A Bayesian implementable social choice function may not be truthfully implementable

Wu, Haoyang

11 September 2016

A Bayesian implementable social choice function may not be truthfully implementable

Haoyang Wu *

Department of Economics, Xi'an Jiaotong University, Xi'an 710049, China.

Abstract

The revelation principle is a fundamental theorem in many economics fields such as game theory, mechanism design and auction theory etc. In this paper, I construct an example to show that a social choice function which can be implemented in Bayesian Nash equilibrium is not truthfully implementable. The key point is that agents pay cost in the indirect mechanism, but pay nothing in the direct mechanism. As a result, the revelation principle may not hold when agent's cost cannot be neglected in the indirect mechanism.

JEL codes: D70

Key words: Revelation principle; Game theory; Mechanism design; Auction theory.

1 Introduction

The revelation principle plays an important role in microeconomics theory and has been applied to many other fields such as auction theory, game theory etc. According to the wide-spread textbook given by Mas-Colell, Whinston and Green (Page 884, Line 24 [1]): "*The implication of the revelation principle is ... to identify the set of implementable social choice functions in Bayesian Nash equilibrium, we need only identify those that are truthfully implementable.*" Related definitions about the revelation principle can be seen in Appendix, which are cited from Section 23.B and 23.D of MWG's textbook[1].

* Corresponding author.
Email address: `18621753457@163.com, Tel: 86-18621753457` (Haoyang Wu).

However, in this paper, I will construct a simple labor model to show that the revelation principle may not hold when agent's cost cannot be neglected in the indirect mechanism. Section 2 is the main part of this paper, and Section 3 draws conclusions.

2 Main results

2.1 A labor model

Here we consider a simple labor model which uses some ideas from the first-price sealed auction model in Example 23.B.5 [1] and the signaling model in Section 13.C [1]. There are one firm and two workers. The firm wants to hire a worker, and two workers compete for this job offer. Worker 1 and Worker 2 differ in the number of units of output they produce if hired by the firm, which is denoted by productivity type.

For simplicity, we make the following assumptions:
1) The possible productivity types of two workers are: θ_L and θ_H, where $\theta_H > \theta_L > 0$. Each worker i's productivity θ_i $(i = 1, 2)$ is a random variable chosen independently, and is private information for each worker.
2) Before confronting the firm, each worker gets some education. The possible levels of education are: e_L and e_H, where $e_H > 0$, $e_L = 0$. Each worker's education is observable to the firm. Education does nothing for a worker's productivity.
3) The cost of obtaining education level e for a worker of some type θ is given by a function $c(e, \theta) = e/\theta$. That is, the cost of education is lower for a high-productivity worker.

The model's outcome can be represented by a vector (y_1, y_2), where y_i denotes the probability that worker i gets the job offer with wage $w > 0$. Recall that the firm does not know the exact productivity types of two workers, but its aim is to hire a worker with productivity as high as possible. This aim can be represented by a social choice function $f(\vec{\theta}) = (y_1(\vec{\theta}), y_2(\vec{\theta}))$, in which $\vec{\theta} = (\theta_1, \theta_2)$,

$$y_1(\vec{\theta}) = \begin{cases} 1, & \text{if } \theta_1 > \theta_2 \\ 0.5, & \text{if } \theta_1 = \theta_2 \\ 0, & \text{if } \theta_1 < \theta_2 \end{cases}, \quad y_2(\vec{\theta}) = \begin{cases} 1, & \text{if } \theta_1 < \theta_2 \\ 0.5, & \text{if } \theta_1 = \theta_2 \\ 0, & \text{if } \theta_1 > \theta_2 \end{cases} \tag{1}$$

In order to implement the above $f(\vec{\theta})$, the firm designs an indirect mechanism $\Gamma = (S_1, S_2, g)$ as follows:

1) A random move of nature determines the productivity of workers: $\theta_1, \theta_2 \in \{\theta_L, \theta_H\}$.
2) Conditional on his type θ_i, each worker $i = 1, 2$ chooses his education level as a bid $b_i : \{\theta_L, \theta_H\} \to \{0, e_H\}$. The strategy set S_i is the set of all possible bids $b_i(\theta_i)$, and the outcome function g is defined as:

$$g(b_1, b_2) = (p_1, p_2) = \begin{cases} (1, 0), & \text{if } b_1 > b_2 \\ (0.5, 0.5), & \text{if } b_1 = b_2 \\ (0, 1), & \text{if } b_1 < b_2 \end{cases} \tag{2}$$

where p_i $(i = 1, 2)$ is the probability that worker i gets the offer.

Let u_0 be the utility of the firm, and u_1, u_2 be the utilities of worker $1, 2$ respectively, then $u_0(b_1, b_2) = p_1\theta_1 + p_2\theta_2 - w$, and for $i, j = 1, 2$, $i \neq j$,

$$u_i(b_i, b_j; \theta_i) = \begin{cases} w - b_i/\theta_i, & \text{if } b_i > b_j \\ 0.5w - b_i/\theta_i, & \text{if } b_i = b_j \\ -b_i/\theta_i, & \text{if } b_i < b_j \end{cases} \tag{3}$$

The individual rationality (IR) constraints are: $u_i(b_i, b_j; \theta_i) \geq 0$, $i = 1, 2$.

2.2 f is Bayesian implementable

Proposition 1: If $w \in (2e_H/\theta_H, 2e_H/\theta_L)$, the social choice function $f(\vec{\theta})$ can be implemented by the indirect mechanism Γ in Bayesian Nash equilibrium.
Proof: Consider a separating strategy, *i.e.*, workers with different productivity types choose different education levels,

$$b_1(\theta_1) = \begin{cases} e_H, & \text{if } \theta_1 = \theta_H \\ 0, & \text{if } \theta_1 = \theta_L \end{cases}, \quad b_2(\theta_2) = \begin{cases} e_H, & \text{if } \theta_2 = \theta_H \\ 0, & \text{if } \theta_2 = \theta_L \end{cases}. \tag{4}$$

Now let us check whether this separating strategy yields a Bayesian Nash equilibrium. Assume $b_j^*(\theta_j)$ takes this form, *i.e.*,

$$b_j^*(\theta_j) = \begin{cases} e_H, & \text{if } \theta_j = \theta_H \\ 0, & \text{if } \theta_j = \theta_L \end{cases}, \tag{5}$$

then consider worker i's problem $(i \neq j)$. For each $\theta_i \in \{\theta_L, \theta_H\}$, worker i solves the maximization problem $\max_{b_i} h(b_i, \theta_i)$, where by Eq (3) the object function is

$$h(b_i, \theta_i) = (w - b_i/\theta_i)P(b_i > b_j^*(\theta_j)) + (0.5w - b_i/\theta_i)P(b_i = b_j^*(\theta_j)) - (b_i/\theta_i)P(b_i < b_j^*(\theta_j)) \tag{6}$$

We discuss this maximization problem in four different cases:
1) Suppose $\theta_i = \theta_j = \theta_L$, then $b_j^*(\theta_j) = 0$ by Eq (5).

$$\begin{aligned} h(b_i, \theta_i) &= (w - b_i/\theta_L)P(b_i > 0) + (0.5w - b_i/\theta_L)P(b_i = 0) - (b_i/\theta_L)P(b_i < 0) \\ &= \begin{cases} w - e_H/\theta_L, & \text{if } b_i = e_H \\ 0.5w, & \text{if } b_i = 0 \end{cases} \end{aligned}$$

Thus, if $w < 2e_H/\theta_L$, then $h(e_H, \theta_i) < h(0, \theta_i)$, which means the optimal value of $b_i(\theta_i)$ is 0. In this case, $b_i^*(\theta_L) = 0$.

2) Suppose $\theta_i = \theta_L$, $\theta_j = \theta_H$, then $b_j^*(\theta_j) = e_H$ by Eq (5).

$$\begin{aligned} h(b_i, \theta_i) &= (w - b_i/\theta_L)P(b_i > e_H) + (0.5w - b_i/\theta_L)P(b_i = e_H) - (b_i/\theta_L)P(b_i < e_H) \\ &= \begin{cases} 0.5w - e_H/\theta_L, & \text{if } b_i = e_H \\ 0, & \text{if } b_i = 0 \end{cases} \end{aligned}$$

Thus, if $w < 2e_H/\theta_L$, then $h(e_H, \theta_i) < h(0, \theta_i)$, which means the optimal value of $b_i(\theta_i)$ is 0. In this case, $b_i^*(\theta_L) = 0$.

3) Suppose $\theta_i = \theta_H$, $\theta_j = \theta_L$, then $b_j^*(\theta_j) = 0$ by Eq (5).

$$\begin{aligned} h(b_i, \theta_i) &= (w - b_i/\theta_H)P(b_i > 0) + (0.5w - b_i/\theta_H)P(b_i = 0) - (b_i/\theta_H)P(b_i < 0) \\ &= \begin{cases} w - e_H/\theta_H, & \text{if } b_i = e_H \\ 0.5w, & \text{if } b_i = 0 \end{cases} \end{aligned}$$

Thus, if $w > 2e_H/\theta_H$, then $h(e_H, \theta_i) > h(0, \theta_i)$, which means the optimal value of $b_i(\theta_i)$ is e_H. In this case, $b_i^*(\theta_H) = e_H$.

4) Suppose $\theta_i = \theta_j = \theta_H$, then $b_j^*(\theta_j) = e_H$ by Eq (5).

$$\begin{aligned} h(b_i, \theta_i) &= (w - b_i/\theta_H)P(b_i > e_H) + (0.5w - b_i/\theta_H)P(b_i = e_H) - (b_i/\theta_H)P(b_i < e_H) \\ &= \begin{cases} 0.5w - e_H/\theta_H, & \text{if } b_i = e_H \\ 0, & \text{if } b_i = 0 \end{cases} \end{aligned}$$

Thus, if $w > 2e_H/\theta_H$, then $h(e_H, \theta_i) > h(0, \theta_i)$, which means the optimal value of $b_i(\theta_i)$ is e_H. In this case, $b_i^*(\theta_H) = e_H$.

From the above four cases, it can be seen that if the wage $w \in (2e_H/\theta_H, 2e_H/\theta_L)$, the strategy $b_i^*(\theta_i)$ of worker i

$$b_i^*(\theta_i) = \begin{cases} e_H, & \text{if } \theta_i = \theta_H \\ 0, & \text{if } \theta_i = \theta_L \end{cases} \tag{7}$$

is the optimal response to the strategy $b_j^*(\theta_j)$ of worker j ($j \neq i$) given in Eq (5). Therefore, the strategy profile $(b_1^*(\theta_1), b_2^*(\theta_2))$ is a Bayesian Nash equilibrium of the game induced by Γ.

Now let us investigate whether the wage $w \in (2e_H/\theta_H, 2e_H/\theta_L)$ satisfies the individual rationality (IR) constraints. Following Eq (3) and Eq (7), the (IR) constraints are changed into: $0.5w - b_H/\theta_H > 0$. Obviously, $w \in (2e_H/\theta_H, 2e_H/\theta_L)$ satisfies the (IR) constraints.

In summary, if $w \in (2e_H/\theta_H, 2e_H/\theta_L)$, then by Eq(2) and Eq(7), for any $\vec{\theta} = (\theta_1, \theta_2)$, where $\theta_1, \theta_2 \in \{\theta_L, \theta_H\}$, there holds:

$$g(b_1^*(\theta_1), b_2^*(\theta_2)) = \begin{cases} (1,0), & \text{if } \theta_1 > \theta_2 \\ (0.5, 0.5), & \text{if } \theta_1 = \theta_2 \\ (0,1), & \text{if } \theta_1 < \theta_2 \end{cases}, \tag{8}$$

which is just the social choice function $f(\vec{\theta})$ given in Eq (1). **Q.E.D.**

2.3 f is not truthfully implementable in Bayesian Nash equilibrium

Before we discuss the truthful implementation problem, let us first cite the basic idea behind the revelation principle given in MWG's textbook (Page 884, Line 16, [1]): "If in mechanism $\Gamma = (S_1, \cdots, S_I, g(\cdot))$, each agent finds that, when his type is θ_i, choosing $s_i^*(\theta_i)$ is his best response to the other agents' strategies, then if we introduce a *mediator* who says 'Tell me your type, θ_i, and I will play $s_i^*(\theta_i)$ for you', each agent will find truth telling to be an optimal strategy given that all other agents tell the truth. That is, truth telling will be a Bayesian Nash equilibrium of this direct revelation game".

Although this basic idea looks reasonable, there indeed exists an assumption behind the mediator's announcement "*I will play $s_i^*(\theta_i)$ for you.*" The underlying assumption is the following cost condition:
Cost condition: when the mediator plays $s_i^*(\theta_i)$ for each agent i, he also pays the cost which would be spent by agent i himself when carrying out the strategy $s_i^*(\theta_i)$ in the original mechanism Γ.

Obviously, only when the cost condition holds will the mediator's announcement be credible to the agents. *Otherwise none of agents will believe that the mediator can play $s_i^*(\theta_i)$ for him, which means no agent will attend the direct revelation game.* Let us take a look at the proof of revelation principle given in Appendix Proposition 23.D.1. In Eq (23.D.3), agents pay costs by themselves when playing strategies $s_i^*(\theta_i)$ ($i = 1, \cdots, I$) in the original mechanism

Γ. As a comparison, in Eq (23.D.4), the direct mechanism works. As specified in Definition 23.B.5, now the strategy set of each agent i is just the same as his type set, $S_i = \Theta_i$. At this time, all things that each agent has to do are to simply announce a type, and each agent does not need to pay any cost by himself.

However, the cost condition is indeed a *paradox*. On one hand, as we have seen above, the cost condition is the cornerstone for the direct revelation mechanism to start up. On the other hand, it is just the cost condition itself that makes a Bayesian implementable social choice function may not be truthfully implementable. I will prove the latter through Proposition 2.

Proposition 2: The social choice function $f(\vec{\theta})$ given in Eq (1) is not truthfully implementable in Bayesian Nash equilibrium.
Proof: Consider the direct revelation mechanism $\Gamma_{direct} = (\Theta_1, \Theta_2, f(\vec{\theta}))$, in which $\Theta_1 = \Theta_2 = \{\theta_L, \theta_H\}$, $\vec{\theta} \in \Theta_1 \times \Theta_2$. The timing steps of Γ_{direct} are as follows:
1) A random move of nature determines the productivity types of workers: $\theta_i \in \Theta_i$ $(i = 1, 2)$, and each worker i announces a type $\hat{\theta}_i \in \Theta_i$ to a mediator. Here $\hat{\theta}_i$ may not be his true type θ_i.
2) The mediator plays the strategy $b_i^*(\hat{\theta}_i)$ $(i = 1, 2)$ for each agent i, and submits the bids to the firm:

$$b_i^*(\hat{\theta}_i) = \begin{cases} e_H, & \text{if } \hat{\theta}_i = \theta_H \\ 0, & \text{if } \hat{\theta}_i = \theta_L \end{cases}$$

3) The firm performs the outcome function $g(b_1, b_2)$, and hires the winner.

As we have discussed, according to the cost condition, each worker i does not need to pay the cost b_i/θ_i by himself when playing in the direct mechanism. Hence, the utility function of each worker $i = 1, 2$ is changed from Eq (3) to the follows:

$$u_i(\hat{\theta}_i, \hat{\theta}_j; \theta_i) = \begin{cases} w, & \text{if } \hat{\theta}_i > \hat{\theta}_j \\ 0.5w, & \text{if } \hat{\theta}_i = \hat{\theta}_j \text{ }, i \neq j \\ 0, & \text{if } \hat{\theta}_i < \hat{\theta}_j \end{cases} \quad (9)$$

The utility matrix of worker i and j can be expressed as follows.

$\hat{\theta}_i$ \ $\hat{\theta}_j$	θ_L	θ_H
θ_L	$[0.5w, 0.5w]$	$[0, w]$
θ_H	$[w, 0]$	$[0.5w, 0.5w]$

Obviously, the dominant strategy for each worker i is to definitely announce $\hat{\theta}_i = \theta_H$, no matter what his true productivity type θ_i is. The unique outcome of Γ_{direct} is that each worker has the same probability 0.5 to get the job offer.

Thus, the unique equilibrium in the direct revelation mechanism Γ_{direct} is the dominant Nash equilibrium $(\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H)$. Consequently, the social choice function $f(\vec{\theta})$ is not truthfully implementable in Bayesian Nash equilibrium. **Q.E.D.**

3 Conclusions

From Section 2, it can be seen that:
1) In the indirect mechanism Γ, the utility function of each worker $i = 1, 2$ is given by Eq (3). The cost b_i/θ_i is the key item that makes the separating strategy profile $(b_1^*(\theta_1), b_2^*(\theta_2))$ be a Bayesian Nash equilibrium, when the wage $w \in (2e_H/\theta_H, 2e_H/\theta_L)$. Thus, the social choice function f can be implemented in Bayesian Nash equilibrium.

2) In Section 2.3, I point out that the cost condition is underlied in the traditional explanation of revelation principle. Following the cost condition, the utility function of each worker i is represented by Eq (9). According to the utility matrix of workers, the unique equilibrium of the game induced by Γ_{direct} is the dominant Nash equilibrium $(\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H)$. Hence, the social choice function f cannot be truthfully implemented in Bayesian Nash equilibrium.

In summary, the revelation principle may not hold when agent's cost cannot be neglected in the indirect mechanism.

Appendix: Definitions in Section 23.B and 23.D [1]

Consider a setting with I agents, indexed by $i = 1, \cdots, I$. Each agent i privately observes his type θ_i that determines his preferences. The set of possible types of agent i is denoted as Θ_i. The agent i's utility function over the outcomes in set X given his type θ_i is $u_i(x, \theta_i)$, where $x \in X$.

Definition 23.B.1: A *social choice function* 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 $g : S_1 \times \cdots \times S_I \to X$.

Definition 23.B.5: A *direct revelation mechanism* is a mechanism in which $S_i = \Theta_i$ for all i and $g(\theta) = f(\theta)$ for all $\theta \in \Theta_1 \times \cdots \times \Theta_I$.

Definition 23.D.1: 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 $\theta_i \in \Theta_i$,

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

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

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* if $s_i^*(\theta_i) = \theta_i$ (for all $\theta_i \in \Theta_i$ and $i = 1, \cdots, I$) is a Bayesian Nash equilibrium of the direct revelation mechanism $\Gamma = (\Theta_1, \cdots, \Theta_I, f(\cdot))$. That is, if for all $i = 1, \cdots, I$ and all $\theta_i \in \Theta_i$,

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i], \qquad (23.D.1)$$

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

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.
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 $\theta_i \in \Theta_i$,

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

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

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

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 $\theta_i \in \Theta_i$,

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

for all $\hat{\theta}_i \in \Theta_i$. But, this is precisely condition (23.D.1), 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.