A costly Bayesian implementable social choice function may not be truthfully implementable

Wu, Haoyang

11 September 2016

A costly Bayesian implementable social choice function may not be truthfully implementable

Haoyang Wu *

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

Abstract

The revelation principle is a fundamental theorem in many economics fields. In this paper, we construct a simple labor model to show that a social choice function which can be implemented costly in Bayesian Nash equilibrium may not be truthfully implementable. The key point is the strategy cost condition given in Section 4: each agent pays cost when performing strategy in the indirect mechanism, but will not pay the strategy cost in the direct mechanism. As a result, the revelation principle may not hold when agents' strategies are costly 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).

In the traditional literatures of mechanism design, costs are usually ignored during the process of a mechanism. Until recently, some researchers began to investigate costs occurred in a mechanism. For every type θ and every type $\hat{\theta}$ an agent might misreport, Kephart and Conitzer [2] define a cost function as $c(\theta, \hat{\theta})$ for doing so. Traditional mechanism design is just the case where $c(\theta, \hat{\theta}) = 0$ everywhere, and partial verification is a special case where $c(\theta, \hat{\theta}) \in \{0, \infty\}$ [3–5]. Kephart and Conitzer [2] proposed that when reporting truthfully is costless and misreporting can be costly, the revelation principle can fail to hold.

Despite these accomplishments, up to now, people seldom consider two different kinds of costs simultaneously: 1) *strategy cost*, which is occurred when agents play strategies in an indirect mechanism; and 2) *misreporting cost*, which is occurred when agents report types falsefully in a direct mechanism. It is usually assumed that an agent can report truthfully with zero cost. The aim of this paper is to investigate the justification of revelation principle when the two kinds of costs are considered simultaneously. By constructing a simple labor model, we show that the revelation principle may not hold when agents' strategies are costly in the indirect mechanism.

The paper is organized as follows. In Section 2, we construct a social choice function f and an indirect mechanism, where agents' strategies are costly. In Section 3, we prove f can be implemented by the indirect mechanism in Bayesian Nash equilibrium. In Section 4, we propose a strategy cost condition by analyzing the basic idea behind the revelation principle. In Section 5, we prove that f is not truthfully implementable in Bayesian Nash equilibrium, which contradicts the revelation principle. Finally, Section 6 draws conclusions.

2 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 $e_L = 0$, $e_H > 0$. Each worker's education is observable to the firm. Education does nothing for a worker's productivity.
3) The strategy 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 strategy cost of education is lower for a high-productivity worker.
4) The misreporting cost for a low-productivity worker to report the high productivity type θ_H is a fixed value $c' > 0$. In addition, a high-productivity worker is assumed to report the low productivity type θ_L with zero cost.

The labor model's outcome is 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 types of two 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 $1, 2$ in the indirect mechanism Γ 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 item "$-b_i/\theta_i$" occurred in Eq (3) is just the strategy cost paid by agent

i of type θ_i when he performs the strategy $b_i(\theta_i)$ in the indirect mechanism.

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

3 f is Bayesian implementable

Proposition 1: If $w \in (2e_H/\theta_H, 2e_H/\theta_L)$, the social choice function $f(\vec{\theta})$ given in Eq (1) 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 a 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). $\square$

4 Strategy cost condition

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 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, it should be emphasized that there indeed exists a condition behind the mediator's announcement "*Tell me your type*, θ_i, *I will play* $s_i^*(\theta_i)$ *for you.*" The underlying condition is denoted as the following strategy cost condition:

Strategy cost condition: After receiving each agent i's type θ_i, in order to play $s_i^*(\theta_i)$ for agent i, the mediator must also pay the strategy cost which would be paid by agent i himself when carrying out $s_i^*(\theta_i)$ in the original mechanism.

Obviously, only when the strategy cost condition holds will the mediator's announcement "*I will play* $s_i^*(\theta_i)$ *for you*" be credible to the agents. Otherwise none of agents will believe the mediator's announcement and hence no agent will attend the direct mechanism, which means the direct mechanism cannot start up.

There is another viewpoint to consider the justification of strategy cost condition. Let us take a look at the proof of revelation principle given in Appendix Proposition 23.D.1. In Eq (23.D.3), the original mechanism Γ works: each agent i pays the strategy cost by himself when carrying out $s_i^*(\theta_i)$ $(i = 1, \cdots, I)$, and the designer carries out the outcome function g. As a comparison, in Eq (23.D.4), the direct mechanism works: at this time the strategy set of agent i is just his type set, $S_i = \Theta_i$, and the designer carries out the outcome function f. Hence, all things that each agent i has to do in the direct mechanism are only to announce a type θ_i, which requires no strategy cost at all. Put differently, *by Definition 23.B.5 of the direct mechanism, each agent i only announces a type θ_i and does not carrying out $s_i^*(\theta_i)$, hence does not need to pay any strategy cost by himself.* It should be noted that if some agent misreports his type in the direct mechanism, then he will pay the misreporting cost as we have assumed before.

Some possible questions to the strategy cost condition are as follows:
Q1: In the above explanation of direct mechanism, the mediator is actually a virtual role and does not exist at all.

A1: The notion "mediator" occurred in the strategy cost condition can be replaced by the notion "designer", and the following discussions are the same.

Q2: In the direct mechanism, after each agent i announces his type θ_i, he will still pay the strategy cost by himself.
A2: This viewpoint is in contrast to Definition 23.B.5 of the direct mechanism (See Appendix). It should be emphasized that by Definition 23.B.5, the action $s_i^*(\theta_i)$ is *illegal* for agent i. Thus, it is *wrong* to claim that agent i will still pay the strategy cost related to the illegal action $s_i^*(\theta_i)$ in the direct mechanism.

Q3: The designer may define the direct mechanism more generally. In particular, The designer defines a new mechanism in which each agent reports his type, then the mechanism suggest to them which action to take, and the final outcome of the mechanism depends on both the report and the action (*i.e.*, education level in this paper).
A3: As Myerson pointed out in Ref [6], the concepts of direct mechanism and revelation principle are in the field of static or one-stage games. However, the new mechanism is in the field of dynamic or multistage games and hence is irrelevant to our discussion.

Q4: Let us consider the equilibrium in the indirect mechanism. Given the equilibrium, there is a mapping from vectors of agents' types into outcomes. Now let us take that mapping to be a revelation game. It will be the case that no type of any agent can make an announcement that differs from his true type and do better.
A4: This viewpoint ignores the strategy costs occurred in the mechanism. Similar to the above analysis of the proof of Proposition 23.D.1, the strategy costs occurred in the equilibrium in the indirect mechanism are paid by agents themselves. But, consider the mapping from vectors of agents' types into outcomes, at this time each agent will not pay the strategy cost any more by Definition 23.B.5. As a result, the utility function of each agent may be changed (See Eq (9) in Section 5), and some agent may find it beneficial for him to differ from his true type.

To sum up, the strategy cost condition is the cornerstone for the direct revelation mechanism to start up. However, in Section 5, we will show that *it is the strategy cost condition itself that makes a Bayesian implementable social choice function may not be truthfully implementable, which eventually contradicts the revelation principle.*

5 f is not truthfully implementable in Bayesian Nash equilibrium

Proposition 2: If the misreporting cost $c' \in (0, 0.5w)$, 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 reports 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.

According to the strategy cost condition, each worker i does not need to pay the strategy cost b_i/θ_i by himself anymore in the direct mechanism. As assumed before, the only cost needed to pay in the direct mechanism is the misreporting cost c' for a low-productivity worker to report the high productivity type θ_H. For worker i $(i = 1, 2)$, if his type is $\theta_i = \theta_L$, his utility function will be as follows:

$$u_i(\hat{\theta}_i, \hat{\theta}_j; \theta_i = \theta_L) = \begin{cases} w - c', & \text{if } (\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_L) \\ 0.5w - c', & \text{if } (\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H) \\ 0.5w, & \text{if } (\hat{\theta}_i, \hat{\theta}_j) = (\theta_L, \theta_L) \\ 0, & \text{if } (\hat{\theta}_i, \hat{\theta}_j) = (\theta_L, \theta_H) \end{cases}, \ i \neq j. \tag{9}$$

If worker i's type is $\theta_i = \theta_H$, his utility function will be as follows:

$$u_i(\hat{\theta}_i, \hat{\theta}_j; \theta_i = \theta_H) = \begin{cases} w, & \text{if } (\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_L) \\ 0.5w, & \text{if } (\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H), or (\theta_L, \theta_L) \\ 0, & \text{if } (\hat{\theta}_i, \hat{\theta}_j) = (\theta_L, \theta_H) \end{cases}, \ i \neq j. \tag{10}$$

It can be seen that the strategy cost "$-b_i/\theta_i$" occurred in Eq (3) disappears in Eq (9) and Eq (10). Now, following Eq (9) and Eq (10), we discuss the utility matrix of worker i and j in four cases.

1) The true types of worker i and j are $\theta_i = \theta_H$, $\theta_j = \theta_H$.

$\hat{\theta}_i$ \ $\hat{\theta}_j$	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 worker i and j is to truthfully report, *i.e.*, $\hat{\theta}_i = \theta_H$, $\hat{\theta}_j = \theta_H$. Thus, the unique Nash equilibrium is $(\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H)$.

2) The true types of worker i and j are $\theta_i = \theta_L$, $\theta_j = \theta_H$.

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

It can be seen that: the dominant strategy for worker j is still to truthfully report $\hat{\theta}_j = \theta_H$; and if the misreporting cost $c' < 0.5w$, the dominant strategy for worker i is to misreport $\hat{\theta}_i = \theta_H$, otherwise agent i should truthfully report. Thus, under the condition $c' < 0.5w$, the unique Nash equilibrium is $(\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H)$.

3) The true types of worker i and j are $\theta_i = \theta_H$, $\theta_j = \theta_L$.

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

It can be seen that: the dominant strategy for worker i is still to truthfully report $\hat{\theta}_i = \theta_H$; and if the misreporting cost $c' < 0.5w$, the dominant strategy for worker j is to misreport $\hat{\theta}_j = \theta_H$, otherwise agent j should truthfully report. Thus, under the condition $c' < 0.5w$, the unique Nash equilibrium is $(\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H)$.

4) The true types of worker i and j are $\theta_i = \theta_L$, $\theta_j = \theta_L$.

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

It can be seen that: if the misreporting cost $c' < 0.5w$, the dominant strategy for both worker i and worker j is to misreport, *i.e.*, $\hat{\theta}_i = \theta_H$, $\hat{\theta}_j = \theta_H$, otherwise both agents should truthfully report. Thus, under the condition $c' < 0.5w$, the unique Nash equilibrium is $(\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H)$.

To sum up, under the condition $0 < c' < 0.5w$, the unique Nash equilibrium of the game induced by the direct mechanism is $(\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H)$, and the unique outcome of Γ_{direct} is that each worker has the same probability 0.5 to get the job offer.

Consequently, the social choice function $f(\vec{\theta})$ is not truthfully implementable in Bayesian Nash equilibrium. □

6 Conclusions

In this paper, we discuss the justification of revelation principle through a simple labor model in which agents pay strategy costs during the process of an indirect mechanism. The main characteristics of the labor model are as follows: 1) In the indirect mechanism, carrying out strategy is costly, *i.e.*, worker of type θ_H pays the strategy cost e_H/θ_H when obtaining education level e_H; 2) The productivity type of worker is private information and not observable to the firm; 3) Misreporting a higher type is also costly, *i.e.*, a low-productivity worker can pretend to be a high-productivity worker with the misreporting cost c'.

The major difference between this paper and traditional literatures is focused on the strategy cost condition given in Section 4. 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), in which the strategy 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 if 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) Following the strategy cost condition, in the direct mechanism, the utility function of each worker i is changed from Eq (3) to Eq (9) and Eq (10). Under the condition $c' \in (0, 0.5w)$, the unique Nash equilibrium of the game induced by the direct mechanism is $(\hat{\theta}_i, \hat{\theta}_j) = (\theta_H, \theta_H)$. Thus, the social choice function f is not truthfully implemented in Bayesian Nash equilibrium.

In summary, the revelation principle may not hold when agents' strategies are costly 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 $f : \Theta_1 \times \cdots \times \Theta_I \rightarrow X$ 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 $g : S_1 \times \cdots \times S_I \rightarrow X$.

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.

Acknowledgments

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

References

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

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

[3] J. Green and J.J. Laffont, Partially verifiable information and mechanism design. Review of Economic Studies, vol.53, 447-456, 1986.

[4] L. Yu, Mechanism design with partial verification and revelation principle. Autonomous Agents and Multi-Agent Systems, vol.22, 217-223, 2011.

[5] V. Auletta, *et al*, Alternatives to truthfulness are hard to recognize. Autonomous Agents and Multi-Agent Systems, vol.22, 200-216, 2011.

[6] R.B. Myerson, Multistage games with communication, Econometrica, vol.54, No.2, 323-358, 1986.