly implement the SCC with socially responsible agents under Assumption n.

This paper is organized as follows. Section 2 presents notation including assumptions on partially honest agents. Section 3 reports our main results about double implementation with partially honest agents and related literature. Section 4 discusses implications for four problems. Section 5 reports a result about double implementation with socially responsible agents and related literature. Section 6 provides concluding remarks. Appendix A includes the proof of Theorem 1 and Appendix B proposes the proof of Theorem 2.

2 Notation

Let A be the arbitrary set of outcomes and $N = \{1, ..., n\}$ be the set of agents. Let R_i be a preference ordering for agent $i \in N$ over A, whose asymmetric and symmetric components are P_i and I_i, respectively. Let $\mathcal{R}_i$ be the set of preference orderings admissible for agent $i \in N$. Let $R = (R_1, ..., R_n)$ be a preference profile and $\mathcal{R} = \times_{i \in N} \mathcal{R}_i$. Let $\mathcal{D} = \times_{i \in N} \mathcal{D}_i \subseteq \mathcal{R}$ where $\mathcal{D}_i \subseteq \mathcal{R}_i$ for each $i \in N$ be a *domain*.

A *social choice correspondence (SCC)* is a mapping $F : \mathcal{D} \twoheadrightarrow A$ that specifies a nonempty subset $F(R) \subseteq A$ for each $R \in \mathcal{D}$. Given an SCC F, an outcome $a \in A$ is *F-optimal at $R \in \mathcal{D}$* if $a \in F(R)$. An SCC F is *single-valued* if $|F(R)| = 1$.

A *mechanism* Γ consists of a pair (M, g) where $M = \times_{i \in N} M_i$, M_i is the *message* (or *strategy*) *space of agent* $i \in N$, and $g : M \longrightarrow A$ is the *outcome function* mapping each message profile $m \in M$ into an outcome $g(m) \in A$.

2.1 Assumptions on partially honest agents

The literature on mechanism design usually assumes that each agent only cares about the outcome obtained in the mechanism. However, some recent studies assume that some agents may have *intrinsic preferences for honesty.*

Dutta and Sen [5] construct a mechanism in which each agent reports a preference profile and an outcome. Under their mechanism, Dutta and Sen[5] assume that some agent has a small intrinsic preference for telling the true preference profile whom they call a *partially honest* agent. An agent is partially honest if he prefers reporting the true preference profile whenever a lie does not allow him to obtain a more preferred outcome; otherwise, he prefers to announce a message inducing a more preferred outcome. Lombardi and Yoshihara [19] extend Dutta and Sen's notion of partially honesty by introducing a truth-telling correspondence for any mechanism. We follow Lombardi and Yoshihara [19].

Let an SCC F be given. For each $i \in N$ and each mechanism Γ, a *truth-telling correspondence* T_i^Γ is a mapping $T_i^\Gamma : \mathcal{D} \twoheadrightarrow M_i$ that specifies a non-empty set of truth-telling messages $T_i^\Gamma(R) \subseteq M_i$ for each $R \in \mathcal{D}$. Given a mechanism Γ, a truth-telling correspondence T_i^Γ, and $R \in \mathcal{D}$, we say that agent $i \in N$ *behaves truthfully* at $m \in M$ if and only if $m_i \in T_i^\Gamma(R)$.

If there are partially honest agents, we focus on mechanisms in which each agent reports a preference profile and a supplemental message. For each agent $i \in N$, the message space of agent $i \in N$ consists of $M_i = \mathcal{D} \times S_i$, where S_i denotes the set of supplemental messages. For each $i \in N$, $m_i = (R^i, s^i)$ is a truth-telling message if and only if $R^i = R$. Then, a truth-telling correspondence is defined by $T_i^\Gamma(R) = \{R\} \times S_i$ for each $i \in N$ and each $R \in \mathcal{D}$.

For each $i \in N$, each $R \in \mathcal{D}$, each mechanism Γ, and each truth-telling correspondence T_i^Γ, agent i's preference ordering $\succsim_i^R$ over M at $R \in \mathcal{D}$, whose asymmetric and symmetric components are $\succ_i^R$ and $\sim_i^R$ respectively, is defined below.

Definition 1. An agent $i \in N$ is *partially honest* if for each $R \in \mathcal{D}$ and each $(m_i, m_{-i}), (m_i', m_{-i}) \in M$, the following properties hold:
(1) If $m_i \in T_i^\Gamma(R)$, $m_i' \notin T_i^\Gamma(R)$ and $g(m_i, m_{-i}) R_i g(m_i', m_{-i})$, then $(m_i, m_{-i}) \succ_i^R (m_i', m_{-i})$.
(2) In all other cases, $g(m_i, m_{-i}) R_i g(m_i', m_{-i})$ if and only if $(m_i, m_{-i}) \succsim_i^R (m_i', m_{-i})$.

Since an agent who is not partially honest only cares about the outcomes obtained in the mechanism, his preference ordering over M is straightforward to define as follows:

Definition 2. An agent $i \in N$ is *not partially honest* if for each $R \in \mathcal{D}$ and each $(m_i, m_{-i}), (m_i', m_{-i}) \in M$, $g(m_i, m_{-i}) R_i g(m_i', m_{-i})$ if and only if $(m_i, m_{-i}) \succsim_i^R (m_i', m_{-i})$.

We consider the following assumptions:
Assumption 0. There is no partially honest agent in N.
Assumption 1. There exists at least *one* partially honest agent in N.
Assumption 2. There are at least *two* partially honest agents in N.[4]

[4] The traditional literature on mechanism design such as Yamato[28] usually studies Assumption 0. In contrast to the traditional literature, Dutta and Sen [5] and Lombardi and Yoshihara [18] [19] investigate Assumption 1. Moreover, Hagiwara et al.[8] consider Assumption 2.

We introduce our formal definitions of double implementation with partially honest agents under Assumption $k \in \{1,2\}$. For each $k \in \{1,2\}$, let $\mathcal{H}^k = \{S \subseteq N : |S| \geq k\}$. For each $R \in \mathcal{D}$ and each $H \in \mathcal{H}^k$, let $\succsim^{R,H} = (\succsim_1^{R,H}, ..., \succsim_n^{R,H})$ be the preference profile over M such that for each $i \in H$, $\succsim_i^{R,H}$ is defined by Definition 1 and for each $i \in N \backslash H$, $\succsim_i^{R,H}$ is defined by Definition 2.

Let $(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$ be a *game with partially honest agents* induced by a mechanism Γ, a truth-telling correspondence T_i^Γ for each $i \in N$, and a preference profile $\succsim^{R,H}$. A message profile $m \in M$ is a *Nash equilibrium in* $(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$ if for each $i \in N$ and each $m_i' \in M_i$, $(m_i, m_{-i}) \succsim_i^{R,H} (m_i', m_{-i})$. The set of Nash equilibria in $(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$ is denoted by $NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$. Also, the set of Nash equilibrium outcomes in $(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$ is denoted by $NE_A(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H}) = \{a \in A | \exists m \in NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$ with $g(m) = a\}$.

A message $m_i \in M_i$ is *weakly dominated by* $\tilde{m}_i \in M_i$ *at* $\succsim_i^{R,H}$ if $(\tilde{m}_i, m_{-i}) \succsim_i^{R,H} (m_i, m_{-i})$ for each $m_{-i} \in M_{-i}$ and $(\tilde{m}_i, m_{-i}) \succ_i^{R,H} (m_i, m_{-i})$ for some $m_{-i} \in M_{-i}$. A message $m_i \in M_i$ is *undominated at* $\succsim_i^{R,H}$if it is not weakly dominated by any message in M_i *at* $\succsim_i^{R,H}$. A message profile $m \in M$ is an *undominated Nash equilibrium in* $(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$ if for each $i \in N$, $m_i \in M_i$ is undominated at $\succsim_i^{R,H}$ and $m \in M$ is a Nash equilibrium with partially honest agents in $(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$. The set of undominated Nash equilibria in $(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$ is denoted by $UNE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$. Note that $UNE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H}) \subseteq NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$. Also, the set of undominated Nash equilibrium outcomes in $(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$ is denoted by $UNE_A(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H}) = \{a \in A | \exists m \in UNE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$ with $g(m) = a\}$.

Under Assumption 1 or Assumption 2, the mechanism designer knows that there are partially honest agents in N but does not know who these agents are. Hence, the mechanism designer needs to cover all feasible cases of partially honest agents to her knowledge. We amend the standard definition of implementation as follows:

Definition 3. Under Assumption $k \in \{1,2\}$, a mechanism Γ *doubly implements an SCC* F *in Nash equilibria and undominated Nash equilibria with partially honest agents* if for each $R \in \mathcal{D}$ and each $H \in \mathcal{H}^k$, $F(R) = NE_A(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H}) = UNE_A(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$.

3 Main results

We consider sufficient conditions for double implementation with partially honest agents.

For each $i \in N$, each $R_i \in \mathcal{D}_i$, and each $a \in A$, let $L(R_i, a) = \{b \in A \mid aR_ib\}$ be the *lower contour set of* $a \in A$ *for* $i \in N$ *at* $R_i \in \mathcal{D}_i$.

An SCC F satisfies *Maskin monotonicity* if for each R, $R' \in \mathcal{D}$ and each $a \in F(R)$, if for each $i \in N$, $L(R_i, a) \subseteq L(R_i', a)$, then $a \in F(R')$. Maskin monotonicity requires that if an outcome $a \in A$ is F-optimal at some preference profile and the profile is then altered so that in each agent's ordering, the outcome a does not fall below any outcome that was not below before, then the outcome a remains F-optimal at the new profile.

Definition 4. An SCC F satisfies *no veto power* if for each $i \in N$, each $R \in \mathcal{D}$, and each $a \in A$ if for each $j \neq i$, $L(R_j, a) = A$, then $a \in F(R)$.

No veto power says that if an outcome $a \in A$ is at the top of $(n-1)$ agents' preference orderings, then the last agent cannot prevent the outcome a from being F-optimal at the preference profile.

Definition 5. An SCC F satisfies *unanimity* if for each $R \in \mathcal{D}$ and each $a \in A$, if for each $i \in N$, $L(R_i, a) = A$, then $a \in F(R)$.

Unanimity says that if an outcome $a \in A$ is at the top of all agents' preference orderings, then the outcome a is F-optimal at the preference profile.

Our main results are as follows:

Theorem 1. *Let* $n \geq 3$.
(1) *Under Assumption 1, every SCC F satisfying no veto power can be doubly implemented with partially honest agents.*
(2) *Under Assumption 2, every SCC F satisfying unanimity can be doubly implemented with partially honest agents.*

The proof of Theorem 1 is given in Appendix A.

3.1 Related literature

Previous studies provide a necessary condition and a sufficient condition for double implementation under Assumption 0, respectively.

Proposition 1. *(Maskin [22], Yamato [28]) Let $n \geq 3$ and suppose Assumption 0 holds. If an SCC F does not satisfy Maskin monotonicity, it cannot be doubly implemented.*

Proposition 2. *(Jackson et al.[12], Tatamitani[26], and Yamato [27][28]) Let $n \geq 3$ and suppose Assumption 0 holds. Then, every SCC F satisfying Maskin monotonicity and no veto power can be doubly implemented.*

For our results and previous results, we summarize sufficient conditions for double implementation with partially honest agents under Assumption $k \in \{0, 1, 2\}$ in Figure 1 below.

	Assumption 0	$\rightarrow$	Assumption 1	$\rightarrow$	Assumption 2
Nash Implementation	Maskin [22]		Dutta and Sen [5]		Kimya [17]
	Maskin monotonicity *no veto power*		*no veto power*		*unanimity*
$\downarrow$					
Double Implementation	Jackson et al.[12] Tatamitani[26] Yamato [27][28]		This paper (Theorem 1 (1))		This paper (Theorem 1 (2))
	Maskin monotonicity *no veto power*		*no veto power*		*unanimity*

Figure 1

The following remark provides a difficulty of double implementability under Assumption 0.

Remark. The strong Pareto correspondence **SP** satisfies unanimity but violates Maskin monotonicity and no veto power for some domain:

Strong Pareto correspondence (SP): $\mathbf{SP}(R) = \{a \in A : \nexists b \in A$ such that for each $i \in N$, bR_ia, and for some $i \in N, bP_ia\}$

The following example represents that the strong Pareto correspondence violates Maskin monotonicity and no veto power.

Example 1. Consider the following example. There are three agents, $N = \{1, 2, 3\}$, two outcomes, $A = \{a, b\}$, and two possible preference profiles, $\mathcal{D} = \{R, R'\}$. Preferences are given by:

R_1	R_2	R_3		R_1	R_2	R_3
a, b	a	b		a, b	a	a, b
	b	a			b	

The strong Pareto correspondence evaluated at each preference profile is as follows: $\mathbf{SP}(R) = \{a, b\}$ and $\mathbf{SP}(R') = \{a\}$. Since Maskin monotonicity and no veto power imply that we must have $b \in \mathbf{SP}(R')$, the strong Pareto correspondence fails the conditions.

By Proposition 1, the strong Pareto correspondence cannot be doubly implemented under Assumption 0. On the other hand, since the strong Pareto correspondence satisfies unanimity, the strong Pareto correspondence can be doubly implemented with partially honest agents under Assumption 2.■

As one of studies in behavioral mechanism design, Kartik et al.[15] investigate a implementation problem in two rounds of strictly dominated strategies. This is also a study of double implementation.[5] However, there are three differences and therefore their result does not imply ours: First, while they consider a single-valued SCC F, we consider a set-valued SCC. Second, they assume that there is *separable punishment*: there are a function $x : \Theta \to A$, agent i, and agent $j \neq i$ such that for each $R, R' \in \mathcal{D}$, $F(R')I_jx(R')$ and $F(R')P_ix(R')$. Third, if there are at least three agents, the agents except for agent $i \in N$ have preferences for honesty on M and the mechanism designer knows that the other agent has a preference for honesty on M. This is stronger than our assumptions: Assumption 1 or Assumption 2. Although their mechanism is simpler than ours, they use a mechanism in which agent i who does not have a preference for honesty on M is a virtual dictator. Therefore, they consider a problem in which the dictator want to induce the socially optimal outcome and all the other agents have preferences for honesty so that the mechanism is applicable for less problems than ours.

4 Implications

In this section, we derive a number of corollaries in problems of allocating an infinitely divisible resource, coalitional games, general problems of one-to-one matching, and voting

[5] I am grateful to Bhaskar Dutta for pointing out this fuct.

games. Since all SCCs considered here violate Maskin monotonicity, the SCCs cannot be doubly implemented under Assumption 0 by Proposition 1. On the other hand, by our results, the SCCs can be doubly implemented with partially honest agents.

4.1 Problems of allocating an infinitely divisible resource

We consider a problem of allocating an infinitely divisible resource among a group of agents. A problem of allocating an infinitely divisible resource is a triple $(N, A(M), R)$. The first component $N = \{1, ..., n\}$ with $n \geq 3$ is a set of agents among whom an amount $M \in \mathbb{R}_{++}$ of an infinitely divisible resource has to be allocated. Note that we do not assume that the resource can be disposed of. Given $M \in \mathbb{R}_{++}$, an *allocation for* M is a list $a \in \mathbb{R}^N_+$ such that $\Sigma_{i \in N} a_i = M$. The second component $A(M) = \{a \in \mathbb{R}^N_+ | \Sigma_{i \in N} a_i = M\}$ is the set of *allocations.* The third component $R = (R_1, ..., R_n)$ where R_i is a preference ordering for agent $i \in N$ over A is a preference profile. Let P_i and I_i be the asymmetric and symmetric components of R_i, respectively.

We consider a situation in which the mechanism designer does not know agents' preferences. This situation is modeled by the triple $(N, A, \mathcal{D})$, which we refer to as a *division problem environment of an infinitely divisible resource.*

In addition to a domain $\mathcal{D} = \mathcal{R}$, we focus on domains satisfying the following restrictions:

Single-plateaued preferences. Given $(R_i, M) \in \mathcal{R}_i \times \mathbb{R}_{++}$, let $T(R_i, M) = \{a \in [0, M] | aR_i b \text{ for each } b \in [0, M]\}$ be the *top set for* (R_i, M). Let $T(R_i, M) = [\underline{\mathrm{T}}(R_i, M), \bar{T}(R_i, M)]$ be such that $\bar{\mathrm{T}}(R_i, M) = max\ T(R_i, M)$ and $\underline{\mathrm{T}}(R_i, M) = min\ T(R_i, M)$. A preference ordering $R_i \in \mathcal{R}_i$ is *single-plateaued on* $[0, M]$ if there is an interval $[\underline{\mathrm{T}}(R_i, M), \bar{T}(R_i, M)] \subseteq [0, M]$ such that for each $a, b \in [0, M]$, if $b < a \leq \underline{\mathrm{T}}(R_i, M)$ or $\bar{\mathrm{T}}(R_i, M) \leq a < b$, then $aP_i b$; if $\underline{\mathrm{T}}(R_i, M) \leq a \leq b \leq \bar{T}(R_i, M)$, then $aI_i b$. Let $\mathcal{D}^{SPL}_i$ be the set of single-plateaued preference orderings on $[0, M]$ for agent $i \in N$ and $\mathcal{D}^{SPL} = \times_{i \in N} \mathcal{D}^{SPL}_i$ be the single-plateaued domain on $[0, M]$.

Single-dipped preferences. A preference ordering $R_i \in \mathcal{R}_i$ is *single-dipped on* $[0, M]$ if there is a point $d(R_i) \in [0, M]$ such that for each $a, b \in [0, M]$, if $a < b \leq d(R_i)$ or $d(R_i) \leq b < a$, then $aP_i b$. Let $\mathcal{D}^{SD}_i$ be the set of single-dipped preference orderings on $[0, M]$ for agent $i \in N$ and $\mathcal{D}^{SD} = \times_{i \in N} \mathcal{D}^{SD}_i$ be the single-dipped domain on $[0, M]$.

Let us give an example of an SCC in a problem of allocating an infinitely divisible resource.

Strong Pareto correspondence (SP): $\mathbf{SP}(R) = \{a \in A(M) | \nexists b \in A(M)$ such that for each $i \in N$, $b_i R_i a_i$, and for some $i \in N, b_i P_i a_i\}$

If $\mathcal{D} = \mathcal{R}$ or $\mathcal{D}^{SPL}$, it is well-known that the strong Pareto correspondence violates Maskin monotonicity. If $\mathcal{D} = \mathcal{D}^{SD}$, Inoue and Yamamura[11] (Remark 1) show that any selection from the strong Pareto correspondence does not satisfy Maskin monotonicity. By Proposition 1, if $\mathcal{D} = \mathcal{R}$, $\mathcal{D}^{SPL}$, or $\mathcal{D}^{SD}$, the strong Pareto correspondence cannot be doubly implemented under Assumption 0.

On the other hand, it is well-known that the strong Pareto correspondence satisfies unanimity but violates no veto power. We conclude that under Assumption 2, the strong

Pareto correspondence can be doubly implemented with partially honest agents.

Corollary 1. *Let $(N, A(M), \mathcal{D})$ be a division problem environment of an infinitely divisible resource where $\mathcal{D} = \mathcal{R}$, $\mathcal{D}^{SPL}$, or $\mathcal{D}^{SD}$. Under Assumption 2, the strong Pareto correspondence* **SP** *can be doubly implemented with partially honest agents.*

4.2 Coalitional games

We consider a coalitional game. A coalitional game (N, A, R, v) contains a finite set of agents N with $n \geq 3$, a non-empty set of outcomes A, a preference profile $R \in \mathcal{D}$, and a characteristic function $v : 2^N \backslash \{\phi\} \to 2^A$, which assigns for each coalition $S \in 2^N \backslash \{\phi\}$ a subset of outcomes.

We consider a situation in which the mechanism designer knows the characteristic function v, but she does not know agents' preferences. This situation is modeled by the four-tuple $(N, A, \mathcal{D}, v)$, which we refer to as a *coalitional game environment.*

Let us give an example of an SCC in a coalitional game.

Given a coalitional game (N, A, R, v), an outcome $a \in A$ is *weakly blocked by S* if there is $b \in v(S)$ such that $bR_i a$ for each $i \in S$, and $bP_i a$ for some $i \in S$.

Strong core correspondence(SC): $\mathbf{SC}(R) = \{a \in v(N) : a$ is not weakly blocked by any coalition $S\}$

We say that $(N, A, \mathcal{D}, v)$ is *a coalitional game environment with non-empty strong core* if $\mathbf{SC}(R) \neq \phi$ for each $R \in \mathcal{D}$.

Lombardi and Yoshihara[19] show that the strong core correspondence does not satisfy Maskin monotonicity.[6] By Proposition 1, the strong core correspondence cannot be doubly implemented under Assumption 0.

On the other hand, it is well-known that the strong core correspondence satisfies unanimity but violates no veto power. We conclude that under Assumption 2, the strong core correspondence can be doubly implemented with partially honest agents.

Corollary 2. *Let $(N, A, \mathcal{D}, v)$ be a coalitional game environment with non-empty strong core. Under Assumption 2, the strong core correspondence* ***SC*** *can be doubly implemented with partially honest agents.*

4.3 General problems of one-to-one matching

We consider a general problem of one-to-one matching (Sönmez [25], Ehlers [6]). A *generalized matching problem* is a triple (N, S, R). The first component N is a finite set of agents with $n \geq 3$. The second component $S = (S_i)_{i \in N}$ is a profile of subsets of N with $i \in S_i$ for each $i \in N$. The last component $R = (R_1, ..., R_n)$ where R_i is a preference ordering for agent $i \in N$ over S_i is a preference profile. Let P_i and I_i be the asymmetric and symmetric components of R_i, respectively. Let $\mathcal{R}_i$ be the set of all preference orderings for agent $i \in N$ and $\mathcal{R} = \times_{i \in N} \mathcal{R}_i$. Given $i \in N$, let $\tilde{\mathcal{R}}_i$ denote the set of all preference orderings for agent i under which agent i is indifferent between at

[6] Moreover, Lombardia and Yoshihara[19] show that the strong core correspondence can not be implemented in Nash equilibria with partially honest agents under Assumption 1.

most two distinct assignments and $\tilde{\mathcal{R}} = \times_{i \in N} \tilde{\mathcal{R}}_i$. Throughout the paper, we fix a domain $\mathcal{D} = \times_{i \in N} \mathcal{D}_i$ where $\mathcal{D}_i$ for each $i \in N$ such that $\tilde{\mathcal{R}} \subseteq \mathcal{D} \subseteq \mathcal{R}$.

A *matching* is a bijection $\mu : N \to N$ such that each agent i's assignment $\mu(i)$ belongs to his set of possible assignments S_i. Given $T \subseteq N$, let $\mu(T) = \{\mu(i) | i \in T\}$ denote the *set of assignments of the agents in T at μ.* Let $\mathcal{M}$ denote the *set of all matchings.* Let μ^I denote the matching such that for each $i \in N$, $\mu(i) = i$. We specify a subset $\mathcal{M}^f$ of $\mathcal{M}$ as the *set of feasible matchings.* We always require that $\mu^I \in \mathcal{M}^f$ and for each $i \in N$, $S_i = \{\mu(i) | \mu \in \mathcal{M}^f\}$. In the context of matching problems, the set of allocations A is the set of feasible matchings $\mathcal{M}^f$.

We consider a situation in which the mechanism designer does not know agents' preferences. This situation is modeled by the triple $(N, \mathcal{M}^f, \mathcal{D})$, which we refer to as a *generalized matching problem environment.*

Given a preference ordering R_i of an agent $i \in N$, initially defined over S_i, we extend it to the set of feasible matchings $\mathcal{M}^f$ in the following natural way: agent i prefers the matching μ to the matching μ' if and only if he prefers his assignment under μ to his assignment under μ'. Slightly abusing notation, we use the same symbols to denote preferences over possible assignments and preferences over feasible matchings.

An SCC is a mapping $F : \mathcal{D} \twoheadrightarrow \mathcal{M}^f$ that specifies a non-empty subset $F(R) \subseteq \mathcal{M}^f$ for each $R \in \mathcal{D}$.

Let us give an example of an SCC in a generalized matching problem.

A *coalition structure* is a set $\mathcal{T} \subseteq 2^N \backslash \{\phi\}$ such that for each $i \in N$, $\{i\} \in \mathcal{T}$. Given $R \in \mathcal{D}$, $T \in \mathcal{T}$, and $\mu \in \mathcal{M}^f$, we say that *coalition T blocks μ under R* if for some $\tilde{\mu} \in \mathcal{M}^f$, (1) $\tilde{\mu}(T) = T$, (2) for each $i \in T$, $\tilde{\mu}(i) R_i \mu(i)$, and (3) for some $j \in T$, $\tilde{\mu}(i) P_i \mu(i)$.

Strong $\mathcal{T}$-core correspondence ($\mathbf{SC}^{\mathcal{T}}$): $\mathbf{SC}^{\mathcal{T}}(R) = \{\mu \in \mathcal{M}^f \mid$ there is no $T \in \mathcal{T}$ that blocks μ under $R\}$.

Ehlers [6] shows that the strong $\mathcal{T}$-core correspondence does not satisfy Maskin monotonicity. By Proposition 1, the strong $\mathcal{T}$-core correspondence cannot be doubly implemented under Assumption 0.

On the other hand, it is well-known that the strong $\mathcal{T}$-core correspondence satisfies unanimity but violates no veto power. We conclude that under Assumption 2, the strong $\mathcal{T}$-core correspondence can be doubly implemented with partially honest agents.

Corollary 3. *Let $(N, \mathcal{M}^f, \mathcal{D})$ be a generalized matching problem environment. Under Assumption 2, strong $\mathcal{T}$-core correspondence $\mathbf{SC}^{\mathcal{T}}$ can be doubly implemented with partially honest agents.*

4.4 Voting games

We consider a voting game. A *voting game* (N, A, R) contains a finite set of agents N with $n \geq 3$, a non-empty finite set of outcomes A, and a preference profile $R \in \mathcal{D}$.

We consider a situation in which the mechanism designer does not know agents' preferences. This situation is modeled by the triple $(N, A, \mathcal{D})$, which we refer to as a *voting game environment.*

Let us give an example of an SCC in a voting game.

For each $R \in \mathcal{D}$ and each $a, b \in A$, we write $aD(R)b$ if a strict majority of agents prefer a to b.

Top-cycle correspondence (TC): TC$(R) = \cap\{B \subseteq A \mid a \in B, b \notin B$ implies $aD(R)b\}$.

Palfley and Srivastava[24] show that the top-cycle correspondence does not satisfy Maskin monotonicity. By Proposition 1, the top-cycle correspondence cannot be doubly implemented under Assumption 0.

On the other hand, it is well-known that the top-cycle correspondence satisfies no veto power. We conclude that under Assumption 2, top-cycle correspondence can be partially honest doubly implemented.

Corollary 4. *Let $(N, A, \mathcal{D})$ be a voting game environment. Under Assumption 1,* the top-cycle correspondence **TC** *can be doubly implemented with partially honest agents.*

Let us give the other example of an SCC in a voting game.

For each $R \in \mathcal{D}$, let $B^i(a, R) = k$ if $a \in A$ is the k'th most preferred outcome.

Borda correspondence ($\mathbf{F}_B$): $\mathbf{F}_B(R)$ $= \{a \in A : \Sigma_{i \in N} B^i(a, R) \leq \Sigma_{i \in N} B^i(b, R)$ for each $b \in A\}$.

The following example represents that the Borda correspondence violates Maskin monotonicity and no veto power.

Example 5. Consider the following example. There are three agents, $N = \{1, 2, 3\}$, two outcomes, $A = \{a, b, c\}$, and two possible preference profiles, $\mathcal{D} = \{R, R'\}$. Preferences are given by :

R_1	R_2	R_3	R'_1	R'_2	R'_3
a	b	c	a	a, b, c	c
b	c	b	b, c		b
c	a	a			a

The Borda correspondence evaluated at each preference profile is as follows: $\mathbf{F}_B(R) = \{b\}$ and $\mathbf{F}_B(R') = \{c\}$. Since Maskin monotonicity and no veto power imply that we must have $b \in \mathbf{F}_B(R')$ and $a \in \mathbf{F}_B(R')$, respectively, the Borda correspondence fails the conditions. By Proposition 1, the Borda correspondence $\mathbf{F}_B$ cannot be doubly implemented under Assumption 0.■

On the other hand, it is well-known that the Borda correspondence satisfies unanimity. We conclude that under Assumption 2, the Borda correspondence can be doubly implemented with partially honest agents.

Corollary 5. *Let $(N, A, \mathcal{D})$ be a voting game environment. Under Assumption 2, the Borda correspondence F_B can be doubly implemented with partially honest agents.*

5 A simple mechanism for double implementation with socially responsible agents

We have considered a truth-telling correspondence with regard to a preference profile following Dutta and Sen [5]. On the other hand, motivated by some studies (Gneezy[7], Hurkens and Kartik[10], Matsushima[21], and Doğan[4]), Hagiwara et al.[8] consider a truth-telling with regard to an outcome.

They design a simple and natural mechanism for implementation in Nash equilibria with socially responsible agents which they call the *outcome mechanism* $\Gamma^O = (M, g)$. The message space of agent $i \in N$ consists of $M_i = A \times N$. Denote an element of M_i by $m_i = (a^i, k^i)$. The outcome function $g : M \to A$ is defined as follows:

Rule 1 : If there is $i \in N$ such that for each $j \neq i$, $m_j = (a, k^j)$, then $g(m) = a$.

Rule 2 : In all other cases, $g(m) = a^{i^*}$, where $i^* = (\Sigma_{i\in N} k^i)(mod\ n) + 1$.

In the outcome mechanism, the mechanism designer expects each agent to report a socially desirable outcome. Then, some agent may strictly prefer to report a socially desirable outcome at the true preference profile to the mechanism designer whenever announcing a socially undesirable outcome does not change the outcome to a more preferred outcome. They call such an agent a *socially responsible* agent. Given an SCC F, let $\succsim_i^{F(R)}$ be a preference ordering for agent $i \in N$ over M at $F(R)$, whose asymmetric and symmetric components are $\succ_i^{F(R)}$ and $\sim_i^{F(R)}$, respectively.

Definition 6. An agent $i \in N$ is *socially responsible* if for each $R \in \mathcal{R}$ and each $(m_i, m_{-i}), (m'_i, m_{-i}) \in M$ such that $m_i = (a^i, k^i)$ and $m'_i = (a'^i, k'^i)$, the following properties hold:
(1) If $a^i \in F(R)$, $a'^i \notin F(R)$, and $g(m_i, m_{-i}) R_i g(m'_i, m_{-i})$, then $(m_i, m_{-i}) \succ_i^{F(R)} (m'_i, m_{-i})$.
(2) In all other cases, $(m_i, m_{-i}) \succsim_i^{F(R)} (m'_i, m_{-i})$ if and only if $g(m_i, m_{-i}) R_i g(m'_i, m_{-i})$.

We consider the following assumption:

Assumption *n*. There are n socially responsible agents in N.[7]

Let $(\Gamma^O, \succsim^{F(R)})$ be a *game with socially responsible agents* induced by the outcome mechanism Γ^O and a preference profile $\succsim^{F(R)}$.

Definition 3. Under Assumption n, a mechanism Γ *doubly implements an SCC F with socially responsible agents* if for each $R \in \mathcal{D}$, $F(R) = NE_A(\Gamma^O, \succsim^{F(R)}) = UNE_A(\Gamma^O, \succsim^{F(R)}$).

We show that if an SCC F is single-valued, the outcome mechanism Γ^O doubly implements F satisfying unanimity with socially responsible agents.

Theorem 2. *Under Assumption n, the outcome mechanism Γ^O doubly implements any single-valued SCC F satisfying unanimity with socially responsible agents.*

The proof of Theorem 2 is given in Appendix B.

[7]Some studies such as Kimya [17] introduce Assumption n for partially honest agents.

5.1 Related literature

For Nash implementation with socially responsible agents, Hagiwara et al.[8] provide the following result under the assumptions:

Assumption 1. There exists at least *one* socially responsible agent in N.
Assumption 2. There are at least *two* socially responsible agents in N.

Proposition 3. (Hagiwara et al.[8]) *Let* $n \geq 3$.
(1) *Under Assumption 1, the outcome mechanism* Γ^O *implements* F *satisfying no veto power in Nash equilibria with socially responsible agents.*
(2) *Under Assumption 2, the outcome mechanism* Γ^O *implements* F *satisfying unanimity in Nash equilibria with socially responsible agents.*

The following example shows that if $|F(R)| > 1$ for some $R \in \mathcal{D}$, the outcome mechanism cannot doubly implement the SCC F with socially responsible agents.

Example 3. Consider the following example. There are three agents, $N = \{1, 2, 3\}$ such that all agents are socially responsible, three outcomes, $A = \{a, b, c\}$, and two admissible preference profiles, $\mathcal{D} = \{R, R'\}$. Preferences are given by:

R_1	R_2	R_3		R'_1	R'_2	R'_3
a	c	c		a	c	c
b, c	b	b		b	a	b
	a	a		c	b	a

Define the SCC F as follows: $F(R) = \{a, b\}$, and $F(R') = \{a\}$. Note that the SCC F satisfies unanimity, so that by Proposition 3 (2), the SCC can be implemented in Nash equilibria with socially responsible agents by their outcome mechanism under Assumption 2.

There exist two Nash equilibrium outcomes in $(\Gamma^O, \succsim^{F(R)})$, a and b. However, it is easy to see that $m_1 = (b, k^1)$ is weakly dominated by $m'_1 = (a, k^1)$ at $\succsim_1^{F(R)}$. Therefore, $F(R) = \{a, b\} = NE_A(\Gamma^O, \succsim^{F(R)})$ but $UNE_A(\Gamma^O, \succsim^{F(R)}) = \{a\}$.$\blacksquare$

We give an example to show that in the outcome mechanism, even if two agents are socially responsible, there may be Nash equilibrium outcomes in which agents use weakly dominated messages, and hence the set of undominated Nash equilibrium outcomes may be a proper subset of the set of Nash equilibrium outcomes. Although Assumption n is stronger, the outcome mechanism solves some problems with respect to Dutta and Sen's mechanism such as the assumption of complete information.[8]

Example 4. Consider the following example. There are three agents, $N = \{1, 2, 3\}$ such that agent 2 and agent 3 are socially responsible, three outcomes, $A = \{a, b.c\}$, and two admissible preference profiles, $\mathcal{D} = \{R, R'\}$. Preferences are given by:

[8]Hagiwara et al.[8] show that in Example, as long as an event is common knowledge, all agents can commonly know the set of Nash equilibria. See Hagiwara et al.[8].

R_1	R_2	R_3		R'_1	R'_2	R'_3
a	c	c		a	c	c
b	b	b		b	a	b
c	a	a		c	b	a

Define the SCC F as follows: $F(R) = \{c\}$, and $F(R') = \{a\}$. Note that the SCC F satisfies unanimity, so that by Proposition 3 (2), the SCC can be implemented in Nash equilibria with socially responsible agents by their outcome mechanism under Assumption 2.

There exists a unique Nash equilibrium outcome with socially responsible agents in $(\Gamma^O, \succsim^{F(R)})$, c. If $m_i = (c, k^i)$ for each $i \in N$, $m \in NE(\Gamma^O, \succsim^{F(R)})$ and it is easy to see that the message m_2 and m_3 is not weakly dominated by any message in M_2 and M_3 *at* $\succsim_2^{R,H}$ and $\succsim_3^{R,H}$, respectively. On the other hand, it is easy to see that $m_1 = (c, k^1)$ is weakly dominated by $m'_1 = (a, k^1)$ at $\succsim_1^{R,H}$. Therefore, $F(R) = \{c\} = NE_A(\Gamma^O, \succsim^{F(R)})$ but $UNE_A(\Gamma^O, \succsim^{F(R)}) = \emptyset.\blacksquare$

For our result and a previous result in which we focus on single-valued SCCs, we summarize sufficient conditions for double implementation with socially responsible agents under Assumption $k \in \{1, n\}$ in Figure 2 below.

	Assumption 1	$\longrightarrow$	Assumption n
Nash Implementation	Hagiwara et al.[8] *no veto power*		Hagiwara et al.[8] *unanimity*
$\downarrow$			
Double Implementation	This paper (Example 4) $\times$		This paper (Theorem 2) *unanimity*

Figure 2

This setting is similar to that of Kartik et al.[15], but not same: while they consider a problem in which the dictator strictly prefers the socially optimal outcome and all the other agents have preferences for honesty, we assume that all agents just want to report the socially optimal outcome. Therefore, we can apply to more problems than Kartik et al.[15] because we do not use an assumption of preferences on A. Moreover, they assume that any agent can observe all the other agents' preferences. However, in the outcome mechanism, the complete information assumption can be weakened.

6 Appendix A

Proof of Theorem 1 (1): Let F be an SCC satisfying no veto power. We construct a mechanism $\Gamma = (M, g)$. For each $i \in N$, the message space of agent $i \in N$ consists of $M_i = \mathcal{D} \times A \times A \times \{-n, ..., -1, 0, 1, ..., n\}$. Denote an element of M_i by $m_i = (R^i, a^i, b^i, k^i)$.

For each agent $i \in N$ and each $R_i \in \mathcal{D}_i$, define $\bar{\mathrm{b}}(R_i)$ and $\underline{\mathrm{b}}(R_i)$ as follows: (1) if there exist $b, c \in A$ such that bP_ic, then let $\bar{\mathrm{b}}(R_i) = b$ and $\underline{\mathrm{b}}(R_i) = c$; (2) otherwise, pick any $b, c \in A$ with $b \neq c$, let $\bar{\mathrm{b}}(R_i) = b$ and $\underline{\mathrm{b}}(R_i) = c$.

The outcome function $g : M \longrightarrow A$ is defined as follows:

Rule 1 : If there exists $i \in N$ such that for each $j \neq i$, $m_j = (R, a, \cdot, j)$ where $a \in F(R)$, then $g(m) = a$.

Rule 2 : If there exists $i \in N$ such that for each $j \neq i$, $m_j = (R, a, \cdot, -i)$ where $a \in F(R)$, then

$$g(m) = \begin{cases} \bar{\mathrm{b}}(R_i) \text{ if } m_i = (R, a, \bar{b}(R_i), i) \\ \underline{\mathrm{b}}(R_i) \text{ if } m_i \neq (R, a, \bar{b}(R_i), i) \text{ with } k^i \leq 0 \text{ or } k^i = i. \end{cases}$$

Rule 3 : In all other cases, $g(m) = a^{i^*}$, where $i^* = (\Sigma_{i \in N} \max\{0, k^i\})(mod\ n) + 1$.

For each $i \in N$ and each $R \in \mathcal{D}$, a truth-telling correspondence is defined by $T_i^\Gamma(R, F) = \{R\} \times A \times A \times \{-n, ..., -1, 0, 1, ..., n\}$.

The proof consists of three lemmata.

Lemma 1. Let $R \in \mathcal{D}$, $H \in \mathcal{H}^1$, and $a \in F(R)$ be given. If for each $i \in N$, $m_i = (R, a, \bar{\mathrm{b}}(R_i), i)$, then $m \in NE(\Gamma, (T_i^\Gamma)_{i \in N}, \succsim^{R,H})$.

Proof: For each $i \in N$, let $m_i = (R, a, \bar{\mathrm{b}}(R_i), i)$. By Rule 1, $g(m) = a$. No unilateral deviation can change the outcome and $m_i \in T_i^\Gamma(R)$ for each $i \in N$. Hence, $m \in NE(\Gamma, (T_i^\Gamma)_{i \in N}, \succsim^{R,H})$.■

Lemma 2. Let $R \in \mathcal{D}$, $H \in \mathcal{H}^1$, and $a \in F(R)$ be given. For each $i \in N$, $m_i = (R, a, \bar{b}(R_i), i)$ is undominated at $\succsim_i^{R,H}$.

Proof: First, suppose that there exist $b, c \in A$ with bP_ic. Then, $\bar{\mathrm{b}}(R_i)P_i\underline{\mathrm{b}}(R_i)$. We show that for each $\tilde{m}_i \neq m_i$, there exists $\tilde{m}_{-i} \in M_{-i}$ such that $(m_i, \tilde{m}_{-i}) \succ_i^{R,H} (\tilde{m}_i, \tilde{m}_{-i})$. There are two cases to consider.

Case 1. $\tilde{k}^i \leq 0$ or $\tilde{k}^i = i$.

For each $j \neq i$, let $\tilde{m}_j = (R, a, \cdot, -i)$. By Rule 2, $g(m_i, \tilde{m}_{-i}) = \bar{\mathrm{b}}(R_i)$ and $g(\tilde{m}_i, \tilde{m}_{-i}) = \underline{\mathrm{b}}(R_i)$, so that $(m_i, \tilde{m}_{-i}) \succ_i^{R,H} (\tilde{m}_i, \tilde{m}_{-i})$.

Case 2. $\tilde{k}^i > 0$ and $\tilde{k}^i \neq i$.

Define $\tilde{m}_{-i} \in M_{-i}$ as follows: for some $j \neq i$, $\tilde{m}_j = (R', a', \underline{\mathrm{b}}(R_i), j-1)$, for some $h \neq i, j$, $\tilde{m}_h = (R'', a'', \underline{\mathrm{b}}(R_i), \tilde{k}^h)$, and for any other ℓ, $\tilde{m}_\ell = (\cdot, \cdot, \underline{\mathrm{b}}(R_i), \tilde{k}^\ell)$, where $(R, a) \neq (R', a') \neq (R'', a'')$ and $(\Sigma_{h \neq i,j}\tilde{k}^h + i + j - 1)(mod\ n) + 1 = i$ with $\tilde{k}^h \geq 0$ for $h \neq i, j$. By Rule 3, $g(m_i, \tilde{m}_{-i}) = \bar{\mathrm{b}}(R_i)$ and $g(\tilde{m}_i, \tilde{m}_{-i}) = \underline{\mathrm{b}}(R_i)$, so that $(m_i, \tilde{m}_{-i}) \succ_i^{R,H} (\tilde{m}_i, \tilde{m}_{-i})$.

Next, suppose that for each $b, c \in A$, bI_ic. Obviously, m_i is undominated *at* $\succsim_i^{R,H}$.■

Lemma 3. For each $R \in \mathcal{D}$ and each $H \in \mathcal{H}^1$, $NE_A(\Gamma, (T_i^\Gamma)_{i \in N}, \succsim^{R,H}) \subseteq F(R)$.

Proof: There are two cases to consider.

Case 1. For each $i \in N$, $m_i = (R', a, \cdot, i)$ such that $R' \neq R$ and $a \in F(R')$.

We show that if $g(m) \notin F(R)$, then $m \notin NE(\Gamma, (T_i^\Gamma)_{i \in N}, \succsim^{R,H})$. Under Assumption 2, there exists a partially honest agent $h \in H$. Let $m'_h = (R, a'^h, b'^h, k'^h)$. By the definition of the truth-telling correspondence, $m_h \notin T_h^\Gamma(R)$ and $m'_h \in T_h^\Gamma(R)$. By Rule 1, $g(m'_h, m_{-h}) = a$ so that $g(m'_h, m_{-h}) = g(m)$. Since $h \in H$, $(m'_h, m_{-h}) \succ_h^{R,H} (m_h, m_{-h})$.

Hence, $m \notin NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$.

Case 2. There are $i, j \in N$ $(i \neq j)$ such that $R^i \neq R^j$.

Let the outcome be some $b \in A$. Then, any one of $(n-1)$ agents can deviate, precipitate the modulo game, and be the winner of the modulo game. Clearly, if the original announcement is to be a Nash equilibrium with partially honest agents, then it must be the case that $L(R_i, b) = A$ for $(n-1)$ agents. Then since F satisfies no veto power, $b \in F(R)$.■

Proof of Theorem 1 (2): Let F be an SCC satisfying unanimity. We use the same mechanism $\Gamma = (M, g)$ as the proof of Theorem 1 (1) and for each $i \in N$ and each $R \in \mathcal{D}$, a truth-telling correspondence is defined by $T_i^\Gamma(R) = \{R\} \times A \times A \times \{-n, ..., -1, 0, 1, ..., n\}$.

The proof consists of three lemmata.

Lemma 4. Let $R \in \mathcal{D}$, $H \in \mathcal{H}^2$, and $a \in F(R)$ be given. If for each $i \in N$, $m_i = (R, a, \bar{\mathrm{b}}(R_i), i)$, then $m \in NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$.

Lemma 5. Let $R \in \mathcal{D}$, $H \in \mathcal{H}^2$, and $a \in F(R)$ be given. For each $i \in N$, $m_i = (R, a, \bar{b}(R_i), i)$ is undominated *at* $\succsim_i^{R,H}$.

The proof of Lemma 4 and Lemma 5 are omitted. It follows from the same reasoning as Lemma 1 and Lemma 2, respectively.

Lemma 6. For each $R \in \mathcal{D}$ and each $H \in \mathcal{H}^2$, $NE_A(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H}) \subseteq F(R)$.

Proof: We show that if $g(m) \notin F(R)$, then $m \notin NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$. There are four cases to consider.

Case 1. For each $i \in N$, $m_i = (R', a, \cdot, i)$ such that $R' \neq R$ and $a \in F(R')$.

By the same argument as Case 1 of Lemma 3, $m \notin NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$.

Case 2. There is $i \in N$ such that for each $j \neq i$, $m_j = (R', a, \cdot, j)$ such that $R' \neq R$ and $a \in F(R')$, and $m_i \neq (R', a, \cdot, i)$.

By Rule 1, $g(m) = a \in F(R')$ such that $R' \neq R$. Under Assumption 2, since $|H| \geq 2$ there exists a partially honest agent $h \in H \backslash \{i\}$.[9] Without loss of generality, let $i = 1$ and $h = 2$. Let $m_2' = (R, a'^2, b'^2, k'^2)$ be such that $(\Sigma_{j\neq 2} k^j + k'^2)(mod\ n) + 1 = 3$. By the definition of the truth-telling correspondence, $m_2 \notin T_2^\Gamma(R)$ and $m_2' \in T_2^\Gamma(R)$. By Rule 3, $g(m_2', m_{-2}) = a^3 = a$ so that $g(m_2', m_{-2}) = g(m)$. Since agent 2 is partially honest, $(m_2', m_{-2}) \succ_2^{R,H} (m_2, m_{-2})$. Hence, $m \notin NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$.

Case 3. Rule 2 is applied.

Suppose $g(m) \notin F(R)$. Since F satisfies unanimity, there is $\ell \in N$ and $b \in A$ such that $bP_\ell g(m)$. Suppose $\ell = i$. Let $m_i' = (\cdot, \cdot, b, i-1)$ if $i \neq 1$ and $m_i' = (\cdot, \cdot, b, n)$ if $i = 1$. By Rule 3, $g(m_i', m_{-i}) = b$ so that $g(m_i', m_{-i}) P_i g(m)$. Otherwise (i.e. $\ell \neq i$), if agent ℓ deviate to $m_\ell' = (\cdot, \cdot, b, k'^\ell) \neq m_\ell$ such that $(\Sigma_{j\neq\ell} k^j + k'^\ell)(mod\ n) + 1 = \ell$, then by Rule

[9]Note that under Assumption 1, there is no partially honest agent in $N\backslash\{i\}$ when $|H| = 1$ and agent i is partially honest.

3, $g(m'_\ell, m_{-\ell}) = b$ so that $g(m'_\ell, m_{-\ell})P_\ell g(m)$. Whether agent ℓ is partially honest or not, $(m'_\ell, m_{-\ell}) \succ_\ell^{R,H} (m_\ell, m_{-\ell})$. Hence, $m \notin NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$.

Case 4. Rule 3 is applied.

Suppose $g(m) \notin F(R)$. Since F satisfies unanimity, there is $i \in N$ and $b \in A$ such that $bP_i g(m)$. Let $m'_i = (\cdot, \cdot, b, k'^i) \neq m_i$ be such that $(\Sigma_{j\neq i} k^j + k'^i)(mod\ n) + 1 = i$. By Rule 3, $g(m'_i, m_{-i}) = b$ so that $g(m'_i, m_{-i})P_i g(m)$. Whether agent i is partially honest or not, $(m'_i, m_{-i}) \succ_i^{R,H} (m_i, m_{-i})$. Hence, $m \notin NE(\Gamma, (T_i^\Gamma)_{i\in N}, \succsim^{R,H})$.■

7 Appendix B

Proof of Theorem 2: Let F be an single-valued SCC satisfying unanimity. For each $R \in \mathcal{R}$, let $a \in A$ be such that $F(R) = \{a\}$.

The proof consists of three lemmata.

Lemma 7. If for each $i \in N$, $m_i = (a, k^i)$, then $m \in NE(\Gamma^O, \succsim^{F(R)})$.

Proof: For each $i \in N$, let $m_i = (a, k^i)$. By Rule 1, $g(m) = a$. No unilateral deviation can change the outcome and $a^i = a \in F(R)$ for each $i \in N$. Hence, $m \in NE(\Gamma^O, \succsim^{F(R)})$.■

Lemma 8. For each $i \in N$, $m_i = (a, k^i)$ is undominated at $\succsim_i^{F(R)}$.

Proof: First, suppose that there exist $b, c \in A$ with $bP_i a$ and $bP_i c$. We show that for each $\tilde{m}_i \neq m_i$, there exists $\tilde{m}_{-i} \in M_{-i}$ such that $(m_i, \tilde{m}_{-i}) \succ_i^{F(R)} (\tilde{m}_i, \tilde{m}_{-i})$. There are two cases to consider.

Case 1. $\tilde{m}_i = (b, \cdot)$ such that $b \neq a$.

For each $j \neq i$, let $\tilde{m}_j = (b, \cdot)$ By Rule 1, $g(m_i, \tilde{m}_{-i}) = g(\tilde{m}_i, \tilde{m}_{-i}) = b$. Since $a \in F(R)$ and $b \notin F(R)$, $(m_i, \tilde{m}_{-i}) \succ_i^{F(R)} (\tilde{m}_i, \tilde{m}_{-i})$.

Case 2. $\tilde{m}_i = (a, \tilde{k}^i)$ such that $\tilde{k}^i \neq k^i$.

Define $\tilde{m}_{-i} \in M_{-i}$ as follows: for some $j \neq i$, $\tilde{m}_j = (b, \bar{k}^j)$, and for any other $h \neq i, j$, $\tilde{m}_h = (c, \bar{k}^h)$, where $b \neq c$ and $(\Sigma_{j\neq i} \tilde{k}^j + k^i)(mod\ n) + 1 = j$. By Rule 3, $g(m_i, \tilde{m}_{-i}) = b$ and $g(\tilde{m}_i, \tilde{m}_{-i}) = a$ or c, so that $(m_i, \tilde{m}_{-i}) \succ_i^{F(R)} (\tilde{m}_i, \tilde{m}_{-i})$.

Next, suppose that for each $b, c \in A$, $bI_i c$ or $a \in L(R_i, a) = A$. Obviously, m_i is undominated at $\succsim_i^{F(R)}$.■

Lemma 9. For each $R \in \mathcal{D}$, $NE_A(\Gamma^O, \succsim^{F(R)}) \subseteq F(R)$.

Proof: We show that if $g(m) \notin F(R)$, then $m \notin NE(\Gamma^O, \succsim^{F(R)})$. There are two cases to consider.

Case 1. For each $b \neq a$, Rule 1 is applied.

By Rule 1, $g(m) = b$. Let $m'_i = (a, \cdot)$. By Rule 1, $g(m'_i, m_{-i}) = b$ so that $g(m'_i, m_{-i}) = g(m)$. Since i is a socially responsible agent, $(m'_i, m_{-i}) \succ_i^{F(R)} (m_i, m_{-i})$. Hence, $m \notin NE(\Gamma^O, \succsim^{F(R)})$.

Case 2. In all other cases, Rule 2 is applied.

Suppose $g(m) \notin F(R)$. Since F satisfies unanimity, there is $i \in N$ and $b \in A$ such that $bP_i g(m)$. Let $m'_i = (b, k'^i) \neq m_i$ be such that $(\Sigma_{j \neq i} k^j + k'^i)(mod\ n) + 1 = i$. By Rule 2, $g(m'_i, m_{-i}) = b$ so that $g(m'_i, m_{-i}) P_i g(m)$. Then, $(m'_i, m_{-i}) \succ_i^{F(R)} (m_i, m_{-i})$. Hence, $m \notin NE(\Gamma^O, \succsim^{F(R)})$.■

References

[1] Attiyeh G, Franciosi R, and Isaac R.M (2000), "AExperiments with the pivotal process for providing public goods" *Public Choice* 102: 95–114.

[2] Cason T, Saijo T, Sjöström T, and Yamato T (2006), "ASecure implementation experiments: Do strategy-proof mechanisms really work?" *Games and Economic Behavior* 57: 206-235.

[3] Corchón L, Herrero C (2004), "A decent proposal" *Spanish Economic Review* 4: 1236-1239.

[4] Doğan B (2013), "Eliciting the Socially Optimal Allocation from Responsible Agents" *mimeo.*

[5] Dutta B, Sen A (2012), "Nash implementation with partially honest individuals" *Games and Economic Behavior* 74: 154-169.

[6] Ehlers T (2004), "Monotonic and implementable solutions in generalized matching problems" *Journal of Economic Theory* 114: 358-369.

[7] Gneezy U (2005), "ADeception: The role of consequences" *American Economic Review* 95: 384-394.

[8] Hagiwara M, Yamamura H, and Yamato T (2017), "Implementation with socially responsible agents" *Economic Theory Bulletin*, forthcoming.

[9] Harstad R M (2000), "ADominant strategy adoption and bidders' experience with pricing rules" *Experimental Economics* 3: 261–280.

[10] Hurkens S, Kartik N (2009), "Would I lie to you? On social preferences and lying aversion" *Experimental Economics* 12: 180-192.

[11] Inoue F, Yamamura H (2017), "The binary mechanism for the allocation problem with single-dipped preferences" *mimeo.*

[12] Jackson M, Palfrey T, and Srivastava S (1994), "AUndominated Nash Implementation in Bounded Mechanisms" *Games and Economic Behavior* 6: 474-501.

[13] Kagel J H, Levin D (1993), "AIndependent private value auctions: Bidder behavior in first-, second- and third-price auctions with varying number of bidders" *Economic Journal* 103: 868–879.

[14] Kagel J H, Harstad R M, Levin D (1987), "AInformation impact and allocation rules in auctions with affiliated private values: A laboratory study" *Econometrica* 55: 1275-1304.

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

[16] Kawagoe T, Mori T (2001), "Can the pivotal mechanism induce truth-telling? An experimental study" *Public Choice* 108: 331–354.

[17] Kimya M (2015), "Nash Implementation and Tie-Breaking Rules" *mimeo.*

[18] Lombardi M, Yoshihara N (2011), "Partially-honest Nash implementation: characterization results" *mimeo.*

[19] Lombardi M, Yoshihara N (2014), "Nash implementation with partially-honest agents: A full characterization" *mimeo.*

[20] Lombardi M, Yoshihara N (2017), "Natural implementation with semi-responsible agents in pure exchange economies" *International Journal of Game Theory*, forthcoming.

[21] Matsushima H (2008), "Behavioral aspects of implementation theory" *Economics Letters* 100: 161-164.

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

[23] Mukherjee S, Muto N, and Ramaekers E (2017), "Implementation in undominated strategies with partially honest agents" *Games and Economic Behavior* 104: 613-631.

[24] Palfrey T, Srivastava S (1991), "Nash implementation using undominated strategies" *Econometrica* 59:479-501.

[25] Sönmez T (1996), "Implementation in generalized matching problems" *Journal of Mathematical Economics* 26: 429-439.

[26] Tatamitani Y (1993), "Double implementation in Nash and undominated Nash equilibria in social choice environments" *Economic Theory* 3: 109–117.

[27] Yamato T (1993), "Double implementation in Nash and undominated Nash equilibria" *Journal of Economic Theory* 59: 311–323.

[28] Yamato T (1999), "Nash implementation and double implementation: equivalence theorems" *Journal of Mathematical Economics* 31: 215–238.