

MECHANISM DESIGN: SOME DEFINITIONS AND RESULTS

Mechanism design is the science of designing rules of a game to achieve a specific outcome, even though each participant may be self-interested. This is done by setting up a structure in which each player has an incentive to behave as the designer intends.

Mechanism designers commonly try to achieve the following basic outcomes: truthfulness, individual rationality, budget balance, and social welfare. However, it is impossible to guarantee optimal results for all four outcomes simultaneously in many situations.

DEFINITIONS

Let s_i be player i 's strategy, S_i her strategy set, and Θ_i her type space.

Mechanism: A *mechanism*, $\Gamma = (s_1, \dots, s_I, g(\cdot))$, is a collection of I strategy sets, (s_1, \dots, s_I) , and an outcome function, $g: s_1 \times \dots \times s_I \rightarrow X$, where X is the outcome set

Social choice function: A *social choice function* is a function $f: \Theta_1 \times \dots \times \Theta_I \rightarrow X$ that, for each possible profile of the agents' types $(\theta_1 \times \dots \times \theta_I)$, assigns a collective choice $f(\theta_1, \dots, \theta_I) \in X$

Implementation: The mechanism Γ *implements* social choice function $f(\cdot)$ (in Bayesian Nash equilibrium) if there is an equilibrium strategy profile, σ^* , that induces a Bayesian Nash equilibrium in the game associated with Γ s.t. (such that) $g(\sigma_1^*(\theta_1), \dots, \sigma_I^*(\theta_I)) = f(\theta_1, \dots, \theta_I) \forall$ (for all) $(\theta_1, \dots, \theta_I) \in \Theta_1 \times \dots \times \Theta_I$

Direct Revelation Mechanism: A *direct revelation mechanism* is a mechanism in which $S_i = \Theta_i \forall i$ and $g(\theta) = f(\theta) \forall \theta \in \Theta_1 \times \dots \times \Theta_I$

Truthful Implementation: The social choice function is *truthfully implementable* if truth telling by each agent constitutes an equilibrium of Γ

RESULTS

The Revelation Principle For Bayesian Nash Equilibrium: Suppose \exists (there exists) a mechanism, $\Gamma = (s_1, \dots, s_I, g(\cdot))$, that implements the social choice function $f(\cdot)$ in Bayesian Nash equilibrium. Then $f(\cdot)$ is truthfully implementable in Bayesian Nash equilibrium.

The intuition of the Revelation Principle is simple: If a player protects her private information using a particular strategy, manipulating the rules of the game to exogenously provide the same protection of her private information **encourages the player to reveal her private information**.

Is it possible to design a mechanism such that truth telling is optimal?

Vickrey-Clarke-Groves Mechanisms: A *Vickrey-Clarke-Groves (VCG) mechanism* is a direct mechanism in which (i) the 'total-value-maximising' outcome is always chosen, (ii) each player pays an individual-specific payment, p_i , equal to the opportunity cost of their presence, and (iii) payments are equal to the sum of the individual specific payment and a common payment, h_{-i} , which is the payment made if they report they are indifferent.

The VCG mechanism maintains the incentive to bid truthfully when multiple identical goods are being auctioned (in contrast to, say, a uniform price auction). It demands **each player in the auction pays the opportunity cost that their presence introduces to all the other players**. However, the VCG mechanism is vulnerable to collusion by losing bidders, to shill bidders, and it does not necessarily maximise seller revenues. It is also only generally functional for quasi-linear preferences.

Which mechanisms provide the same *expected* revenue?

Revenue Equivalence: Any allocation mechanism in which (i) the bidder with the highest value always wins, (ii) the bidder with the lowest possible value expects zero surplus, (iii) where all bidders are risk neutral, and (iv) drawn from a strictly increasing and atomless distribution, will lead to the same expected revenue for the seller (and each type of agent can expect the same surplus across auction types).

Although we will not present a formal proof here, the intuition is as follows: By the Revelation Principle, we know that **the social choice function that is (indirectly) implemented by the equilibrium of any auction procedure must be Bayesian incentive compatible**. Thus, we can establish the result by showing that if two Bayesian incentive compatible social choice functions in this auction setting have the same transfer functions ($y_1(\theta), \dots, y_I(\theta)$) and the same payoff values, ($u_1(\theta_1), \dots, u_I(\theta_I)$), then **they generate the same expected revenue for the seller**. To show this we can

derive an expression for the seller's expected revenue from an arbitrary Bayesian incentive compatible mechanism.

Note that the proof of revenue equivalence can be derived *directly* from the incentive compatibility constraint.

If we can design truth-telling mechanisms all of which yield equal revenues, why does mechanism design matter? The conditions under which the VCG mechanisms function and under which the revenue equivalence result holds are restrictive. In many cases, we will not even be able to guarantee an efficient allocation using *any* mechanism.

Myerson-Satterthwaite Impossibility Theorem: Consider a bilateral trade setting in which the buyer and the seller are risk neutral, the valuations θ_1 and θ_2 are independently drawn from the intervals $[\underline{\theta}_1, \bar{\theta}_1] \subset \mathfrak{R}$ and $[\underline{\theta}_2, \bar{\theta}_2] \subset \mathfrak{R}$ with strictly positive densities and $(\underline{\theta}_1, \bar{\theta}_1) \cap (\underline{\theta}_2, \bar{\theta}_2) \neq \emptyset$. Then there is *no* Bayesian incentive compatible social choice function that is ex post efficient and gives every buyer type and every seller type non-negative expected gains from participation.

Again, no formal proof is provided. The intuition is as follows: The gains of trade in a private value auction cannot everywhere be positive since **the Bayesian Nash equilibrium is for parties to report untruthfully so to increase their expected payoff**. By doing this they create regions of types in which the individual rationality constraint is violated for some participant.

Mechanisms fail to guarantee efficiency due to the presence of private information. Once signals held by various agents (their private information) is correlated, there may be scope for regaining an efficient outcome, but is not assured.

Cremer and McLean (1988): If the signals held by various agents are correlated, it is possible to design transfer devices whereby the belief held by agent i is elicited for free. When j 's belief about i 's signal completely determines the signal held by i , a welfare-maximising allocation can be (approximately) implemented.

Jehiel-Moldovanu (2001): In general, in multi-object auctions with heterogeneous objects and common rather than private values, no mechanism whatsoever can allocate the goods efficiently if either complementarities, substitutability, or allocative externalities are present.

MATHEMATICAL NOTE

In his 'Putting Auction Theory to Work', Paul Milgrom argues that "Many of the key results of mechanism design theory can be derived from the envelope theorem and stated as a restriction on a derivative or a restriction on an integral." To approach mechanism design from this perspective, it is important to understand the Integral Form Envelope Theorem, and its applications.

The Integral Form Envelope Theorem (Milgrom): Suppose that $u(x, \cdot): [0, 1] \rightarrow \mathfrak{R}$ has the properties that:

- i) \exists a real valued function $u_2(x, t)$ s.t. $\forall x \in X, [a, b] \in [0, 1], u(x, b) - u(x, a) = \int_a^b u_2(x, s) ds$
- ii) \exists an integrable function $b: [0, 1] \rightarrow \mathfrak{R}$ (that is $\int b(s) ds < \infty$) s.t. $|u_2(x, t)| \leq b(t) \forall x \in X$ and almost all $t \in [0, 1]$
- iii) $X^*(t) \equiv \operatorname{argmax}_{x \in X} u(x, t) \neq \emptyset$ for almost all $t \in [0, 1]$

Then for any selection $x^*(t)$ from $X^*(t)$, $V(t) = u(x^*(t), t) = u(x^*(0), 0) + \int u_2(x^*(s), s) ds$

Applying the Integral Form Envelope Theorem to the special case of quasi-linear preferences and Bayes-Nash equilibria yields:

Myerson's lemma: Consider a standard independent private values model, and suppose that σ is a Bayes-Nash equilibrium of the game corresponding to $(\Omega, N, S, \omega, [0, 1]^N, v, \pi)$ (standard definitions apply; see book for details) with full performance (x, p) . Then the expected payoff satisfy $V^i(\tau) = V^i(0) + \int E^i[z(\tau) | t^i=s] (dv^i/ds) ds$.

Prepared for MICT PhD course, Department of Economics, UCL
by Daniel Rogger, UCL PhD candidate