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

Wu, Haoyang

15 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 mechanism design theory, a designer would like to implement a desired social choice function which specifies her favorite outcome for each possible profile of agents' types. To do so, the designer constructs a mechanism which describes each agent's feasible strategy set and the outcome function. 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 notion of Bayesian Nash equilibrium of a mechanism should be based on a three-parameter profit function instead of the conventional two-parameter utility function when strategies of agents are costly actions. 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 [1–4], 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

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

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

and unknown to the designer. In order to implement a social choice function in Bayesian Nash equilibrium, the designer constructs a mechanism which 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) (see MWG's Book, Page 883, Line 8, [1]). The distinction between the two formats is that: the former format of strategy is *a real action* which naturally requires some action cost to be performed realistically, whereas the latter format of strategy is *a message of action plan* which is reported by each agent to the designer and hence doesn't need action 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 notion of Bayesian Nash equilibrium of a mechanism should be based on a three-parameter profit function instead of a two-parameter utility function when strategies of agents are costly actions (*i.e.*, Definition 2). Next, we derive the main result (*i.e.*, Proposition 1): Given a social choice function f which is 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. We consider two possible arguments and give replies. 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 shown 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 $s_i(\theta_i)$, agent i with type θ_i shall spend some action cost (or make some effort which can be quantified as some action cost).
Case 2: If the format of strategy $s_i(\theta_i)$ is a message, then $s_i(\theta_i)$ is not a real action and hence doesn't need action cost to be performed realistically. □

In the following discussions, we will focus on the former case and investigate what would happen to the notion of Bayesian Nash equilibrium of 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. If each agent i's strategy $s_i(\theta_i) : \Theta_i \to S_i$ in the game induced by Γ is a costly action, 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.*, $c_i(s_i, \theta_i) > 0$ for each $s_i \in S_i$, $\theta_i \in \Theta_i$. Suppose the outcome yielded by Γ is $x \in X$ and agent i's utility is denoted by a two-parameter function $u_i(x, \theta_i) : X \times \Theta_i \to \mathcal{R}$, then each agent i's profit is defined by a three-parameter *profit function* $p_i(x, s_i, \theta_i) : X \times S_i \times \Theta_i \to \mathcal{R}$,

$$p_i(x, s_i, \theta_i) = u_i(x, \theta_i) - c_i(s_i, \theta_i). \qquad (1)$$

Question 1: Someone may argue that when each agent performs a costly strategy action, then the meaning of his utility has already changed, and has included the action cost. Thus, it is not necessary to introduce another notion of profit function to make confusion.

Answer 1: Generally speaking, there are two versions of utility function which are usually used in the literature of game theory and mechanism design:
1) *Three-parameter version*: For example, in Section 13.C (Page 450, the fourth line from the bottom, [1]), Mas-Colell, Whinston and Green use a three-parameter function $u(w, e|\theta) = w - c(e, \theta)$ to denote the utility of a type θ agent who plays a strategy (*i.e.*, choosing education level e) and receives an outcome (*i.e.*, the wage w), where $c(e, \theta)$ denotes the agent's cost of obtaining education level e. Obviously, *the three-parameter utility function of agent i does already include his strategy cost*, and indeed is agent i's profit function.

2) *Two-parameter version*: For example, in Section 23.B (Page 858, the fifth line from the bottom, [1]), Mas-Colell, Whinston and Green use a two-parameter function $u_i(x, \theta_i)$ to denote the utility of agent i with type θ_i after obtaining an outcome $x \in X$. Obviously, *the two-parameter utility function does not include any item to represent the action cost which may be spent by agent i to obtain the outcome x*. When agent i's strategy s_i is a costly action, *i.e.*, $c_i(s_i, \theta_i) > 0$, we must replace the two-parameter utility function $u_i(x, \theta_i)$ by the three-parameter profit function $p_i(x, s_i, \theta_i)$ to exactly describe how much each agent i benefits from the game induced by a mechanism. □

According to MWG's 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 $\theta_i \in \Theta_i$, $\hat{s}_i \in S_i$,

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

Note 2: As given above, the notion of Bayesian Nash equilibrium of a mechanism in the literature is always based on a two-parameter utility function $u_i(x, \theta_i) : X \times \Theta_i \to \mathcal{R}$. The parameters $\hat{s}_i, s_i^*, s_{-i}^*$ appeared in inequality (2) are used by the designer to compute the outcome $g(\cdot) \in X$, and does not act as parameters of agent i's 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.*, $c_i(s_i, \theta_i) > 0$. Then, as pointed out in Answer 1, the two-parameter utility function $u_i(x, \theta_i)$ neglects agent i's cost and cannot describe his net 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 $\theta_i \in \Theta_i$,

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

i.e.,

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

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

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

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 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 $\bar{\Gamma} = (\bar{S}_1, \cdots, \bar{S}_I, \bar{g}(\cdot))$, each agent i independently chooses his report strategy $\bar{s}_i(\cdot) : \Theta_i \to \Theta_i$, and the report type $\bar{s}_i(\theta_i)$ does not need to be his true type θ_i. Hence, the format of each agent i's strategy is a message, and it is reasonable to assume each agent i plays his strategy costlessly. [4] Thus, each agent's utility in the direct mechanism is just equal to his profit. Obviously, although the notion of Bayesian Nash equilibrium of a mechanism should be revised to Definition 2 when each agent's strategy is a costly action, the conventional notion of Bayesian incentive compatibility still holds as inequality (4) specifies.

Note 4: In a direct mechanism, the only thing that the designer gets from each agent i is the reported 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 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 $c_i(s_i, \theta_i)$. [5]

---

[3] Here we use a bar symbol to distinguish a direct mechanism from an indirect mechanism.

[4] Some researchers investigated misreporting costs in a direct mechanism [5,6], which are possibly spent by agents when reporting a false type. It should be noted that the misreporting cost is irrelevant to this paper. Our result holds no matter whether there exists the misreporting cost or not. Hence, we simply omit the misreporting cost in this paper.

[5] 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 $\Gamma = (S_1, \cdots, S_I, g(\cdot))$. Thus, each agent i with type θ_i also spends strategy cost $c_i(s_i, \theta_i)$ in the indirect mechanism. However, this argument requires each agent to do beyond the framework of the direct mechanism, since strategy $s_i(\theta_i) \in S_i$ is meaningless and not defined in the direct mechanism $\bar{\Gamma} = (\bar{S}_1, \cdots, \bar{S}_I, \bar{g}(\cdot))$.

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 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 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 $\theta_i \in \Theta_i$,

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

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

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

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 (4). Consequently, it cannot be inferred that there exists a direct mechanism that can truthfully implement f in Bayesian Nash equilibrium. □

Question 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 $s^* = (s_1^*, \cdots, s_I^*)$. 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 strategy action $s_i^*(\theta_i)$, the same as what he would perform in the indirect mechanism Γ.

Answer 2: It should be noted 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, which means that the corresponding $s_i^*(\hat{\theta}_i) \in S_i$ is not always equal to $s_i^*(\theta_i)$. Thus, after the designer receives an arbitrary report profile $(\hat{\theta}_1, \cdots, \hat{\theta}_I)$, in order to exactly know which $s_i^*(\hat{\theta}_i)$ should be suggested to each agent, the designer must know not only the specific $s_i^*(\theta_i)$, but also 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, nor his private strategy function $s_i^*(\cdot) : \Theta_i \to S_i$. [6] Therefore, the so-called direct revelation game does not hold. □

3 Conclusion

This paper mainly investigates the notion of Bayesian Nash equilibrium of a mechanism when the format of each agent's strategy is an action. The work is also relevant to the possible failure 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 a three-parameter profit function rather than the conventional two-parameter utility function (see Definition 2). This is the key point why the revelation principle may fail, and this failure is different from the current discussions of possible failures of the revelation principle.

Acknowledgments

The author is grateful to Fang Chen, Hanyue, Hanxing and Hanchen for their great support.

[6] Otherwise, assume to the contrary that the designer knows each agent i's strategy function $s_i^*(\cdot) : \Theta_i \to S_i$, 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.

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.