The revelation principle fails when the format of each agent's strategy is an action

Wu, Haoyang

Wan-Dou-Miao Research Lab

8 December 2020

The revelation principle fails when the format of each agent's strategy is an action

Haoyang Wu

Wan-Dou-Miao Research Lab, Shanghai, China.

Abstract

In mechanism design theory, a designer would like to implement a social choice function which specifies her favorite outcome for each possible profile of agents' private types. The revelation principle asserts that if a social choice function can be implemented by a mechanism in equilibrium, then there exists a direct mechanism that can truthfully implement it.

This paper aims to propose a failure of the revelation principle. We point out that in any game the format of each agent's strategy is either an informational message or a realistic action, and the action format is very common in many practical cases. The main result is that: For any given social choice function, if the mechanism which implements it has action-format strategies, then *"honest and obedient"* will no longer be the Bayesian Nash equilibrium of the direct mechanism, actually the social choice function can only be implemented *"dishonestly and disobediently"* in Bayesian Nash equilibrium by the direct mechanism. Consequently, the revelation principle fails when the format of each agent's strategy is an action.

Key words: Mechanism design; Revelation principle.

1 Introduction

In the framework of mechanism design theory [1–3], there are one designer and some agents labeled as $1, \cdots, I$. [1] Suppose that the designer would like to implement a social choice function which specifies her favorite outcome for each possible profile of agents' types, and each agent's type is modeled as his privacy. In order to implement a social choice function in equilibrium, the

---

Email address: 18621753457@163.com (Haoyang Wu).

[1] In this paper, the designer is always denoted as "She", and the agent is denoted as "He".

designer constructs a mechanism which specifies each agent's feasible strategy set (*i.e.*, the allowed actions of each agent) and an outcome function (*i.e.*, a rule for how agents' actions get turned into a social choice).

The revelation principle is an important theorem in mechanism design theory. It asserts that if a social choice function can be implemented by a mechanism in equilibrium, then it is truthfully implementable. So far, there have been several criticisms on the revelation principle: Bester and Strausz [4] pointed out that the revelation principle may fail because of imperfect commitment; Epstein and Peters [5] proposed that the revelation principle fails in situations where several mechanism designers compete against each other. Kephart and Conitzer [6] proposed that when reporting truthfully is costless and misreporting is costly, the revelation principle can fail to hold.

Different from these criticisms, this paper aims to propose another failure of the revelation principle. The paper is organized as follows: Section 2 analyses two formats of strategy and points out that the action format is very common in many practical cases. Section 3 proposes the main result, *i.e.*, the revelation principle fails when each agent's strategy is action-format. Section 4 draws conclusions. Notations about mechanism design theory are given in Appendix, which are cited from Ref [1].

2   Two formats of strategy

Note 1: In any game, the format of each agent's strategy is either an informational message or a realistic action. □

Although the note looks naive, it is not trivial. The reason why we highlight the distinction of two formats of strategy is that the revelation principle does not hold for the case of action-format strategies, as will be discussed deeply in Section 3. For simplification, in the following discussions we simply assume that in any game all agents' strategies are of the same format, *i.e.*, we omit the case in which some agents' strategies are message-format and other agents' strategies are action-format. Next, we will deeply investigate the two formats of strategy respectively.

2.1   Case 1: Mechanism with message-format strategies

Definition 1: A message-format strategy is represented by an informational message. The information itself contained in the message is just the strategy, which does not need to be carried out realistically in the mechanism.

Practically, only in some restricted cases can each agent's strategy be described as pure information and represented by an informational message. For example, let us consider a chess game, then each player's strategy is message-format, since it is a strategic plan about how to play chess. Similarly, the strategy in a war simulation game is also message-format, since it contains military *plans* of players.

Definition 2: Given a social choice function f, suppose a mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ implements it in Bayesian Nash equilibrium with message-format strategies. To clearly describe the message format, we denote each agent i's strategy set S_i as M_i, and each strategy function as $m_i(\cdot) : \Theta_i \to M_i$, where Θ_i is agent i's type set. The outcome function $g(\cdot)$ is denoted as $g_m(\cdot) : M_1 \times \cdots \times M_I \to X$, where the parameters are message-format strategies and X is the set of outcomes. Hence, the mechanism Γ is rewritten as $\Gamma_m = (M_1, \cdots, M_I, g_m(\cdot))$. The game induced by Γ_m is denoted as G_m, which works in a one-stage manner:
Step 1: By using the strategy function $m_i(\cdot) : \Theta_i \to M_i$, each agent i with private type θ_i sends a message $m_i(\theta_i)$ to the designer. [2]
Step 2: The mechanism Γ_m yields the outcome $g_m(m_1(\theta_1), \cdots, m_I(\theta_I))$.
Here, each agent i's utility is denoted as $u_i(g_m(m_1(\theta_1), \cdots, m_I(\theta_I)), \theta_i)$.

Definition 3: Suppose the game G_m has a Bayesian Nash equilibrium, denoted as $m^*(\cdot) = (m_1^*(\cdot), \cdots, m_I^*(\cdot))$, *i.e.*, for all i and all $\theta_i \in \Theta_i$, $\hat{m}_i \in M_i$,

$$E_{\theta_{-i}}[u_i(g_m(m_i^*(\theta_i), m_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g_m(\hat{m}_i, m_{-i}^*(\theta_{-i})), \theta_i)|\theta_i].$$

Consider this equilibrium, there is a compound mapping $g_m(m^*(\cdot)) : \Theta \to X$ from agents' possible types $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_I) \in \Theta$ into the outcome $g_m(m^*(\hat{\theta}))$, which is equal to $f(\hat{\theta})$ for any $\hat{\theta} \in \Theta$. Based on the compound mapping, we construct a direct mechanism $\bar{\Gamma}_m = (\Theta_1, \cdots, \Theta_I, g_m(m^*(\cdot)))$ with message-format strategies. [3]

Definition 4: The direct mechanism $\bar{\Gamma}_m$ induces a *one-stage* direct game $\bar{G}_m$, which works as follows:
Step 1: Each agent i with private type θ_i individually reports an arbitrary

---

[2] In the following discussions, we denote each agent i's true type as θ_i, and his any possible type as $\hat{\theta}_i \in \Theta_i$.
[3] Although $g_m(m^*(\hat{\theta})) = f(\hat{\theta})$ for any $\hat{\theta} \in \Theta$, the outcome function of the constructed direct mechanism $\bar{\Gamma}_m$ must be the compound mapping $g_m(m^*(\cdot)) : \Theta \to X$ instead of $f(\cdot) : \Theta \to X$. The reason is straightforward: if the outcome function of $\bar{\Gamma}_m$ is simply written as $f(\cdot)$, then $\bar{\Gamma}_m$ will become a naive direct mechanism $(\Theta_1, \cdots, \Theta_I, f(\cdot))$, which is irrelevant to the mechanism $\Gamma_m = (M_1, \cdots, M_I, g_m(\cdot))$ which implements f in equilibrium. Indeed, the formula $g_m(m^*(\hat{\theta})) = f(\hat{\theta})$ (for any $\hat{\theta} \in \Theta$) is the result of the mechanism Γ_m, and the naive direct mechanism $(\Theta_1, \cdots, \Theta_I, f(\cdot))$ cannot implement $f(\cdot)$ in Bayesian Nash equilibrium at all.

type $\hat{\theta}_i \in \Theta_i$. Here, each agent i does not need to be "*honest*", *i.e.*, $\hat{\theta}_i$ can be different from agent i's private type θ_i.

Step 2: By using the equilibrium strategy functions $m^*(\cdot) = (m_1^*(\cdot), \cdots, m_I^*(\cdot))$, the direct mechanism $\bar{\Gamma}_m$ calculates $m^*(\hat{\theta}) = (m_1^*(\hat{\theta}_1), \cdots, m_I^*(\hat{\theta}_I))$, and then yields the outcome $g_m(m^*(\hat{\theta}))$.

Note 2: Obviously, the calculated results $m^*(\hat{\theta}) = (m_1^*(\hat{\theta}_1), \cdots, m_I^*(\hat{\theta}_I))$ are pure information, and hence are message-format. Actually, only when each agent i's strategy $m_i(\cdot)$ is message-format can $m^*(\hat{\theta})$ be *legal* parameters of the outcome function $g_m(\cdot)$.

Note 3: By Definition 3, $m^*(\cdot) = (m_1^*(\cdot), \cdots, m_I^*(\cdot))$ is the equilibrium of the game G_m, then $m_i^*(\theta_i)$ is the optimal choice for each agent i given that all other agents send $m_{-i}^*(\theta_{-i})$. Therefore, in the direct game $\bar{G}_m$, each agent i will find truth-telling $\hat{\theta}_i = \theta_i$ to be the optimal choice given that the others agents tell the truth $\hat{\theta}_{-i} = \theta_{-i}$, and the final outcome will be $g_m(m^*(\theta))$, which is equal to $f(\theta)$ for all $\theta \in \Theta$. Thus, for the case of message-format strategies, truth-telling is a Bayesian Nash equilibrium of the direct game $\bar{G}_m$. Consequently, *the revelation principle holds when each agent's strategy is message-format.* □

2.2 Case 2: Mechanism with action-format strategies

Definition 5: An action-format strategy is represented by a realistic action, which should be performed practically.

In many common cases each agent's strategy must be described as a realistic action instead of an informational message. For example, let us consider a tennis game, then each player's strategy is his realistic action of playing tennis, but not any imaginary plan about how to play tennis. Compared with the message-format strategy in a war simulation game, the strategy in a real war is action-format, since it contains military *actions* of armies.

Definition 6: Given a social choice function f, suppose a mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ implements it in Bayesian Nash equilibrium with action-format strategies. To clearly describe the action format, we denote each agent i's strategy set S_i as A_i, and each strategy function as $a_i(\cdot) : \Theta_i \to A_i$. The outcome function $g(\cdot)$ is denoted as $g_a(\cdot) : A_1 \times \cdots \times A_I \to X$, where the parameters are action-format strategies. Hence, the mechanism Γ is rewritten as $\Gamma_a = (A_1, \cdots, A_I, g_a(\cdot))$. The game induced by Γ_a is denoted as G_a, which works in a one-stage manner:

Step 1: By using the action-format strategy function $a_i(\cdot) : \Theta_i \to A_i$, each agent i with private type θ_i performs the action-format strategy $a_i(\theta_i)$. [4]

---

[4] For the case of action-format strategies, the designer *observes* the performance

Step 2: The mechanism Γ_a yields the outcome $g_a(a_1(\theta_1), \cdots, a_I(\theta_I))$. [5]
Here, each agent i's utility is denoted as $u_i(g_a(a_1(\theta_1), \cdots, a_I(\theta_I)), \theta_i)$.

Definition 7: Suppose the game G_a has a Bayesian Nash equilibrium $a^*(\cdot) = (a_1^*(\cdot), \cdots, a_I^*(\cdot))$, *i.e.*, for all i and all $\theta_i \in \Theta_i$, $\hat{a}_i \in A_i$,

$$E_{\theta_{-i}}[u_i(g_a(a_i^*(\theta_i), a_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g_a(\hat{a}_i, a_{-i}^*(\theta_{-i})), \theta_i)|\theta_i].$$

Consider this equilibrium, there is a compound mapping $g_a(a^*(\cdot)) : \Theta \to X$ from agents' possible types $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_I) \in \Theta$ into the outcome $g_a(a^*(\hat{\theta}))$, which is equal to $f(\hat{\theta})$ for any $\hat{\theta} \in \Theta$. Based on the compound mapping, we construct a direct mechanism $\bar{\Gamma}_a = (\Theta_1, \cdots, \Theta_I, g_a(a^*(\cdot)))$ with action-format strategies.

Definition 8: According to Myerson [2], the direct mechanism $\bar{\Gamma}_a$ induces a *multistage* direct game $\bar{G}_a$, which works as follows:
Step 1: Each agent i with private type θ_i individually reports an arbitrary type $\hat{\theta}_i \in \Theta_i$. Here each agent does not need to be *"honest"*, *i.e.*, $\hat{\theta}_i$ can be different from θ_i.
Step 2: The direct mechanism $\bar{\Gamma}_a$ returns a suggestion to each agent i, here the suggestion is just a message-format description of action $a_i^*(\hat{\theta}_i) \in A_i$. In order to emphasize the format of the suggestion is a message, we denote the suggestion as $a_i^m(\hat{\theta}_i)$;
Step 3: Each agent i individually performs an action $\hat{a}_i \in A_i$. Here each agent i does not need to be *"obedient"*, *i.e.*, $\hat{a}_i$ can be different from $a_i^m(\hat{\theta}_i)$.
Step 4: After observing all actions $\hat{a}_1, \cdots, \hat{a}_I$ have been performed, the direct mechanism $\bar{\Gamma}_a$ yields the outcome $g_a(\hat{a}_1, \cdots, \hat{a}_I)$.
Here, each agent i's utility is denoted as $u_i(g_a(\hat{a}_1, \cdots, \hat{a}_I), \theta_i)$.

Note 4: Generally speaking, each agent's private type is his privacy and is valuable to him. Consider Step 1 in Definition 8, each agent i reports an arbitrary type either honestly or dishonestly. Note that choosing to be honest or dishonest is each agent's independent and private choice, which cannot be controlled by the designer. *From each agent's perspective, if the utility of truthtelling is not strictly larger than the utility of false-telling, then any reasonable agent will certainly prefer false-telling*, because false-telling always hides his privacy. □

Note 5: Consider Step 3 in Definition 8, each agent performs an action after receiving a suggestion, either obediently or disobediently. Note that choosing to be obedient or disobedient is each agent's independent and open choice,

----

of each agent's action. As a comparison, for the case of message-format strategies, the designer *receives* each agent's message.
[5] If in the mechanism Γ_a some agent i only declares a message about how to perform an action but does not realistically perform it, then this declaration is meaningless.

which cannot be controlled by the designer. Since the designer is not a dictator in the framework of mechanism design theory, she has no power to punish any disobedient agent. Obviously, each agent will choose the choice which yields the higher utility. □

3   Main results

Consider the multistage direct game $\bar{G}_a$ induced by the direct mechanism $\bar{\Gamma}_a$ described in Definition 8. According to Myerson [2], the strategy "*honest and obedient*" is the Bayesian Nash equilibrium of the game $\bar{G}_a$: *i.e.*, each agent i not only honestly discloses his private type in Step 1 (*i.e.*, $\hat{\theta}_i = \theta_i$), but also obeys the suggestion in Step 3 (*i.e.*, $\hat{a}_i = a_i^m(\theta_i)$). However, in this section we will point out that Myerson's conclusion will not hold when each agent's strategy is of an action format.

Proposition 1: For any given social choice function $f(\cdot) : \Theta \rightarrow X$, suppose there is a mechanism that implements it in Bayesian Nash equilibrium, and each agent's strategy is of an action format. Then f will not be truthfully implementable, *i.e.*, in the multistage direct game induced by the corresponding direct mechanism, "*honest and obedient*" will no longer be the Bayesian Nash equilibrium.

Proof: Suppose a mechanism $\Gamma_a = (A_1, \cdots, A_I, g_a(\cdot))$ implements the social choice function $f(\cdot) : \Theta \rightarrow X$ in Bayesian Nash equilibrium, and the format of each agent's strategy is an action. By Definition 6 it induces a one-stage game G_a, the equilibrium of which is denoted as $a^*(\cdot) = (a_1^*(\cdot), \cdots, a_I^*(\cdot))$, and the outcome $g_a(a^*(\theta)) = f(\theta)$ for any $\theta \in \Theta$. By Definition 7 and 8, there is a corresponding direct mechanism $\bar{\Gamma}_a = (\Theta_1, \cdots, \Theta_I, g_a(a^*(\cdot)))$ and a multistage direct game $\bar{G}_a$. In the direct game $\bar{G}_a$, each agent i with private type θ_i independently makes two decisions: in Step 1 he chooses to be "*honest*" or "*dishonest*", and in Step 3 he chooses to be "*obedient*" or "*disobedient*".

For any agent i, if all other agents $1, \cdots, i-1, i+1, \cdots, I$ choose to be "*honest*" (*i.e.*, $\hat{\theta}_{-i} = \theta_{-i}$), in the following discussions we investigate which choice agent i should choose in Step 1 of $\bar{G}_a$, either "*honest*" or "*dishonest*".

Case 1: Suppose agent i chooses to be "*honest*" in Step 1 of $\bar{G}_a$, *i.e.*, $\hat{\theta}_i = \theta_i$. Then in Step 2 of $\bar{G}_a$, the suggestion to agent i will be $a_i^m(\theta_i)$, and the suggestions to the rest agents will be $a_{-i}^m(\theta_{-i})$. Since the Bayesian Nash equilibrium strategy of G_a is $a^*(\cdot) = (a_1^*(\cdot), \cdots, a_I^*(\cdot))$, then in Step 3 of $\bar{G}_a$, choosing to be "*obedient*" (*i.e.*, $\hat{a}_i = a_i^*(\theta_i)$) will be the optimal choice of agent i if all other agents choose to be "*obedient*" (*i.e.*, $\hat{a}_{-i} = a_{-i}^*(\theta_{-i})$). In Step 4 of $\bar{G}_a$, the direct mechanism $\bar{\Gamma}_a$ will yield the outcome $g_a(a^*(\theta))$, which is equal to

$f(\theta)$ for all $\theta \in \Theta$.

Case 2: Suppose in Step 1 of $\bar{G}_a$, agent i wants to hide his privacy and chooses to be "*dishonest*", *i.e.*, $\hat{\theta}_i \neq \theta_i$. Obviously, in Step 2 of $\bar{G}_a$, the suggestion to agent i will be $a_i^m(\hat{\theta}_i) \neq a_i^m(\theta_i)$, and the suggestions to the others will still be $a_{-i}^m(\theta_{-i})$. Therefore, if all other agents choose "*obedient*" (*i.e.*, performing $\hat{a}_{-i} = a_{-i}^*(\theta_{-i})$), then the optimal choice of agent i in Step 3 of $\bar{G}_a$ should be "*disobedient*" (*i.e.*, not obeying the suggestion $a_i^m(\hat{\theta}_i)$ but still performing $\hat{a}_i = a_i^*(\theta_i)$). Although the designer can find the agent i is disobedient in Step 4 of $\bar{G}_a$, she cannot punish him according to Note 5. Consequently, in Step 4 of $\bar{G}_a$, the direct mechanism $\bar{\Gamma}_a$ will still yield the outcome $g_a(a^*(\theta))$, which is equal to $f(\theta)$ for all $\theta \in \Theta$.

It should be emphasized that although in Case 2 the agent i chooses to be "*disobedient*" in Step 3 of $\bar{G}_a$, what he really performs is still consistent with his private type, *i.e.*, $\hat{a}_i = a_i^*(\theta_i)$, the same as what agent i performs in Case 1. As a result, the social choice function $f(\cdot)$ is still implemented in Bayesian Nash equilibrium by the constructed direct mechanism $\bar{\Gamma}_a = (\Theta_1, \cdots, \Theta_I, g_a(a^*(\cdot)))$.

From agent i's perspective, his privacy is disclosed in Case 1 and hidden in Case 2. Thus, Case 2 will be strictly better than Case 1, because agent i can obtain the same outcome $g_a(a^*(\theta))$ in the two cases but protects his privacy in Case 2. Consequently, "*honest and obedient*" will no longer be the Bayesian Nash equilibrium of the direct mechanism. Furthermore, since Case 2 holds for any agent i, it can be generalized to every agent as follows.

Case 3: Suppose each agent i with private type θ_i chooses to be "*dishonest*" in Step 1 of $\bar{G}_a$ (*i.e.* reporting a false type $\hat{\theta}_i \neq \theta_i$), and then chooses to be "*disobedient*" in Step 3 of $\bar{G}_a$, *i.e.*, not obeying the designer's suggestion $a_i^m(\hat{\theta}_i)$ but still performing the action $\hat{a}_i = a_i^*(\theta_i)$ which is consistent with his private type. Since the Bayesian Nash equilibrium strategy of G_a is $a^*(\cdot) = (a_1^*(\cdot), \cdots, a_I^*(\cdot))$, then the final outcome of $\bar{G}_a$ will still be $g_a(a^*(\theta))$, which is equal to $f(\theta)$ for all $\theta \in \Theta$. Put differently, $f(\cdot)$ is implemented by the constructed direct mechanism $\bar{\Gamma}_a = (\Theta_1, \cdots, \Theta_I, g_a(a^*(\cdot)))$ in Bayesian Nash equilibrium, *not truthfully, but dishonestly and disobediently.*

Compared with Case 1, in Case 3 each agent obtains the same outcome $g_a(a^*(\theta))$ and at the same time protects his privacy. Therefore, "*dishonest and disobedient*" will be strictly better than "*honest and obedient*" from each agent's perspective.

To sum up, when the format of each agent's strategy is an action, "*dishonest and disobedient*" instead of "*honest and obedient*" will be the Bayesian Nash equilibrium of the direct mechanism $\bar{\Gamma}_a = (\Theta_1, \cdots, \Theta_I, g_a(a^*(\cdot)))$, which means the revelation principle does not hold. □

4   Conclusions

In this paper, we propose that in any game there are two formats of strategy (*i.e.*, an informational message or a realistic action), and the action format is very common in many practical cases. In Section 2.1 we point out that the revelation principle holds when the format of each agent's strategy is a message. However, when the format of each agent's strategy is an action, "*honest and obedient*" will no longer be the Bayesian Nash equilibrium of the direct mechanism, actually the social choice function can only be implemented "*dishonestly and disobediently*" in Bayesian Nash equilibrium by the direct mechanism. Therefore, the revelation principle fails when each agent's strategy is of action format.

Appendix: Notations and Definitions [1]

Let us consider a setting with one designer and I agents indexed by $i = 1, \cdots, I$. Each agent i privately observes his *type* θ_i that determines his preference over elements in an outcome set X. The set of possible types for agent i is denoted as Θ_i. The vector of agents' types $\theta = (\theta_1, \cdots, \theta_I)$ is drawn from set $\Theta = \Theta_1 \times \cdots \times \Theta_I$ according to probability density $\phi(\cdot)$, and each agent i's *utility function* over the outcome $x \in X$ given his type θ_i is $u_i(x, \theta_i) \in \mathbb{R}$.

Definition 23.B.1 A *social choice function* (SCF) is a function $f : \Theta_1 \times \cdots \times \Theta_I \to X$ that, for each possible profile of the agents' types $\theta_1, \cdots, \theta_I$, assigns a collective choice $f(\theta_1, \cdots, \theta_I) \in X$.

Definition 23.B.3 A *mechanism* $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ is a collection of I strategy sets $S_1, \cdots, S_I$ and an outcome function $g : S_1 \times \cdots \times S_I \to X$.

The mechanism combined with possible types $(\Theta_1, \cdots, \Theta_I)$, the probability density $\phi(\cdot)$ over the possible realizations of $\theta \in \Theta_1 \times \cdots \times \Theta_I$, and utility functions $(u_1, \cdots, u_I)$ defines a Bayesian game of incomplete information. The strategy function of each agent i in the game induced by Γ is a private function $s_i(\theta_i) : \Theta_i \to S_i$. Each strategy set S_i contains agent i's possible strategies. The outcome function $g(\cdot)$ describes the rule for how agents' strategies get turned into a social choice.

Definition 23.B.5 A *direct mechanism* is a mechanism $\bar{\Gamma} = (\bar{S}_1, \cdots, \bar{S}_I, \bar{g}(\cdot))$ in which $\bar{S}_i = \Theta_i$ for all i and $\bar{g}(\theta) = f(\theta)$ for all $\theta \in \Theta$. [6]

---

[6] The bar symbol is used to distinguish the direct mechanism from the indirect mechanism.

Definition 23.D.1 A strategy profile $s^*(\cdot) = (s_1^*(\cdot), \cdots, s_I^*(\cdot))$ is a *Bayesian Nash equilibrium* of mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ if, for all i and all $\theta_i \in \Theta_i$, $\hat{s}_i \in S_i$, there exists

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(\hat{s}_i, s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i].$$

Definition 23.D.2 The mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ *implements the social choice function* $f(\cdot)$ *in Bayesian Nash equilibrium* if there is a Bayesian Nash equilibrium of Γ, $s^*(\cdot) = (s_1^*(\cdot), \cdots, s_I^*(\cdot))$, such that $g(s^*(\theta)) = f(\theta)$ for all $\theta \in \Theta$.

Definition 23.D.3 The social choice function $f(\cdot)$ is *truthfully implementable in Bayesian Nash equilibrium* (or *Bayesian incentive compatible*) if $\bar{s}_i^*(\theta_i) = \theta_i$ (for all $\theta_i \in \Theta_i$ and $i = 1, \cdots, I$) is a Bayesian Nash equilibrium of the direct mechanism $\bar{\Gamma} = (\Theta_1, \cdots, \Theta_I, f(\cdot))$. That is, if for all $i = 1, \cdots, I$ and all $\theta_i \in \Theta_i$, $\hat{\theta}_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i].$$

Proposition 23.D.1: (*The Revelation Principle for Bayesian Nash Equilibrium*) Suppose that there exists a mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ that implements the social choice function $f(\cdot)$ in Bayesian Nash equilibrium. Then $f(\cdot)$ is truthfully implementable in Bayesian Nash equilibrium.

References

[1] A. Mas-Colell, M.D. Whinston and J.R. Green, Microeconomic Theory. Oxford University Press, 1995.

[2] R. Myerson, Optimal coordination mechanisms in generalized principal-agent problems. Journal of Mathematical Economics, vol.10, 67-81, 1982.

[3] Y. Narahari et al, Game Theoretic Problems in Network Economics and Mechanism Design Solutions. Springer, 2009.

[4] H. Bester and R. Strausz, Contracting with imperfect commitment and the revelation principle: The single agent case. Econometrica, Vol.69, No.4, 1077-1098, 2001.

[5] Epstein and Peters, A revelation principle for competing mechanisms. Journal of Economic Theory, 88, 119-160 (1999).

[6] A. Kephart and V. Conitzer, The revelation principle for mechanism design with reporting cost. In Proceedings of the ACM Conference on Electronic Commerce (EC), Maastricht, The Netherlands, 2016.