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

Wu, Haoyang

Wan-Dou-Miao Research Lab

15 November 2020

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

Haoyang Wu *

Wan-Dou-Miao Research Lab, Room 301, Building 3, 718 WuYi Road, 200051, 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. At first we point out that in any game the format of each agent's strategy is either an abstract message or a real action. For any given social choice function, if the mechanism which implements it in Bayesian Nash equilibrium has action-format strategies, then "honest and obedient" will not be an equilibrium strategy in the corresponding direct mechanism. Consequently, the revelation principle fails.

Key words: Mechanism design; Revelation principle.

1 Introduction

In the framework of mechanism design theory [1–4], 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 designer constructs a mechanism which specifies each agent's feasible strategy

* Corresponding author.
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".

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 [5] pointed out that the revelation principle may fail because of imperfect commitment;Epstein and Peters [6] proposed that the revelation principle fails in situations where several mechanism designers compete against each other. Kephart and Conitzer [7] proposed that when reporting truthfully is costless and misreporting is costly, the revelation principle can fail to hold.

Different from these criticisms on the revelation principle, this paper aims to propose another failure of the revelation principle. We point out that each agent's strategy is of two formats, *i.e.*, a message or an action. Indeed, the format of each agent's strategy plays an important role to the correctness of revelation principle. The paper is organized as follows: Section 2 discusses the motivation of this paper, Section 3 proposes the main result. In Section 4, we propose the bug in the proof of revelation principle given by Mas-Colell *et al* [1]. Section 5 draws conclusions. Notations and proof about mechanism design theory and the revelation principle are given in Appendix 1 and Appendix 2, which are cited from MWG's book [1].

2 Motivation

Note 1: In any game, the format of each agent's strategy is either an abstract message or a real action. [2]

It should be emphasized that only in some restricted cases (*e.g.*, chess, war simulation game and so on) can each agent's strategy be described as pure information and represented by an abstract message. On the other hand, in many practical cases each agent's strategy cannot be described as pure information but must be described as a real action. For example, a war simulation game only contains military plans of players, but a real war contains military actions of armies.

[2] Although Note 1 looks naive, it is not trivial. The reason why we emphasize the two formats of strategy is that the revelation principle will not hold for the case of action-format strategies. We will deeply discuss it in Section 3. For simplification, in the following discussions we simply assume that in any game each agent's strategy is 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.

An interesting example is the auction. At first sight each bidder's bid is pure information and looks like a message-format strategy, however in many practical cases only the bid information itself is not enough to be a full strategy: besides announcing the message-format bid, the winner must perform a real action (*e.g.*, paying money to the auctioneer) in order to really finish the auction. Hence, in many practical cases, the auction is an action-format game.

Next we will deeply investigate the distinction of two formats of strategy.

2.1 Case 1: Mechanism with message-format strategies

Definition 1: A message-format strategy of an agent in a mechanism is a strategy represented by an abstract message, which only contains pure information and does not need to be performed realistically. For example, let us consider a chess game, then each player's strategy is his message-format plan about how to play chess.

Definition 2: Given a social choice function f, suppose a mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ implements it in equilibrium with message-format strategies. To clearly describe the case of message-format strategies, we denote each strategy set S_i as M_i, and each agent i's strategy function is denoted as $m_i(\cdot) : \Theta_i \to M_i$, in which Θ_i is agent i's type set. The outcome function $g(\cdot)$ is denoted as $g_m(\cdot) : M_1 \times \cdots \times M_I \to X$. Hence, the mechanism Γ is denoted 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 strategy functions $(m_1(\cdot), \cdots, m_I(\cdot))$, agents $1, \cdots, I$ with private types $(\theta_1, \cdots, \theta_I)$ send messages $(m_1(\theta_1), \cdots, m_I(\theta_I))$ to the designer. Here, there is no disclosure of each agent i's true type θ_i. [3]
Step 2: The mechanism Γ_m yields the outcome $g_m(m_1(\theta_1), \cdots, m_I(\theta_I))$.
Here, the input parameters of the outcome function $g_m(\cdot)$ are message-format strategies $m_1(\theta_1), \cdots, m_I(\theta_I)$.

Definition 3: Suppose the game G_m has an equilibrium $(m_1^*(\cdot), \cdots, m_I^*(\cdot))$. Consider this equilibrium, there is a compound mapping from agents' possible types $(\hat{\theta}_1, \cdots, \hat{\theta}_I) \in \Theta$ into the outcome $g_m(m_1^*(\hat{\theta}_1), \cdots, m_I^*(\hat{\theta}_I))$, which is equal to $f(\hat{\theta}_1, \cdots, \hat{\theta}_I)$. Based on the compound mapping, we define a corresponding direct mechanism $\bar{\Gamma}_m = (\Theta_1, \cdots, \Theta_I, g_m(m_1^*(\cdot), \cdots, m_I^*(\cdot)))$.

Note 2: The outcome function of the constructed direct mechanism $\bar{\Gamma}_m$ must be the compound mapping $g_m(m_1^*(\cdot), \cdots, m_I^*(\cdot))$ instead of the social choice function $f(\cdot)$, although the two functions are outcome-equivalent. The reason

[3] To clearly describe the true type of each agent i, we denote it as θ_i, and any possible type of agent i is denoted as $\hat{\theta}_i \in \Theta_i$.

is straightforward: if the outcome function of $\bar{\Gamma}_m$ is simply written as $f(\cdot)$, then this naive direct mechanism will be irrelevant to the mechanism $\Gamma_m = (M_1, \cdots, M_I, g_m(\cdot))$, and hence cannot implement $f(\cdot)$ at all.

Definition 4: The direct mechanism $\bar{\Gamma}_m$ induces a *one-stage* direct game $\bar{G}_m$ as follows:
Step 1: Each agent i with private type θ_i individually reports a message-format type $\hat{\theta}_i \in \Theta_i$, here $\hat{\theta}_i$ does not need to be equal to agent i's true type θ_i.
Step 2: By using the equilibrium strategy functions $m_1^*(\cdot), \cdots, m_I^*(\cdot)$, the direct mechanism $\bar{\Gamma}_m$ calculates $m_1^*(\hat{\theta}_1), \cdots, m_I^*(\hat{\theta}_I)$, and then yields the outcome $g_m(m_1^*(\hat{\theta}_1), \cdots, m_I^*(\hat{\theta}_I))$.

Note 3: It should be emphasized that the calculated results $m_1^*(\hat{\theta}_1), \cdots, m_I^*(\hat{\theta}_I)$ are pure information. Only when each agent i's strategy set S_i is message-format can the calculated results be legal message-format parameters of the outcome function $g_m(\cdot)$.

Note 4: By Definition 3, $(m_1^*(\cdot), \cdots, m_I^*(\cdot))$ is the equilibrium of the game G_m, in which each agent i with private type θ_i will choose $m_i^*(\theta_i)$. Therefore, by Definition 4 in the direct game $\bar{G}_m$ each agent i will find truth-telling $\hat{\theta}_i = \theta_i$ to be the optimal strategy given that all other agents tell the truth $\hat{\theta}_{-i} = \theta_{-i}$. Thus, for the case of message-format strategies, truth-telling is a Bayesian Nash equilibrium of the game $\bar{G}_m$. This conclusion means that *the revelation principle holds when each agent's strategy is of message format.*

2.2 Case 2: Mechanism with action-format strategies

Definition 5: An action-format strategy of an agent in a mechanism is a strategy represented by a realistic action, which should be performed by himself practically. For example, let us consider a tennis game, then each player's strategy is his realistic action of playing tennis, but not any informational plan of how to play tennis.

Definition 6: Given a social choice function f, suppose a mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ implements it in equilibrium with action-format strategies. To clearly describe the case of action-format strategies, we denote each strategy set S_i as A_i, and each agent i's strategy function is denoted as $a_i(\cdot) : \Theta_i \to A_i$. The outcome function $g(\cdot)$ is denoted as $g_a(\cdot) : A_1 \times \cdots \times A_I \to X$. Hence, the mechanism Γ is denoted 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 strategy function $a_i(\cdot)$, each agent i $(i = 1, \cdots, I)$ with private type θ_i performs the action-format strategy $a_i(\theta_i)$, which is observed by the designer. Here, each agent i's private type θ_i is not disclosed.
Step 2: The mechanism Γ_a yields the outcome $g_a(a_1(\theta_1), \cdots, a_I(\theta_I))$.

Here, the input parameters of the outcome function $g_a(\cdot)$ are action-format strategies $a_1(\theta_1), \cdots, a_I(\theta_I)$. Obviously, if some agent i with private type θ_i only declares a plan of action $a_i(\theta_i)$ but does not realistically perform the action, then G_a will not work at all.

Definition 7: Suppose the game G_a has an equilibrium $(a_1^*(\cdot), \cdots, a_I^*(\cdot))$. Consider this equilibrium, there is a compound mapping from agents' possible types $(\hat{\theta}_1, \cdots, \hat{\theta}_I)$ into the outcome $g_a(a_1^*(\hat{\theta}_1), \cdots, a_I^*(\hat{\theta}_I))$, which is equal to $f(\hat{\theta}_1, \cdots, \hat{\theta}_I)$. Based on the compound mapping, we define a corresponding direct mechanism $\bar{\Gamma}_a = (\Theta_1, \cdots, \Theta_I, g_a(a_1^*(\cdot), \cdots, a_I^*(\cdot)))$. [4]

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

Note 5: Different from Step 2 in Definition 4, here the action-format strategies $(\hat{a}_1, \cdots, \hat{a}_I)$ cannot be calculated by the designer but must be performed by each agent realistically.

Note 6: Consider Step 1 in Definition 8, each agent is required to report a type, either honestly or dishonestly. Note that choosing to be honest or dishonest is each agent's private choice and cannot be directly observed by the designer. Since no agent will be punished even if being found false-telling, of course no agent is willing to disclose his privacy unless truth-telling is his strictly optimal choice. Put differently, *if the outcome of truth-telling is equivalent to the outcome of false-telling, then each agent will certainly prefer false-telling*, since false-telling always protects his privacy.

Note 7: Consider Step 3 in Definition 8, each agent performs an action-format strategy, either obediently or disobediently. Note that choosing to be obedient or disobedient is each agent's open choice and can be directly observed by the designer. However, the designer can neither control any agent's decision nor punish any disobedient agent.

[4] Similar to Note 2, here the outcome function of the constructed direct mechanism $\bar{\Gamma}_a$ should be the compound mapping $g_a(a_1^*(\cdot), \cdots, a_I^*(\cdot))$ instead of $f(\cdot)$.

3 Main results

Consider the multistage direct game $\bar{G}_a$ induced by $\bar{\Gamma}_a$ given in Definition 8. Myerson [2] claims that the strategy "honest and obedient" is the Bayesian Nash equilibrium of the direct 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 designer's 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 does not hold when each agent's strategy is of an action format.

Proposition 1: For a given social choice function f, suppose there is a mechanism $\Gamma_a = (A_1, \cdots, A_I, g_a(\cdot))$ that implements it in Bayesian Nash equilibrium, in which each agent's strategy is of an action format. Then f will not be truthfully implementable, *i.e.*, in the multistage direct game $\bar{G}_a$ induced by the corresponding direct mechanism $\bar{\Gamma}_a$, "honest and obedient" is no longer the equilibrium strategy.

Proof: Suppose a mechanism $\Gamma_a = (A_1, \cdots, A_I, g_a(\cdot))$ implements the social choice function f in Bayesian Nash equilibrium, then by Definition 6 it induces a one-stage game G_a. Let the equilibrium strategy of G_a be denoted as $(a_1^*(\theta_1), \cdots, a_I^*(\theta_I))$. According to Definition 8, there are two possible cases for agents $1, \cdots, I$ in Step 1 of the multistage game $\bar{G}_a$:

Case 1: Each agent is honest and obedient
Suppose each agent i chooses to be "*honest*" in Step 1 of the direct game $\bar{G}_a$, *i.e.*, $\hat{\theta}_i = \theta_i$, then in Step 2 of $\bar{G}_a$ the designer's suggestion will be $a_i^m(\theta_i)$. Since the equilibrium strategy of each agent i in G_a is a_i^*, the optimal choice of each agent i in Step 3 of $\bar{G}_a$ is to be "*obedient*", *i.e.*, obey the suggestion $a_i^m(\theta_i)$ and perform the action-format strategy $a_i^*(\theta_i)$. In Step 4, the final outcome will be $g_a(a_1^*(\theta_1), \cdots, a_I^*(\theta_I))$. Note that each agent i's private type is disclosed in this case.

Case 2: At lease one agent is dishonest and disobedient
Suppose at least one agent i chooses to be "*dishonest*" in Step 1 of the direct game $\bar{G}_a$, *i.e.*, $\hat{\theta}_i \neq \theta_i$, then in Step 2 of $\bar{G}_a$ the designer's suggestion to agent i will be $a_i^m(\hat{\theta}_i) \neq a_i^m(\theta_i)$. Since the equilibrium of G_a is $(a_1^*(\cdot), \cdots, a_I^*(\cdot))$, the optimal choice of agent i in Step 3 of $\bar{G}_a$ is to be "*disobedient*" (*i.e.*, *not* to obey the suggestion $a_i^m(\hat{\theta}_i)$ but to perform the action-format strategy $a_i^*(\theta_i)$, which is consistent with his true type θ_i). Note that agent i's private type is *not* disclosed in this case. Furthermore, Case 2 can be generalized to each agent i as follows.

Case 3: Each agent is dishonest and disobedient
Suppose each agent i chooses to be "dishonest" in Step 1 of the direct game $\bar{G}_a$ (*i.e.* reporting a false type $\hat{\theta}_i \neq \theta_i$), and chooses to be "disobedient" in Step 3

of $\bar{G}_a$ (*i.e.*, performing the action-format strategy $a_i^*(\theta_i)$, which is inconsistent with the designer's suggestion $a_i^m(\hat{\theta}_i)$ but consistent with each agent i's private type). By Note 7, although the designer can find each agent i is disobedient in Step 4 of $\bar{G}_a$, she has to yield the outcome $g_a(a_1^*(\theta_1), \cdots, a_I^*(\theta_I))$.

To sum up, "dishonest and disobedient" is outcome-equivalent to "honest and obedient" from each agent's perspective. According to Note 6, each agent i will certainly prefer "dishonest and disobedient". Therefore, "dishonest and disobedient" is the Bayesian Nash equilibrium of the direct game $\bar{G}_a$, and the final outcome is $g_a(a_1^*(\theta_1), \cdots, a_I^*(\theta_I))$, which is equal to $f(theta_1), \cdots, \theta_I)$. Consequently, f cannot be truthfully implemented in Bayesian Nash equilibrium, and hence the revelation principle does not hold when each agent's strategy is of an action format. $\square$

4 Bug in MWG's proof of the revelation principle

Here we cite formula (23.D.3) from Appendix 2 as follow. For all i and all $\theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(s_i^*(\hat{\theta}_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i], \quad (23.D.3)$$

for all $\hat{\theta}_i \in \Theta_i$.

When each agent's strategy is of an action format, formula (23.D.3) should be rewritten as follows.

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

It can be seen that the item of outcome in each agent i's utility u_i is yielded by the compound function $g_a(a_1^*(\cdot), \cdots, a_I^*(\cdot)) : \Theta_1 \times \cdots \times \Theta_I \to X$. The processing sequence of the compound function is that at first each agent i performs his action-format strategy $a_i^*(\cdot)$, and then the designer performs the outcome function $g_a(\cdot)$. Note that during this process each agent's true type θ_i is not disclosed.

Note that $(a_1^*(\cdot), \cdots, a_I^*(\cdot))$ is the equilibrium strategy of the game G_a induced by the mechanism Γ_a, thus $g_a(a_1^*(\hat{\theta}_1), \cdots, a_I^*(\hat{\theta}_I)) = f(\hat{\theta}_1, \cdots, \hat{\theta}_I)$ for all possible $\hat{\theta}_i \in \Theta_i$. However, by Proposition 1, when each agent's strategy is of an action format, f is not truthfully implementable: *i.e.*, it is impossible to construct a direct mechanism $\bar{\Gamma} = (\Theta_1, \cdots, \Theta_I, g_a(a_1^*(\cdot), \cdots, a_I^*(\cdot)))$ to implement $f(\cdot)$ in Bayesian Nash equilibrium. Consequently, formula (23.D.4) cannot be yielded from formula (23.D.3). This is just the bug in MWG's proof.

5 Conclusions

In this paper, we propose that in any game there are two formats of strategy (*i.e.*, an abstract message or a real action). In Section 2.1 we point out that the revelation principle holds when each agent's strategy is of message format. However, when each agent's strategy is of an action format, in the multistage direct game induced by the direct mechanism, "dishonest and disobedient" is the Bayesian Nash equilibrium. Therefore, the revelation principle fails when each agent's strategy is of action format.

Appendix 1: Notations and Definitions

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 $u_i(x, \theta_i) \in \mathbb{R}$.

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.

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 $f(\theta_1, \cdots, \theta_I) \in X$.

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

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

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 $\theta \in \Theta$.

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 $\theta \in \Theta$. [5] 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 $\bar{\Gamma} = (\bar{S}_1, \cdots, \bar{S}_I, \bar{g}(\cdot))$, in which $\bar{S}_i = \Theta_i$, $\bar{g} = f$. That is, if for all $i = 1, \cdots, I$ and all $\theta_i \in \Theta_i$, $\hat{\theta}_i \in \Theta_i$, there exists

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i]. \tag{1}$$

Appendix 2: Proof of the revelation principle

Proposition 23.D.1 [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.

Proof: If $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ implements $f(\cdot)$ in Bayesian Nash equilibrium, then there exists a profile of strategies $s^*(\cdot) = (s_1^*(\cdot), \cdots, s_I^*(\cdot))$ such that $g(s^*(\theta)) = f(\theta)$ for all θ, and for all i and all $\theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(\hat{s}_i, s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i], \tag{23.D.2}$$

for all $\hat{s}_i \in S_i$. Condition (23.D.2) implies, in particular, that for all i and all $\theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(s_i^*(\hat{\theta}_i), s_{-i}^*(\theta_{-i})), \theta_i)|\theta_i], \tag{23.D.3}$$

for all $\hat{\theta}_i \in \Theta_i$. Since $g(s^*(\theta)) = f(\theta)$ for all θ, (23.D.3) means that, for all i and all $\theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i], \tag{23.D.4}$$

for all $\hat{\theta}_i \in \Theta_i$. But, this is precisely the condition for $f(\cdot)$ to be truthfully implementable in Bayesian Nash equilibrium. □

References

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

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

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

[3] R. Serrano, The theory of implementation of social choice rules. SIAM Review 46 (2004) 377-414.

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

[5] 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.

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

[7] 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.