An Introduction to the Revelation Principle

Andrew Clausen <clausen@econ.upenn.edu> *

November 23, 2005

1 Introduction

I am aware of two ways to think about the revelation principle. The first one is about the existence of simple games that are equivalent to more complicated ones. The second one is about implementation of social welfare functions. Both approaches boil down to the same thing, but the first approach has better intuition while the second approach is a useful abstraction for game designers with a particular objective.

2 Intuitive Approach

In my home town, most houses are sold by an English auction. On a sunny Saturday morning, the house is opened to the public. Then, in mid-afternoon, the auctioneer stands on the footpath (sidewalk) outside the house, and begins the auction after talking up the good points of the house. A good auction usually has about 50 people standing on the road with about 10 serious bidders, and the remainder being accompanying family, and curious neighbours.

The auctioneer begins by requesting an initial bid from the public. After a few painful minutes, someone puts in a seriously undervalued bid, which the auctioneer usually accepts with a few insults. Once an initial bid has been accepted, anyone can place a higher bid. The auction ends when no-one wants to place a higher bid. The person who placed the winning bid gets the house, and pays the price equal to his bid.

If we assume that everyone has the same information about the house, it is clearly a Bayes-Nash equilibrium for everyone to keep on placing bids until the leading bid is higher than how much they are prepared to pay. So, we have a game Γ (the English auction game), and a Bayes-Nash equilibrium $s*$. Now, suppose that instead of announcing bids to the auctioneer, you (player i) instead handed the auctioneer your bidding strategy $s*_i$ (along with your private valuation). If the auctioneer collected everyone's bidding strategies, he could then simulate the entire auction on his own, and compute the outcome of the auction. Of course, you might not trust the auctioneer to do this honestly... but this is a separate issue. If you did trust the auctioneer to execute your strategy properly, then you would reveal your private information, and get exactly the same outcome. So announcing your private information is a Bayes-Nash equilibrium of the simulated English auction game Γ'. This is what the first interpretation of the revelation principle says.

That is, for any Bayes-Nash equilibrium $s*$ of any game Γ (of asymmetric information), there is another game Γ' in which each player chooses what (mis)information to announce about their private information, and that the simple-minded strategy profile s_{honest} in which everyone announces their private information is a Bayes-Nash equilibrium with the same outcome as $s*$ in the original game Γ.

Now, the after reading the abstract approach, it will be obvious that this statement is equivalent to the abstract version of the revelation principle. (Note that "mechanism" means a game and a Bayes-Nash equilibrium of it, and that the outcome of the game corresponds to a social choice function.)

*This approach is taken from Peter Bardsley's presentation of this topic. He has not reviewed this document, and is not responsible for any defects.

3 Abstract Approach

Definition 1 (Design Setting). *Let I be a set of players, T be a set of types (private information), p be a common prior probability measure on T, Z be a set of possible outcomes, and g_i be the utility functions for each type/outcome pair. (I, T, p, Z, g) is a design problem if*

- $T = \prod_{i \in I} T_i$.
- $p \in \Delta T$.
- $g_i : Z \times T \to \mathbb{R}$.

Definition 2 (Mechanism). *(A, f, s) is a mechanism for a design setting (I, T, p, Z, g) if*

- $A = \prod_{i \in I} A_i$.
- $f : A \to \Delta Z$.
- $s_i : T_i \to A_i$.

Definition 3 (Social Choice Function). *For any design setting (I, T, p, Z, g), any $\xi : T \to \Delta Z$ is a social choice function.*

Definition 4 (Induced Mechanism Game). *If (A, f, s) is a mechanism for a design setting (I, T, p, Z, g), then the induced Bayesian game is (I, T, A, u), where $u(s, t) = g(f(s(t)), t)$. [$s(t)$ is defined in the obvious way.]*

Definition 5 (Implementation). *A mechanism (A, f, s) implements the social choice function ξ in the design setting (I, T, p, Z, g) if $f \circ s = \xi$ and s is a Bayes-Nash equilibrium of the induced mechanism game.*

Definition 6 (Direct Mechanism). *The direct mechanism for a design setting (I, T, p, Z, g) and social choice function $\xi : T \to \Delta Z$ is (T, ξ, s) where $s_i(t_i) = t_i$.*

Theorem 1 (Revelation Principle). *If ξ is implementable by some mechanism in a design setting, then the direct mechanism implements ξ.*

Proof. Suppose the mechanism (A, f, s) implements ξ in the design setting (I, T, p, Z, g). Then for all $t_i \in T_i$,

$$s_i(t_i) \in \arg\max_{a_i \in A_i} \mathop{\mathrm{E}}_{p(t_{-i}|t_i)} \left[\mathop{\mathrm{E}}_{f(a_i, s_{-i}(t_{-i}))(z)} g(z, t) | t_i \right].$$

But $f(s_i(t_i), s_{-i}(t_{-i})) = \xi(t_i, t_{-i})$ for all $(t_i, t_{-i}) \in T$. Moreover, for all $i \in I$, $s_i(t_i) \subseteq A_i$, so for all $t_i \in T_i$,

$$t_i \in \arg\max_{\hat{t}_i \in T_i} \mathop{\mathrm{E}}_{p(t_{-i}|\hat{t}_i)} \left[\mathop{\mathrm{E}}_{\xi(\hat{t}_i, t_{-i})(z)} g(z, t) | t_i \right].$$

So, truth-telling is a Bayes-Nash equilibrium of the induced direct revelation game. □