Mechanism Design and the Revelation Principle

CS 1951k/2951z

2020-02-5

First, we review the mechanism design formalism, in terms of games of incomplete information. Then, we introduce the notion of an indirect mechanism, presenting the English and Dutch auctions as examples. Finally, we cover the Revelation Principle,[1] which can transform any mechanism into a direct one, ensuring incentive compatibility.

[1] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981; and Roger B Myerson. Incentive compatibility and the bargaining problem. *Econometrica: journal of the Econometric Society*, pages 61–73, 1979

1 *Mechanism Design Framework*

Mechanism design has been referred to as the engineering branch of game theory. It is concerned with designing mechanisms (i.e., games) such that the outcomes that arise when the games are played by rational agents (i.e., the equilibria) achieve some desiderata. The mechanism design framework thus consists of three parts: the mechanism formalism; solution concepts, or equilibria; and possible objectives.

Mechanisms The interaction between the designer of a mechanism and its participants can be modeled as a multi-stage game. The designer moves first by selecting a mechanism. The participants observe the mechanism, and move thereafter. We restrict our present attention to a two-stage game, in which the participants play a simultaneous-move (i.e., one-shot) game in the second stage.

After the designer announces their choice of mechanism, the agents face a **game of incomplete information**. Formally, such a game is denoted by $\Gamma = \langle [n], \{A_i\}_{i\in n}, \{T_i\}_{i\in n}, \Omega, g, \{u_i\}_{i\in n}\rangle$, where $[n] = \{1, \ldots, n\}$ is the set of players (or agents), A_i is the set of actions available to player $i \in [n]$, with $A = \prod_{i=1}^{n} A_i$ as the joint action space; and T_i is the set of types (private information) available to player $i \in [n]$, with $T = \prod_{i=1}^{n} T_i$ as the joint type space. Additionally, a joint distribution F over types is assumed to be common knowledge, known to both the players and the designer.

We define a **strategy** of a player i as a function $s_i : T_i \to A_i$, and use $\mathbf{s_t}$ to denote the vector $(s_1(t_1), \ldots, s_n(t_n))$. The function $g : A \to \Omega$ maps a joint action profile into a space Ω of possible outcomes; that is, $g(\mathbf{s}(\mathbf{t}))$ is the **outcome** when player i of type t_i plays strategy s_i and the remaining players of type $\mathbf{t}_{-i}$ play strategy $\mathbf{s}_{-i}$.[2] Finally, player i's **utility** $u_i : \Omega \times T \to \mathbb{R}$ depends on both the outcome of the game and (in general) all players' types.

[2] For example, the outcome function of an auction maps a profile of bids to an outcome, which comprises an allocation and a payment rule.

Equilibria Given a mechanism Γ, a solution takes the form of a joint strategy profile $\mathbf{s}^*$ that the players are predicted to play. Dominant-

strategy or ex-post Nash when they exist, and otherwise the Bayes-Nash equilibrium solution concept, are usually applied to solve a game. We define these concepts in our current formalism, presently.

Definition 1.1. A strategy vector $\mathbf{s} = (s_i, \mathbf{s}_{-i}) \in S$ is a **Bayes-Nash equilibrium** in a mechanism Γ if no player can increase their ex-ante expected utility by unilaterally changing their strategy:

$$\mathop{\mathbb{E}}_{\mathbf{t}\sim F}\left[u_i(g(\mathbf{s}(\mathbf{t})); \mathbf{t})\right] \geq \mathop{\mathbb{E}}_{\mathbf{t}\sim F}\left[u_i(g(s_i'(\mathbf{t}_i), \mathbf{s}_{-i}(\mathbf{t}_{-i})); \mathbf{t})\right], \quad \forall i \in [n], \forall s_i' \in S_i. \tag{1}$$

Definition 1.2. A strategy vector $\mathbf{s} = (s_i, \mathbf{s}_{-i}) \in S$ is an **ex-post Nash equilibrium** in a mechanism Γ if no player can increase their ex-post expected utility by unilaterally changing their strategy:

$$u_i(g(\mathbf{s}(\mathbf{t})); \mathbf{t}) \geq u_i(g(s_i'(\mathbf{t}_i), \mathbf{s}_{-i}(\mathbf{t}_{-i})); \mathbf{t}), \quad \forall i \in [n], \forall s_i' \in S_i. \tag{2}$$

Definition 1.3. A strategy s_i for player $i \in [n]$ is **dominant** if it is (weakly) optimal, regardless of the other players' actions *and types*: i.e.,

$$u_i(g(s_i(t_i), \mathbf{a}_{-i}); \mathbf{t}) \geq u_i(g(s_i'(t_i), \mathbf{a}_{-i}); \mathbf{t}), \quad \forall s_i' \in S_i, \forall \mathbf{a}_{-i} \in A_{-i}, \forall \mathbf{t} \in T. \tag{3}$$

A strategy vector $\mathbf{s} \in S$ is a **dominant strategy equilibrium** (DSE) if all players play dominant strategies. At a DSE, s_i is (weakly) optimal for player i, regardless of what other players know or do.

Desiderata The criteria by which a mechanism should be evaluated are highly domain dependent. Still, it suffices to mention two alternative means of evaluation. First, it may be the designer's goal to implement a **social choice function** $f : T \to \Omega$, in which case a design could be deemed successful if $g(\mathbf{s}^*(\mathbf{t})) \in f(\mathbf{t})$, for all type profiles $\mathbf{t}$, or partially successful to the extent $g(\mathbf{s}^*(\mathbf{t}))$ intersects $f(\mathbf{t})$. Alternatively, the designer may seek to maximize a numeric function of the solution to the induced game,[3] such as a **social welfare** function $\mathbb{E}_{\mathbf{t}\sim F}\left[W(g(\mathbf{s}^*(\mathbf{t}))\right]$, or a **revenue** function $\mathbb{E}_{\mathbf{t}\sim F}\left[R(g(\mathbf{s}^*(\mathbf{t}))\right]$.[4]

In much of the mechanism design literature, the problem is greatly simplified by reliance on the revelation principle, which argues that the strategic outcome of any mechanism can be replicated by a direct mechanism (i.e., a mechanism in which agents simply report their types). Direct (vs. indirect) mechanisms, and the revelation principle, are the subjects of the remainder of this lecture.

[3] There is an implicit equilibrium selection function that cannot be overlooked; in case the predicted solution is not unique, welfare/revenue could, for example, be computed in either the worst-case or the average case.

[4] Note that within the MD framework these goals could easily be relaxed so that, for example, welfare/revenue is maximized or exceeds some threshold value only with high probability.

2 *Indirect Mechanisms*

Definition 2.1. A **direct mechanism** is one in which the space of possible actions is equal to the space of possible types.

Examples of direct mechanisms include first-, second-, and third-price auctions (assuming the space of possible bids is restricted to the space of possible types). All other mechanisms are called **indirect**. Examples of indirect mechanisms include the **English**, or **ascending**, auction, and the **Dutch**, or **descending** auction.

Example 2.2. The **English auction** consists of a number of rounds. On round $k = 1, 2, \ldots$, the auctioneer offers the good at price $p = k\epsilon$, for some small $\epsilon > 0$, asking all bidders if they are interested in the good at that price. The auction continues so long as more than one bidder is interested. The auction terminates, say at round t, when one or fewer bidders remain interested. If there is one interested bidder at round t, then she wins, paying $t\epsilon$; if there are no interested bidders then a winner is selected at random from the set of interested bidders during round $t - 1$. This winner pays $(t-1)\epsilon$.

In this auction, actions consist of t binary answers to queries "Would you like the good at price p?". In practice, it may be easier for bidders to answer queries like this one, rather than articulate an exact value for a good, which may be part of the reason why this auction format is so widely used.[5]

[5] This is not to say that second-price (i.e., Vickrey) auctions are not used in practice. On the contrary, stamp auctioneers used this mechanism to sell stamps by mail as early as the late 1800s, before Vickrey was born!

Example 2.3. The **Dutch auction** also consists of a number of rounds, but in this case, the price p is initialized high enough so that no bidders are interested. The price is then decremented successively by ϵ until a bidder (or a set of bidders) declares their interest in the item. That bidder is declared the winner (or a tie is broken randomly); the winner receives the item and pays the final price.

Not surprisingly, Dutch auctions are popular in the Netherlands.

3 *Incentive Compatibility*

In direct mechanisms, the equilibrium notions of BNE, EPNE, and DSE correspond to the analogous notions of **Bayesian incentive compatibility** (BIC), **ex-post incentive compatibility** (EPIC), and **dominant strategy incentive compatibility** (DSIC).

Definition 3.1. A (direct) mechanism is BIC iff truthtelling is a Bayes-Nash equilibrium: i.e.,

$$\mathop{\mathbb{E}}_{\mathbf{t}\sim F}\left[u_i(g(t_i, \mathbf{t}_{-i}); \mathbf{t})\right] \geq \mathop{\mathbb{E}}_{\mathbf{t}\sim F}\left[u_i(g(t_i', \mathbf{t}_{-i}); \mathbf{t})\right], \quad \forall i \in [n], \forall t_i' \in T_i. \tag{4}$$

Definition 3.2. A (direct) mechanism is EPIC iff truthtelling is a ex-post Nash equilibrium: i.e.,

$$u_i(g(t_i, \mathbf{t}_{-i}); \mathbf{t}) \geq u_i(g(t_i', \mathbf{t}_{-i}); \mathbf{t}), \quad \forall i \in [n], \forall t_i' \in T_i. \tag{5}$$

Definition 3.3. A (direct) mechanism is DSIC iff truth-telling is a dominant strategy equilibrium: i.e.,

$$u_i(g(t_i, \mathbf{t}_{-i}); \mathbf{t}) \geq u_i(g(t_i', \mathbf{t}_{-i}); \mathbf{t}), \quad \forall i \in [n], \forall t_i' \in T_i, \forall \mathbf{t}_{-i} \in T_{-i}. \quad (6)$$

Remark 3.4. In a direct mechanism, EPIC is equivalent to DSIC.

Proof. DSIC implies EPIC, so it suffices to show that EPIC also implies DSIC. Let M be an EPIC, one-shot mechanism such that the space of possible actions equals the space of possible types.

Since M is EPIC, for all bidders $i \in N$ and for all (true) type profiles $\mathbf{t} \in T$, truthful bidding satisfies

$$u_i(t_i, \mathbf{t}_{-i}) \geq u_i(t_i', \mathbf{t}_{-i}), \quad \forall t_i' \in T_i.$$

Our goal is to show that M is also DSIC: i.e., for all bidders $i \in N$ and for all (true) types $t_i \in T_i$,

$$u_i(t_i, \mathbf{b}_{-i}) \geq u_i(t_i', \mathbf{t}_{-i}), \quad \forall t_i' \in T_i, \forall \mathbf{b}_{-i} \in B_{-i}.$$

But since $B_i = T_i$ by assumption (i.e., the mechanism is direct), it suffices to show: for all bidders $i \in N$ and for all (true) types $t_i \in T_i$,

$$u_i(t_i, \mathbf{t}_{-i}) \geq u_i(t_i', \mathbf{t}_{-i}), \quad \forall t_i' \in T_i, \forall \mathbf{t}_{-i} \in T_{-i}.$$

Fix a bidder i and their true type t_i. For two arbitrary type profiles $\mathbf{t}'_{-i}, \mathbf{t}''_{-i} \in T$, since M is EPIC,

$$u_i(t_i, \mathbf{t}'_{-i}) \geq u_i(t_i', \mathbf{t}_{-i}), \quad \forall t_i' \in T_i.$$

$$u_i(t_i, \mathbf{t}''_{-i}) \geq u_i(t_i', \mathbf{t}_{-i}), \quad \forall t_i' \in T_i.$$

Hence, truthful bidding is a best response for bidder i, assuming others are also bidding truthfully, but regardless their type profiles. In other words, truthful bidding is a best response for bidder i, for *all* other-agent type profiles: i.e., for all bidders $i \in [n]$,

$$u_i(t_i, \mathbf{t}_{-i}) \geq u_i(t_i', \mathbf{t}_{-i}), \quad \forall i \in [n], \forall t_i' \in T_i, \forall \mathbf{t}_{-i} \in T_{-i}.$$

Since bidder i was arbitrary, truthful bidding is a dominant-strategy equilibrium (i.e., it is a best response for all bidders). □

4 *The Revelation Principle*

Theorem 4.1 (Revelation Principle)**.** *Given a mechanism M for which there exists an equilibrium strategy profile $\mathbf{s}$, we can construct an equivalent direct mechanism M^* in which truthtelling is likewise an equilibrium.*

By *equivalence* of two mechanisms M and M', given two corresponding strategy profiles $\mathbf{s}$ and $\mathbf{s}'$, we mean that the same type profile $\mathbf{t}$ yields the exact same outcome in each mechanism: i.e., $g(\mathbf{s}(\mathbf{t})) = g'(\mathbf{s}'(\mathbf{t}))$, for all $\mathbf{t} \in T$. If M' is assumed to be a direct mechanism, then equivalence means $g(\mathbf{s}(\mathbf{t})) = g'(\mathbf{t})$, for all $\mathbf{t} \in T$.

By the revelation principle, if there exists a Bayes-Nash equilibrium in mechanism M, we can produce a BIC mechanism M^*. Furthermore, when bidders play according to these respective equilibria, the outcomes in M and M^* are identical. Likewise, if there exists an ex-post Nash or dominant strategy equilibrium in mechanism M, then we can produce an EPIC or DSIC mechanism M^*, respectively, so that when bidders play according to these respective equilibria, the outcomes in M and M^* are identical.

One further point before we prove the theorem: The revelation principle does not only apply to indirect mechanisms. It applies to *all* mechanisms, and can therefore be applied, for example, to a first-, second-, third-, etc.-price, or even an all-pay auction, to convert them, together with their BNE, into direct, BIC mechanisms.

Proof. Given a (possibly indirect) mechanism M, together with an equilibrium $\mathbf{s}$, we construct a direct mechanism M^* with the corresponding truth-telling equilibrium as follows:

- Elicit all players' types $t_1, t_2, \ldots, t_n$.
- Simulate M by performing i's equilibrium action on his behalf, assuming i's type is t_i.
- Return the outcome produced by M.

We can think of the construction (see Figure 1) as a machine that first asks all players for their types, and then runs the equilibrium strategy on their behalf. Each agent reports a (possibly false) type t_i to the direct mechanism M^*, which simulates $s_i(t_i)$. The agents are incentivized to report their true types, because by doing so, the outcome of M^* is the same as the outcome of M, which by assumption is an equilibrium.

More specifically, if player i lies to the robot (and if they are the only one lying), the robot will run everyone else's equilibrium strategy based on their true types, except for player i's. The outcome will be the exact same outcome as running M, assuming i deviates from its equilibrium strategy. But this deviation was not in i's best interest in M, so likewise, it is not in i's best interest in M^*. □

Example 4.2. Consider a modified second-price auction M in which the winner pays *twice* the second-highest bidder's bid. This auction has a DSE in which the bidders bid half their values. (Why?)

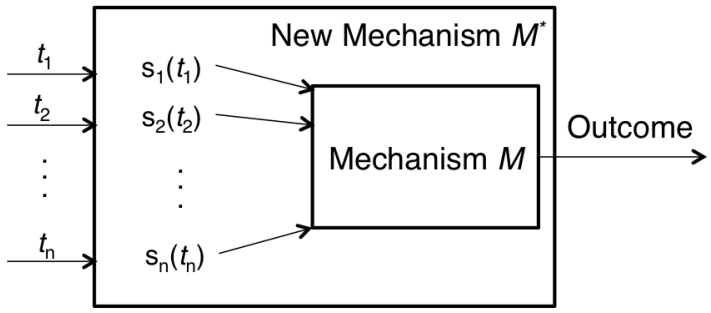

Figure 1: The Revelation Principle

Given M and the aforementioned DSE, the mechanism M^* constructed according to the Revelation Principle, works as follows:

- Elicit all bidders' values $v_1, v_2, \ldots, v_n$.
- For each bidder i, submit the sealed bid $v_i/2$.
- Return the outcome produced by original auction (M).

The mechanism M^* has the following three properties:

1. DSIC: Truth-telling is a dominant-strategy equilibrium.
2. The highest bidder—who, by 1, has the highest value—wins the auction.
3. The winner pays the second-highest bid, which by 1 is also the second-highest value.

Therefore, M^* is functionally identical to the second-price auction.

References

[1] Roger B Myerson. Incentive compatibility and the bargaining problem. *Econometrica: journal of the Econometric Society*, pages 61–73, 1979.

[2] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.