A note on the definition of Bayesian Nash equilibrium of a mechanism when strategies of agents are costly actions

Wu, Haoyang

8 April 2019

A note on the definition of Bayesian Nash equilibrium of a mechanism when strategies of agents are costly actions

Haoyang Wu *

Wan-Dou-Miao Research Lab, Room 301, Building 3, 718 WuYi Road, 200051, Shanghai, China.

Abstract

In the framework of mechanism design theory, in order to implement a desired social choice function in Bayesian Nash equilibrium, a designer constructs a mechanism which specifies her favorite outcome for each possible profile of agents' types. Generally speaking, each agent's strategy has two possible formats: an action, or a message. In this paper, we focus on the former case and claim that the definitions of Bayesian Nash equilibrium of a mechanism and Bayesian incentive compatibility should all be based on a profit function instead of a utility function. Next, we derive the main result: Given a social choice function which can be implemented by an indirect mechanism in Bayesian Nash equilibrium, if all strategies of agents are costly actions, then it cannot be inferred that there exists a direct mechanism that can truthfully implement the social choice function in Bayesian Nash equilibrium.

Key words: Bayesian Nash Equilibrium; Mechanism design; Revelation Principle.

1 Introduction

In the framework of mechanism design theory [1–3], there are one designer and some agents. [1] The designer would like to implement a desired social choice function which specifies her favorite outcome for each possible profile of agents' types. However, each agent's type is modelled as his private property and unknown to the designer. In order to implement a social choice function in Bayesian Nash equilibrium, the designer constructs a mechanism which

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

[1] The designer is denoted as "She", and the agent is denoted as "He".

specifies each agent's 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).

Generally speaking, each agent's strategy has two possible formats: an action, or a message (*i.e.*, a plan of action) (Page 883, Line 8, [1]). The distinction between the two formats is that: the former format of strategy is *a real action* which is performed realistically by each agent and hence naturally needs some action cost, whereas the latter format of strategy is *a message of action plan* which is reported to the designer and hence doesn't need any cost to be performed realistically. In this paper, we focus on the former format of strategy, and investigate what would happen to the notion of Bayesian Nash equilibrium of a mechanism.

The paper is organized as follows. First, we introduce a notion of profit function (*i.e.*, Definition 1), and then claim that the notions of Bayesian Nash equilibrium of a mechanism and Bayesian incentive compatibility should all be based on the profit function instead of the utility function when strategies of agents are costly actions (*i.e.*, Definition 2 and Definition 3). Next, we derive the main result (*i.e.*, Proposition 1): Given a social choice function which can be implemented by an indirect mechanism in Bayesian Nash equilibrium, if all strategies of agents are costly actions, then it cannot be inferred that there exists a direct mechanism that can truthfully implement f in Bayesian Nash equilibrium, which contradicts the revelation principle. Afterwards, we consider a possible argument and give a reply. Section 3 concludes the paper.

2 Theoretical Analysis

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 $\Theta = (\Theta_1, \cdots, \Theta_I)$ 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 $u_i(x, \theta_i)$.

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 $s_i(\cdot) : \Theta_i \to S_i$. Each strategy set S_i contains agent i's possible strategies (*i.e.*, *actions*, or *plans of action*). 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 $f(\theta_1, \cdots, \theta_I) \in X$.

Note 1: As given above, for each agent i with type θ_i, there are two possible formats of his strategy $s_i(\theta_i)$: an action, or a message.
Case 1: If the format of strategy $s_i(\theta_i)$ is an action, then $s_i(\theta_i)$ should be performed by agent i realistically. Hence, it is reasonable to say that in order to perform the action $s_i(\theta_i)$, agent i with type θ_i needs to spend some monetary cost (or make some effort which can be quantified as some monetary cost).
Case 2: If the format of strategy $s_i(\theta_i)$ is a message, then $s_i(\theta_i)$ is just a plan of action but not a realistic action. $\square$

In the following discussions, we will focus on the former case and investigate what would happen in this case to the notion of Bayesian Nash equilibrium in the game induced by a mechanism. To simplify representations, we assume that each agent's action cost is only relevant to his action and private type, and is independent of the game outcome.

Definition 1: For a given social choice function f, consider a mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ that implements it in Bayesian Nash equilibrium. Suppose each agent i's strategy in the game induced by Γ is a costly action $s_i(\theta_i) : \Theta_i \to S_i$, then the corresponding action cost is defined by a *cost function* $c_i(s_i, \theta_i) : S_i \times \Theta_i \to \mathcal{R}^+$, *i.e.*, $c_i(s_i, \theta_i) \geq 0$ for any $s_i \in S_i$, $\theta_i \in \Theta_i$. Let the outcome yielded by the mechanism be $x \in X$ and agent i's utility function be $u_i(x, \theta_i) : X \times \Theta_i \to \mathcal{R}$, then each agent i's profit is defined by a *profit function* $p_i(x, s_i, \theta_i) : X \times S_i \times \Theta_i \to \mathcal{R}$,

$$p_i(x, s_i, \theta_i) = u_i(x, \theta_i) - c_i(s_i, \theta_i). \tag{1}$$

Discussion 1: Someone may argue that when each agent i performs the strategy action s_i with cost $c_i(s_i, \theta_i) \geq 0$, his utility $u_i(x, \theta_i)$ already includes the action cost. Thus, it is not necessary to introduce another notion of profit function $p_i(x, s_i, \theta_i)$ to make confusion.

Answer 1: It should be noted that the utility function $u_i(x, \theta_i)$ contains only two parameters, *i.e.*, the outcome x and the private type θ_i. There is no parameter in $u_i(x, \theta_i)$ to represent the item of action cost. Put differently, the utility function $u_i(x, \theta_i)$ only corresponds to agent i's utility after obtaining the outcome x, but cannot describes his cost spent in performing the action $s_i(\theta_i)$. Consequently, the profit function $p_i(x, s_i, \theta_i)$ should be introduced to describe how much each agent i benefits from the game induced by a mechanism when his strategy is a costly action. $\square$

According to MWG book [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 $\theta_i \in \Theta_i$,

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

for all $\hat{s}_i \in S_i$.

Note 2: As shown above, the conventional notion of Bayesian Nash equilibrium is based on the utility function $u_i(x, \theta_i)$. Suppose that in an indirect mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$, the format of each agent i's strategy $s_i(\theta_i)$ is an action that requires some cost to be performed, *i.e.*, $c_i(s_i, \theta_i) > 0$. As pointed out in Answer 1, the utility function $u_i(x, \theta_i)$ only describes agent i's utility with respect to the outcome x but misses his action cost, hence cannot describe agent i's profit. *Since it is the profit that each rational agent really concerns in a game, the profit function should be introduced to define the Bayesian Nash equilibrium of a mechanism.* [2]

Definition 2: The 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 $\theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[p_i(g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), s_i^*(\theta_i), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[p_i(g(\hat{s}_i, s_{-i}^*(\theta_{-i})), \hat{s}_i, \theta_i)|\theta_i] \qquad (3)$$

i.e.,

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

for all $\hat{s}_i \in S_i$, in which p_i is the profit of agent i given by Eq (1).

According to MWG book [1], 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 $\theta \in \Theta_1 \times \cdots \times \Theta_I$. [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 revelation mechanism

[2] In most of practical cases, the format of each agent's strategy is a costly action. Only in very limited cases (*e.g.*, strategies of agents can be considered as oral announcements) can strategies be viewed costless, and hence by Eq (1) the utility function is equal to the profit function. Therefore, the conventional definition of Bayesian Nash equilibrium based on the utility function holds only in these limited cases.

[3] Here we use a bar symbol to distinguish a direct mechanism from an indirect 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 $\theta_i \in \Theta_i$, $\hat{\theta}_i \in \Theta_i$,

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

Note 3: In the direct mechanism $\bar{\Gamma} = (\bar{S}_1, \cdots, \bar{S}_I,\ \bar{g}(\cdot))$, the format of each agent i's strategy is a message, *i.e.*, each agent i independently determines the report strategy $\bar{s}_i(\cdot) : \Theta_i \to \Theta_i$, and his report type $\bar{s}_i(\theta_i)$ does not need to be his true type θ_i. Let each agent i's *reporting cost* be denoted as $\bar{c}_i(\bar{s}_i, \theta_i) : \Theta_i \times \Theta_i \to \mathcal{R}^+$, *i.e.*, $\bar{c}_i(\bar{s}_i, \theta_i) \geq 0$.[4] Let the outcome yielded by $\bar{\Gamma}$ be $x \in X$, then each agent i's profit in the direct mechanism can be denoted by:

$$\bar{p}_i(x, \bar{s}_i, \theta_i) = u_i(x, \theta_i) - \bar{c}_i(\bar{s}_i, \theta_i). \tag{5}$$

In the literature, it is usually assumed that each agent i can report his true type costlessly, *i.e.*, the truthful report cost $\bar{c}_i(\theta_i, \theta_i) = 0$. However, each agent i's misreporting cost may not be zero, *i.e.*, $\bar{c}_i(\hat{\theta}_i, \theta_i) \geq 0$, $\hat{\theta}_i \neq \theta_i$, $\hat{\theta}_i \in \Theta_i$.[5] Therefore, following Definition 2, the definition of Bayesian incentive compatibility should also be revised as follows.

Definition 3: Suppose in the direct mechanism $\bar{\Gamma} = (\bar{S}_1, \cdots, \bar{S}_I,\ \bar{g}(\cdot))$, each agent i's truthful report cost $\bar{c}_i(\theta_i, \theta_i) = 0$ and misreporting cost $\bar{c}_i(\hat{\theta}_i, \theta_i) \geq 0$, $\hat{\theta}_i \neq \theta_i$, then the social choice function f is *truthfully implementable in Bayesian Nash equilibrium* (or *Bayesian incentive compatible*) if for all $i = 1, \cdots, I$ and all $\theta_i \in \Theta_i$, $\hat{\theta}_i \in \Theta_i$,

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

i.e.,

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[(u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i) - \bar{c}_i(\hat{\theta}_i, \theta_i))|\theta_i]. \tag{6}$$

Note 4: In a direct mechanism, the only thing that the designer knows about each agent i is his report type $\bar{s}_i \in \Theta_i$. After the designer receives $\bar{s}_1, \cdots, \bar{s}_I$ from agents, she has no way to verify whether these reports are truthful or not. All that the designer can do is just to announce $f(\bar{s}_1, \cdots, \bar{s}_I)$ as the outcome. Thus, *in a direct mechanism* $\bar{\Gamma} = (\bar{S}_1, \cdots, \bar{S}_I,\ \bar{g}(\cdot))$, *each agent i with type*

[4] From the perspective of each agent i, reporting a type $\bar{s}_i$ in the direct mechanism is simply to announce a message, and is distinct from the strategy s_i chosen by him in the indirect mechanism. Thus, the reporting cost $\bar{c}_i(\bar{s}_i, \theta_i)$ occurred in a direct mechanism is different from the strategy cost $c_i(s_i, \theta_i)$ occurred in an indirect mechanism.

[5] For example, some researchers investigated misreporting cost [5,6], which are possibly spent by agents when reporting a false type.

θ_i *does not need to perform any strategy* $s_i(\theta_i) \in S_i$ *specified in any indirect mechanism* $\Gamma = (S_1, \cdots, S_I, g(\cdot))$, and consequently does not need to spend any strategy cost $c_i(s_i, \theta_i)$.[6] $\square$.

Proposition 1: For a given social choice function f, suppose that there exists an indirect mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ that implements it in Bayesian Nash equilibrium. If the format of each agent's strategy $s_i \in S_i$ is a costly action, *i.e.*, $c_i(s_i, \theta_i) > 0$, then it cannot be inferred that there exists a direct mechanism that can truthfully implement f in Bayesian Nash equilibrium.

Proof: Consider the given social choice function f, and the indirect mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ that implements it in Bayesian Nash equilibrium, then there exists a profile of strategies $s^*(\cdot) = (s_1^*(\cdot), \cdots, s_I^*(\cdot))$ such that the mapping $g(s^*(\cdot)) : \Theta_1 \times \cdots \times \Theta_I \to X$ from a vector of agents' types $\theta = (\theta_1, \cdots, \theta_I)$ into an outcome $g(s^*(\theta))$ is equal to the desired outcome $f(\theta)$, *i.e.*, $g(s^*(\theta)) = f(\theta)$ for all $\theta \in \Theta_1 \times \cdots \times \Theta_I$.

By Definition 2, for all i and all $\theta_i \in \Theta_i$,

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

for all $\hat{s}_i \in S_i$. Thus, for all i and all $\theta_i \in \Theta_i$, $\hat{\theta}_i \in \Theta_i$

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

Since $g(s^*(\theta)) = f(\theta)$ for all θ, then for all i and all $\theta_i \in \Theta_i$,

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

for all $\hat{\theta}_i \in \Theta_i$. Note that the above inequality cannot infer the inequality (6). Consequently, it cannot be inferred that there exists a direct mechanism that can truthfully implement f in Bayesian Nash equilibrium. $\square$

Discussion 2: Someone may disagree with Note 4 and Proposition 1, and propose a "direct revelation game" as follows. For a given social choice function f, suppose there is an indirect mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$ that implements f in Bayesian Nash equilibrium, and the equilibrium strategy is

[6] Someone may argue that in a direct mechanism $\bar{\Gamma} = (\bar{S}_1, \cdots, \bar{S}_I, \bar{g}(\cdot))$, in addition to choose a type $\bar{s}_i \in \Theta_i$ to report, each agent may also be *willing* to perform an additional strategy $s_i(\theta_i) \in S_i$ as what he would perform in some indirect mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$. However, this argument requires each agent to do beyond the framework of the direct mechanism, since this additional strategy $s_i(\theta_i) \in S_i$ is meaningless and illegal in the direct mechanism $\bar{\Gamma} = (\bar{S}_1, \cdots, \bar{S}_I, \bar{g}(\cdot))$.

$s^* = (s_1^*, \cdots, s_I^*)$. Consider this equilibrium, there is a mapping from vectors of agents' types into outcomes. Now we take the mapping to be a revelation game, *i.e.*, each agent i with private type θ_i independently chooses a type $\hat{\theta}_i \in \Theta_i$ to report to the designer, and the designer suggests each agent an action $s_i^*(\hat{\theta}_i) \in S_i$. Then no type of any agent can benefit by reporting a false type $\hat{\theta}_i \neq \theta_i$ and performing the suggested action $s_i^*(\hat{\theta}_i)$. As a result, truth-telling is the equilibrium strategy of this game, *i.e.*, each agent i reports his true type θ_i and performs the same strategy action $s_i^*(\theta_i)$ as what he would perform in the indirect mechanism.

Answer 2: It should be emphasized that in the direct revelation game, each agent i with private type θ_i can choose an *arbitrary* type $\hat{\theta}_i \in \Theta_i$ to report to the designer. Thus, in order to know which action $s_i^*(\hat{\theta}_i) \in S_i$ should be suggested to each agent, the designer must know not only one specific action $s_i^*(\theta_i)$, but also the full details of each agent i's strategy function $s_i^*(\cdot) : \Theta_i \to S_i$. However, *the designer is always at the information disadvantage in a mechanism*: she does not know each agent i's private type θ_i, or his private strategy function $s_i^*(\cdot) : \Theta_i \to S_i$. All that the designer knows from each agent i is just one report $s_i^*(\theta_i)$. [7] Therefore, the direct revelation game *takes it for granted* that the designer can suggest each agent which strategy action he should play after receiving a report type from each agent, but actually does not hold. $\square$

3 Conclusion

This paper mainly investigates the notion of Bayesian Nash equilibrium of a mechanism when strategies of agents are costly actions. This work is also relevant to the foundation of revelation principle. So far, there have been several discussions on possible failures of the revelation principle. Kephart and Conitzer [6] proposed that when reporting truthfully is costless and misreporting is costly, the revelation principle can fail to hold. Bester and Strausz [7] pointed out that the revelation principle may fail because of imperfect commitment. Martimort and Stole [8] said that the revelation principle does not apply to situations where several mechanism designers compete against each other.

The main result of this paper is that: When strategies of agents are costly actions, the definition of Bayesian Nash equilibrium of a mechanism should be based on the profit function. As a result, the notion of Bayesian incentive

[7] Otherwise, assume to the contrary that the designer knows each agent i's strategy function $s_i^*(\cdot)$, then she can easily infer each agent i's private type θ_i from his report $s_i(\theta_i)$. This case contradicts the basic framework of mechanism design and does not hold.

compatibility should also be revised. This is the key point why the revelation principle may fail when strategies of agents are costly actions.

References

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

[2] R. Myerson, Incentive compatibility and the bargaining problem, Econometrica, vol.47, 61-73, 1979.

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

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

[5] N. Kartik, Strategic Communication with Lying Costs. The Review of Economic Studies, vol.76, 1359-1395, 2009.

[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.

[7] 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.

[8] D. Martimort and L. Stole, The revelation and delegation principles in common agency games. Econometrica, Vol.70, No.4, 1659-1673, 2002.