Bugs in the proofs of revelation principle

Wu, Haoyang

Wan-Dou-Miao Research Lab, Suite 1002, 790 WuYi Road, Shanghai, China.

19 August 2010

Bugs in the proofs of revelation principle

Haoyang Wu *

Wan-Dou-Miao Research Lab, Suite 1002, 790 WuYi Road,
Shanghai, 200051, China.
Tel: 86-13818470974, Fax: 86-29-82668725

Abstract

In the field of mechanism design, the revelation principle has been known for decades. Myerson, Mas-Colell, Whinston and Green gave formal proofs of the revelation principle. However, in this paper, we argue that there are bugs hidden in their proofs.

JEL codes: D7

Key words: Revelation principle; Mechanism design; Implementation theory.

1 Introduction

The revelation principle is well-known in the economic literature. It has several versions of representations [1–3]. See Page 884, Line 24 [3]: "The implication of the revelation principle is ... to identify the set of implementable social choice functions (now in Bayesian Nash equilibrium), we need only identify those that are truthfully implementable."

Although the revelation principle is fundamental in the field of mechanism design, in this paper we will argue that there are bugs in two versions of proofs. The rest of the paper are organized as follows: In Section 2, we will analyse the bug in Myerson's proof [1]. Then, we will point out the bugs in the proofs given by Mas-Colell, Whinston and Green [3].

* Corresponding author.
Email address: hywch@mail.xjtu.edu.cn (Haoyang Wu).

2 The bug in Myerson's proof

In this section, the notation is referred to Ref. [1]. The bug will be showed by the following three claims. We use the capital form to emphasize key words.

Claim 1: For each agent i, t_i and d_i are his private information and unknown to the principal.

Proof: See Page 69, Line 26 [1], "...each type t_i in T_i represents a complete description of all the private information i might have about his environment, his abilities and his preferences. Each private decision-option d_i in D_i may represent, for example, a level of effort which agent i might exert in working for the principal, and which the principal cannot observe or control". Obviously, Claim 1 holds.

Claim 2: For each agent i, the two mappings $\rho_i : T_i \to R_i$ and $\delta_i : M_i \times T_i \to D_i$ are his private information and unknown to the principal.

Proof: See Page 71, Line 19 [1], "In the context of this coordination mechanism $((R_i, M_i)_{i=1}^n, \pi)$, each agent i controls his choice of reporting strategy in R_i as a function of his type, and controls his choice of a decision in D_i as a function of his type and his message received. That is, agent i SELECTS a pair of functions $\rho_i : T_i \to R_i$ and $\delta_i : M_i \times T_i \to D_i$, such that $\rho_i(t_i)$ would be agent i's reporting strategy if i were of type t_i, and $\delta_i(m_i, t_i)$ would be i's final decision in D_i after he received message m_i if his type were t_i".

Generally, each agent i acts independently and self-interestedly when he selects his participation strategy (ρ_i, δ_i). Any agent has incentives to report dishonestly and act disobediently whenever doing so is better for him. Thererfore, the two mappings $\rho_i : T_i \to R_i$ and $\delta_i : M_i \times T_i \to D_i$ must be agent i's private information.

Claim 3: There is a bug in Myerson's proof.

Proof: See Page 74, Line 1 [1], "Given the equilibrium of participation strategies $(\rho_i, \delta_i)_{i=1}^n$, let $\delta^{-1}(d, t)$ be the set of all messages to the agents such that each agent i would respond by choosing decision d_i if his type were t_i. That is,

$$\delta^{-1}(d, t) = \{m | \delta_i(m_i, t_i) = d_i, \text{ for all } i\}."$$

It is implicit WHO IS ABLE to calculate $\delta^{-1}(d, t)$ for any arbitrarily given d and t. We emphasize the ability to do the calculation because anybody who wants to calculate $\delta^{-1}(d, t)$ must be able to get all necessary data. As discussed in Claim 2, the mapping δ_i is private information of agent i and unknown to the principal. It is impossible for the principal to calculate $\delta^{-1}(d, t)$ when he is given some $d \in D$ and $t \in T$.

See Page 74, Line 5 [1], "Then, define $\pi^* : D \times T \to R$ so that

$$\pi^*(d|t) = \sum_{m \in \delta^{-1}(d,t)} \pi(d_0, m|\rho_1(t_1), \cdots, \rho_n(t_n)).$$

π^* is the direct coordination mechanism which simulates the overall effect of the original mechanism with the given participation strategies".

It is also implicit WHO IS ABLE to calculate $\rho_1(t_1), \cdots, \rho_n(t_n)$ for some given $t_1, \cdots, t_n$. As discussed in Claim 2, the mapping ρ_i is private information of agent i and unknown to the principal. Therefore, although in Bayesian Nash equilibrium, each agent i has incentives to truthfully report $\rho_i(t_i)$ to the principal if his type is t_i, it is unreasonable to assume that each agent i has incentives to TRUTHFULLY report $\rho_i(\hat{t}_i)$ to the principal for all possible $\hat{t}_i \in T_i$, $\hat{t}_i \neq t_i$ when his type is t_i. Hence, it is impossible for the principal to calculate $\rho_1(t_1), \cdots, \rho_n(t_n)$ for some given $t_1, \cdots, t_n$.

To sum up, the principal cannot calculate $\delta^{-1}(d,t)$ or $\rho_1(t_1), \cdots, \rho_n(t_n)$. Consequently, the principal cannot calculate π^*. In Page 6, Line 24 [2], Myerson assumed a virtual person (named mediator) to calculate π^*: "The assumption that perfectly trustworthy mediators are available is essential to the mathematical simplicity of the incentive-compatible set". Therefore, π^* can be calculated only by the assumed mediator, NOT BY THE PRINCIPAL. However, in Page 73, Proposition 2 [1], Myerson said: "... there exists an incentive-compatible direct mechanism π^* in which the PRINCIPAL gets the same expected utility...".

That's the bug!

3 The bugs in the proofs by Mas-Colell, Whinston and Green

In this section, the notation follows from Ref. [3].

3.1 The revelation principle for Bayesian Nash equilibrium

To derive Eq (23.D.3), the authors replace $\hat{s}_i$ in Eq (23.D.2) by $s_i^*(\hat{\theta}_i)$ (for all $\hat{\theta}_i \in \Theta_i$). Since Eq (23.D.2) holds for all $\hat{s}_i \in S_i$, it looks reasonable to do so at first sight.

As we have pointed out in Section 2, for each agent $i \in I$, his strategy (here s_i^*) is private information. Therefore, although in Bayesian Nash equilibrium, for all i and all $\theta_i \in \Theta_i$, agent i has incentives to truthfully report $s_i^*(\theta_i)$ to the principal when agent i's type is θ_i (See Eq (23.D.2)), it cannot be deduced that each agent i has incentives to TRUTHFULLY report $s_i^*(\hat{\theta}_i)$ (for all $\hat{\theta}_i \in \Theta_i$, $\hat{\theta}_i \neq \theta_i$) to the principal given that his type is θ_i. Indeed, for all i, the true value of $s_i^*(\hat{\theta}_i)$ (for all $\hat{\theta}_i \in \Theta_i$, $\hat{\theta}_i \neq \theta_i$) is NOT AVAILABLE to the principal.

In Page 884, Line 20 [3], the authors assume a mediator:" ... 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." Therefore, the item $g(s_i^*(\hat{\theta}_i), s_{-i}^*(\theta_{-i}))$ in the right part of Eq (23.D.3) can be calculated only by the assumed mediator, NOT BY THE PRINCIPAL.

However, according to Definition 23.B.1 (Page 859, [3]), the social choice function $f : \Theta_1 \times \cdots \times \Theta_I \to X$ is specified by the principal, not by some assumed mediator. Hence, the item $f(\hat{\theta}_i, \theta_{-i})$ in the right part of Eq (23.D.4) can be calculated by the PRINCIPAL. To sum up, it is impossible to derive Eq (23.D.4) from Eq (23.D.3). That's the bug!

3.2 The revelation principle for dominant strategies

To derive Eq (23.C.5), the authors substitute $s_i^*(\hat{\theta}_i)$, $s_{-i}^*(\theta_{-i})$ for $\hat{s}_i$, s_{-i} in Eq (23.C.4) respectively. Since Eq (23.C.4) holds for all $\hat{s}_i \in S_i$ and all $s_{-i} \in S_{-i}$, it looks reasonable to do so at first sight.

As we have pointed out, for each agent $i \in I$, his strategy s_i^* is private information. Therefore, although for all i and all $\theta_i \in \Theta_i$, agent i has incentives to truthfully report $s_i^*(\theta_i)$ to the principal when his type is θ_i (See Eq (23.C.4)), it cannot be deduced that each agent i has incentives to TRUTHFULLY report $s_i^*(\hat{\theta}_i)$ (for all $\hat{\theta}_i \in \Theta_i$, $\hat{\theta}_i \neq \theta_i$) to the principal given that his type is θ_i. Similar to the discussion in Section 3.1, the item $g(s_i^*(\hat{\theta}_i), s_{-i}^*(\theta_{-i}))$ in the right part of Eq (23.C.5) CANNOT be calculated by the principal.

However, since the social choice function f is specified by the principal, the item $f(\hat{\theta}_i, \theta_{-i})$ (for all $\hat{\theta}_i \in \Theta_i$ and all $\theta_{-i} \in \Theta_{-i}$) in the right part of Eq (23.C.3) CAN be calculated by the principal. To sum up, it is impossible to derive Eq (23.C.3) from Eq (23.C.5). That's the bug!

Acknowledgments

The author is very grateful to Ms. Fang Chen, Hanyue Wu (*Apple*), Hanxing Wu (*Lily*) and Hanchen Wu (*Cindy*) for their great support.

References

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

[2] Myerson, R., http://home.uchicago.edu/ rmyerson/research/revnprl.pdf

[3] Mas-Colell, A., MD Whinston, and JR Green, "Microeconomic Theory", Oxford: Oxford University Press, 1995.