Chapter 2

Equilibrium

Equilibrium attempts to capture what happens in a game when players behave strategically. This is a central concept to these notes as in mechanism design we are optimizing over games to find the games with good equilibria. Here, we review the most fundamental notions of equilibrium. Note that these notions are all static notions in that players are assumed to understand the game and will play once in the game. While this foreknowledge is certainly questionable, some justification can be derived from imagining the game in a dynamic setting where players can learn from past play. Readers should look elsewhere for formal justifications.

This chapter reviews equilibrium in both complete and incomplete information games. As games of incomplete information are the most central to mechanism design special attention will be paid to these. In particular, we will characterize equilibrium when the private information of each agent is single-dimensional and corresponds, for instance, to a value for receiving a good or service. We will show that auctions with the same equilibrium outcome have the same expected revenue. Using this so-called *revenue equivalence* we will describe how to solve for the equilibrium strategies of standard auctions in symmetric setting.

Emphasis of placed on demonstrating the central theories of equilibrium and not on providing the most comprehensive or general results. For that readers are recommended to consult a game theory textbook.

2.1 Complete Information Games

In games of compete information all players are assumed to know precisely the payoff structure of all other players for all possible outcomes of the game. A classic example of such a game is the *prisoner's dilemma*, the story for which is as follows. Two prisoners who have jointly committed a crime, are being interrogated in separate quarters. Unfortunately, the interrogators are unable to prosecute either prisoner without a confession. Each prisoner is offered the following deal: If they confess and their accomplice does not, they will be released and their accomplice will serve the full sentence of ten years in prison. If they both confess, they will share the sentence and serve five years each. If neither confesses, they will both be prosecuted for a minimal offense and each receive a year of prison. This story can be expressed as the following bimatrix game where entry (a, b) represents row player's payoff a and column player's payoff b.

	silent	confess
silent	(-1,-1)	(-10,0)
confess	(0,-10)	(-5,-5)

A simple thought experiment suggests what will happen in this game. Suppose the row player is silent. What should the column player do? Remaining silent as well results in one year of prison while confessing results in immediate release. Clearly confessing is better. Now suppose that the column player confesses. Now what should the row player do? Remaining silent results in ten years of prison while confessing as well results in only five. Clearly confessing is better. In other words, no matter what the column player does, the row player is better of by confessing. The prisoners dilemma is hardly a dilemma at all: the strategy profile (confess, confess) is a dominant strategy equilibrium.

Definition 2.1 A *dominant strategy equilibrium* (DSE) in a complete information game is a strategy profile in which each player's strategy is as least as good as all other strategies regardless of the actions of any other players.

DSE is a strong notion of equibrium and is therefore unsurprisingly rare. For an equilibrium notion to be complete it should identify equilibrium in every game. Another well studied game is *chicken*. The story for chicken is as follows, James Dean and Buzz (in the movie *Rebel without a Cause*) face off at opposite ends of the street. On the signal they race their cars on a collision course towards each other. The options each have are to swerve or to stay their course. Clearly if they both stay their course they crash. If they both swerve (opposite directions) they escape with their lives but the match is a draw. Finally, if one swerves and the other stays, the one that stays is the victor and the other the loses.[1] A reasonable bimatrix game depicting this is the following.

	stay	swerve
stay	(-10,-10)	(1,-1)
swerve	(-1,1)	(0,0)

Again, a simple thought experiment suggests how the players might play. Suppose James Dean is going to stay, what should Buzz do? If Buzz stays they crash and Buzz's payoff is -10, but if Buzz swerves his payoff is only -1. Clearly, of these two options Buzz prefers to swerve. Suppose now that Buzz is going to stay, what should James Dean do? If James Dean stays he wins and his payoff is 1, but if he swerves its a draw and his payoff is zero. Clearly, of these two options James Dean prefers to stay. What we have shown is that the strategy profile (stay, swerve) is a mutual best response, a.k.a., a Nash equilibrium.

[1] The actual chicken game depicted in *Rebel without a Cause* is slightly different from the one described here.

Definition 2.2 A *Nash equilibrium* in a game of complete information is a strategy profile where each players strategy is a best response to the strategies of the other players as given by the strategy profile.

2.2 Incomplete Information Games

Now we turn to the case where the payoff structure of the game is not completely known. In the setting of interest to us, we will assume that each agent has some private information and this information affects the payoffs of the game. We denote by t_i the type of agent i and by $\mathbf{t} = (t_1, \ldots, t_n)$ the profile of types for the n agents in the game.

A *strategy* in a game of incomplete information is a function that maps an agent's type to any of the agent's possible actions in the game. We will denote by $s_i(\cdot)$ the strategy of agent i and $\mathbf{s} = (s_1, \ldots, s_n)$ a *strategy profile*. For example, we already saw reasonable strategies in the English auction were $s_i(v_i) =$ "drop out when the price exceeds v_i" and strategies in the second price auction were $s_i(v_i) =$ "bid $b_i = v_i$." We refer to this latter strategy as *truthtelling*. Both of these strategy profiles are in *dominant strategy equilibrium* for their respective games.

Definition 2.3 A *dominant strategy equilibrium* (DSE) is a strategy profile $\mathbf{s}$ such that for all i, v_i, and $\mathbf{b}_{-i}$ (where $\mathbf{b}_{-i}$ generically refers to the actions of all players but i), agent i's utility is maximized by following strategy $s_i(v_i)$.

Notice that aside from strategies being defined as a map from types to actions, this definition of DSE is identical to the definition of DSE for games of complete information.

2.3 Bayes-Nash Equilibrium

Naturally, many games of incomplete information do not have dominant strategy equilibria. Therefore, we will also need to generalize Nash equilibrium to this setting. Recall that equilibrium is a property of a strategy profile. It is in equilibrium if each agent does not want to change their strategy given the other agents' strategies. Meaning, for an agent i, we want to the fix other agent strategies and let i optimize its strategy (meaning compute their best response for all possible types t_i). This is an ill specified optimization as just knowing the other agents' strategies is not enough to compute a best response. Instead, i's best response depends on i's beliefs on the types of the other agents. The standard economic treatment addresses this by making the *common prior assumption*.

Definition 2.4 The agent types $\mathbf{t}$ are drawn at random from a *prior distribution* $\mathbf{F}$ (a probability distribution over type profiles) and this prior distribution is *common knowledge*.

Notice that a prior $\mathbf{F}$ and strategies $\mathbf{s}$ induces a distribution over the actions of each of the agents. With such a distribution over actions, the problem each agent faces of optimizing their own action is fully specified.

Notice that the distribution $\mathbf{F}$ over $\mathbf{t}$ may generally be correlated. Which means that an agent with knowledge of their own type must do *Bayesian updating* to determine the distribution over the types of the remaining bidders. We denote this conditional distribution as $\mathbf{F}_{-i}|_{t_i}$

Definition 2.5 A *Bayes-Nash equilibrium* (BNE) for a game G and common prior $\mathbf{F}$ is a strategy profile $\mathbf{s}$ such that for all i and t_i, $s_i(t_i)$ is a best response when other agents play $\mathbf{s}_{-i}(\mathbf{t}_{-i})$ when $\mathbf{t}_{-i} \sim \mathbf{F}_{-i}|_{t_i}$.

To illustrate BNE, consider using the first price auction to sell a single item to one of two agents, each with valuation drawn independently and identically from $U[0,1]$, i.e., $\mathbf{F} = F_1 \times F_2$ with $F_i(z) = \mathbf{Pr}[v_i < z] = z$. Here each agent's type is their valuations. We will calculate the BNE of this game by the "guess and verify" technique. First, we quess that there is a symmetric BNE with $s_i(z) = z/2$ for $i \in \{1,2\}$. Second, we calculate agent 1's expected utility with value v_1 and bid b_1 under the standard assumption that the agent's utility u_i is their value less their payment (when they win).

$$\begin{aligned}
\mathbf{E}[u_i] &= (v_1 - b_1) \times \mathbf{Pr}[1 \text{ wins}] \\
&= (v_1 - b_1) \times \mathbf{Pr}[b_1 \geq b_2] \\
&= (v_1 - b_1) \times \mathbf{Pr}[b_1 \geq v_2/2] \\
&= (v_1 - b_1) \times \mathbf{Pr}[2b_1 \geq v_2] \\
&= (v_1 - b_1) \times \mathbf{Pr}[F_2(2b_1)] \\
&= (v_1 - b_1) \times 2b_1 \\
&= 2v_1 b_1 - 2b_1^2.
\end{aligned}$$

Thrid, we optimize agent 1's bid. Agent 1 with value v_1 should maximize this quantity as a function of b_1, and to do so, can differentiate the function and set its derivative equal to zero. The result is $\frac{d}{db_1}(2v_1b_1 - 2b_1^2) = 2v_1 - 4b_1 = 0$ and we can conclude that the optimal $b_1 = v_1/2$. This proves that agent 1 should bid as prescribed if agent 2 does; and vice versa. Finally, we conclude that the guessed strategy profile is in BNE.

2.4 Single-dimensional Games

We will focus on a conceptually simple class of games that is relevant to the auction problems we have already discussed. In this class, each agent's private type is their value for receiving an abstract service. A game has an outcome $\mathbf{x} = (x_1, \ldots, x_n)$ and payments $\mathbf{p} = (p_1, \ldots, p_n)$ where x_i is an indicator for whether agent i indeed received their desired service, i.e., $x_i = 1$ if i is served and 0 otherwise. Price p_i will denote the payment i makes to the mechanism. An agent's valuation can be positive or negative and an agent's payment can be positive or negative. An agent's utility is linear in their valuation and payment and specified by $u_i = v_i x_i - p_i$. Agents are risk-neutral expected utility maximizers.

A game G maps actions $\mathbf{b}$ of agents to an outcome and payment. Formally we will specify these outcomes and payments as:

- $x_i^G(\mathbf{b})$ = outcome to i when actions are $\mathbf{b}$, and
- $p_i^G(\mathbf{b})$ = payment from i when actions are $\mathbf{b}$.

Given a game G and a strategy profile $\mathbf{s}$ we can express the outcome and payments of the game as a function of the valuation profile. From the point of view of analysis this description of the the game outcome is much more relevant. Define

- $x_i(\mathbf{v}) = x_i^G(\mathbf{s}(\mathbf{v}))$, and
- $p_i(\mathbf{v}) = p_i^G(\mathbf{s}(\mathbf{v}))$.

We refer to the former as the *allocation rule* and the latter as the *payment rule* for G and $\mathbf{s}$ (implicit). Finally, for G, $\mathbf{s}$, and $\mathbf{F}$ taken implicitly we can specify the allocation and payment rules from i's perspective when the other agent's values are drawn from the distribution. For $\mathbf{v} \sim F$,

- $x_i(v_i) = \mathbf{Pr}[x_i(v_i) = 1 \mid v_i] = \mathbf{E}[x_i(\mathbf{v}) \mid v_i]$, and
- $p_i(v_i) = \mathbf{E}[p_i(\mathbf{v}) \mid v_i]$.

With linearity of expectation we can combine these with the agents utility function to write

- $u_i(v_i) = v_i x_i(v_i) - p_i(v_i)$.

2.5 Characterization of Bayes-Nash equilibrium

We now discuss what Bayes-Nash equilibria look like. For instance, suppose we were given G, $\mathbf{s}$, and $\mathbf{F}$. From this we can calculate the allocation and payment rules $x_i(v_i)$ and $p_i(v_i)$ of each agent. We want to succinctly describe properties of these allocation and payment rules that characterize BNE.

Theorem 2.6 *G, $\mathbf{s}$, and $\mathbf{F}$ are in BNE if and only if for all i,*

1. *(monotonicity) $x_i(v_i)$ is monotone non-decreasing, and*
2. *(payment identity) $p_i(v_i) = v_i x_i(v_i) - \int_0^{v_i} x_i(z)dz + p_i(0)$,*

where often $p_i(0) = 0$.[2]

Proof: We break this proof into three pieces. First we show, by picture, that the game is in BNE if the characterization holds. Next we will prove that a game is in BNE only if

[2]While this characterization is quite general, we will only prove it for valuation distributions with continuously differentiable support on $[0, \infty)$ where the intuition behind the result is readily apparent. We will also assume that each strategy in strategy profile maps values *onto* actions. This restriction is necessary only for proving the "if" direction where it does not weaken the applicability of the theorem in settings we will consider later.

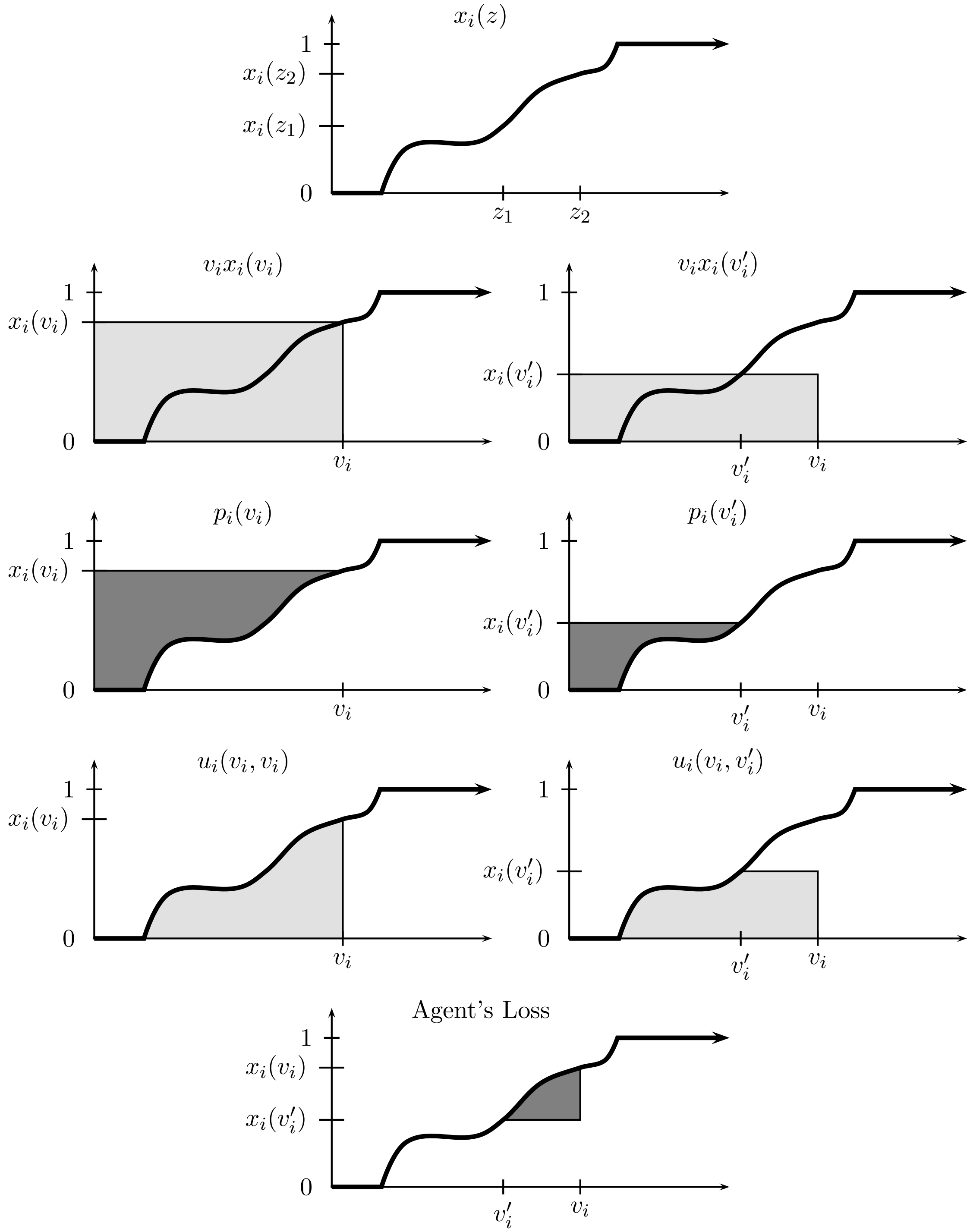

Figure 2.1: The left column shows the valuation, payment, and utility of agent i playing action $s_i(v_i = z_2)$. The right column shows the same for agent i playing action $s_i(v_i' = z_1)$. The final diagram shows the difference between the agent's utility for these strategies. Monotonicity implies this difference is non-negative.

the monotonicity condition holds. Finally we will prove that a game is in BNE only if the payment identity holds.

Note that if agent i with value v_i deviates from the equilibrium and takes action $s_i(v_i')$ instead of $s_i(v_i)$ then agent i will receive outcome and payment $x_i(v_i')$ and $p_i(v_i')$. This motivates the definition,

$$u_i(v_i, v_i') = v_i x_i(v_i') - p_i(v_i'),$$

which corresponds to this deviating agent's utility. Notice that $\mathbf{s}$ is in equilibrium if and only if for all i, v_i, and v_i', $u_i(v_i, v_i) \geq u_i(v_i, v_i')$, i.e., if our deviating agent derives no increased utility. (The "if" direction here follows from the assumption that strategies map values onto actions.)

1. G, $\mathbf{s}$, and $\mathbf{F}$ are in BNE if the characterization holds.

   We prove this by picture. Though the formulaic proof is simple, the pictures provide useful intuition. We consider two possible values z_1 and z_2 with $z_1 < z_2$. Supposing $v_i = z_2$ we argue that agent i does not benefit by simulating their strategy in the case their valuation is the lower $v_i' = z_1$. We leave the proof of the opposite, that when $v_i = z_1$ and the agent is considering simulating the higher strategy $v_i' = z_2$, as an exercise for the reader.

   To start with this proof, we assume that $x_i(v_i)$ is monotone and that $p_i(v_i) = v_i x_i(v_i) - \int_0^{v_i} x_i(z)dz$.

   Consider the diagrams in Figure 2.1. The first diagram (top, center) shows $x_i(\cdot)$ which is indeed monotone as per our assumption. The column on the left show the agent's valuation, $v_i x_i(v_i)$; payment, $p_i(v_i)$, and utility, $u_i(v_i) = v_i x_i(v_i) - p_i(v_i)$, assuming that they follow the BNE strategy $s_i(v_i = z_2)$. The column on the right shows the analogous quantities when the agent follows strategy $s_i(v_i' = z_1)$. The final diagram (bottom, center) shows the difference in the bidders utility for these outcome and payments of these two strategies. Note that as the picture shows, the monotonicity of the allocation function implies that this difference is always non-negative. Therefore, it is there is no incentive for the agent to simulate the strategy of a lower value.

   As mentioned, a similar proof shows that the agent has no incentive to simulate their strategy on a higher value. We conclude that agent i with value v_i (weakly) prefers to play the BNE strategy $s_i(v_i)$.

2. G, $\mathbf{s}$, and $\mathbf{F}$ are in BNE only if the allocation rule is monotone.

   If we are in BNE then for all valuations, v_i and v_i', $u_i(v_i, v_i) \geq u_i(v_i, v_i')$. Expanding we require

   $$v_i x_i(v_i) - p_i(v_i) \geq v_i x_i(v_i') - p_i(v_i').$$

   We now consider z_1 and z_2 and take turns setting $v_i = z_1$, $v_i' = z_2$, and $v_i' = z_1$, $v_i = z_2$. This yields the following two inequalities:

   $$v_i = z_2, v_i' = z_1 \Longrightarrow z_2 x_i(z_2) - p_i(z_2) \geq z_2 x_i(z_1) - p_i(z_1), \text{ and}$$
   $$v_i = z_1, v_i' = z_2 \Longrightarrow z_1 x_i(z_1) - p_i(z_1) \geq z_1 x_i(z_2) - p_i(z_2).$$

Adding these inequalities and canceling the payment terms we have,

$$z_2 x_i(z_2) + z_1 x_i(z_1) \geq z_2 x_i(z_1) + z_1 x_i(z_2).$$

Rearranging,

$$(z_2 - z_1)(x_i(z_2) - x_i(z_1)) \geq 0.$$

Since $z_2 - z_1 > 0$ it must be that $x_i(z_2) - x_i(z_1) \geq 0$, i.e., $x_i(\cdot)$ is monotone non-decreasing.

3. G, $\mathbf{s}$, and $\mathbf{F}$ are in BNE only if the payment rule satisfies the payment identity.

   Now we show that the payment rule must satisfy $p_i(v_i) = v_i x_i(v_i) - \int_0^{v_i} x_i(z)dz$. Fix v_i and recall that $u_i(v_i, z) = v_i x_i(z) - p_i(z)$. Let $u_i'(v_i, z)$ be the partial derivative of $u_i(v_i, z)$ with respect to z. Thus, $u_i'(v_i, z) = v_i x_i'(z) - p_i'(z)$, where $x_i'(\cdot)$ and $p_i'(\cdot)$ are the derivatives of $p_i(\cdot)$ and $x_i(\cdot)$, respectively. Since truthfulness implies that $u_i(v_i, z)$ is maximized at $z = v_i$. It must be that

   $$\begin{aligned} u_i'(v_i, v_i) &= v_i x_i'(v_i) - p_i'(v_i) \\ &= 0. \end{aligned}$$

   This formula must hold true for all values of v_i. For remainder of the proof, we treat this identity formulaically. To emphasize this, substitute $z = v_i$:

   $$z x_i'(z) - p_i'(z) = 0.$$

   Solving for $p_i'(z)$ and then integrating both sides of the equality from 0 to v_i we have,

   $$\begin{aligned} p_i'(z) &= z x_i'(z), \text{ so} \\ \int_0^{v_i} p_i'(z)dz &= \int_0^{v_i} z x_i'(z)dz, \text{ so} \\ p_i(v_i) - p_i(0) &= z x_i(z)\big|_0^{v_i} - \int_0^{v_i} x_i(z)dz \\ &= v_i x_i(v_i) - \int_0^{v_i} x_i(z)dz. \end{aligned}$$

   Adding $p_i(0)$ from both sides of the equality, we conclude that the payment identity must hold.

■

As we conclude the proof of the BNE characterization theorem, it is important to note how little we have assumed of the underlying game. We did not assume it was a single-round, sealed-bid auction. We did not assume that only a winner will make payments. Therefore, we conclude for any potentially wacky, multi-round game the outcomes of all Bayes-Nash equilibria have a nice form.

2.6 Characterization of Dominant Strategy Equilibrium

Dominant strategy equilibrium is a stronger equilibrium concept than Bayes-Nash equilibrium. All dominant strategy equilibria are Bayes-Nash equilibria, but as we have seen, the opposite is not true; for instance, there is no DSE in the first-price auction. Recall that a strategy profile is in DSE if each agent's strategy is optimal for them regardless of what other agents are doing. The DSE characterization theorem below follows from the BNE characterization theorem.

Theorem 2.7 *G and $\mathbf{s}$ are in DSE if and only if for all i,*

1. *(monotonicity) $x_i(\mathbf{v}_{-i}, v_i)$ is monotone non-decreasing in v_i, and*
2. *(payment identity) $p_i(\mathbf{v}_{-i}, v_i) = v_i x_i(\mathbf{v}_{-i}, v_i) - \int_0^{v_i} x_i(\mathbf{v}_{-i}, z)dz + p_i(\mathbf{v}_{-i}, 0)$,*

where often $p_i(0) = 0$.[3]

It was important when discussing BNE to explicitly refer to $x_i(v_i)$ and $p_i(v_i)$ as the probability of allocation and the expected payments as a game played by agents with values drawn from a distribution will inherently have a randomized outcome and payments. In contrast, for games with DSE we can consider outcomes and payments in a non-probabilistic sense. A deterministic game, i.e., one with no internal randomization, will result in deterministic outcomes and payments. For our single-dimensional agent settings where an agent is either served or not served we will have $x_i(\mathbf{v}) \in \{0, 1\}$. This specification along with the monotonicity condition implied by DSE implies that the function $x_i(\mathbf{v}_{-i}, v_i)$ is a step function in v_i. The reader can easily verify that the payment required for such a step function is exactly the critical value, i.e., t_i at which $x_i(\mathbf{v}_{-i}, t_i)$ changes from 0 to 1. This gives the following corollary.

Corollary 2.8 *A deterministic game G and strategies $\mathbf{s}$ are in DSE if and only if for all i,*

1. *(step-function) $x_i(\mathbf{v}_{-i}, v_i)$ steps from 0 to 1 at $t_i = \inf\{z : x_i(\mathbf{v}_{-i}, z) = 1\}$, and*
2. *(critical values) for $p_i(\mathbf{v}_{-i}, v_i) = \begin{cases} t_i & \text{if } x_i(\mathbf{v}_{-i}, v_i) = 1 \\ 0 & \text{otherwise} \end{cases} + p_i(\mathbf{v}_{-i}, 0)$,*

where often $p_i(0) = 0$.

Notice that the above theorem deliberately skirts around a subtle tie-breaking issue. Consider the truthtelling DSE of the second-price auction on two agents. What happens when $v_1 = v_2$? One agent should win and pay the other's value, but, as this results in a utility of zero, from the perspective of utility maximization both agents are indifferent as to which of them it is. One natural tie-breaking rule is the lexicographical one, i.e., in favor of agent 1 winning. For this rule 1 wins when $v_1 \in [v_2, \infty)$ and 2 wins when $v_2 \in (v_1, \infty)$. The critical values are $t_1 = v_2$ and $t_2 = v_1$. We will usually prefer randomized tie-breaking rule because of its symmetry.

[3] Again we assume strategies map values onto actions. This is necessary for the "if" direction and a notational convenience for the "only if" direction.

2.7 Revenue Equivalence

We are now ready to make one of the most significant observations in auction theory. Namely, mechanisms with the same outcome in BNE have the same expected revenue. In fact, not only do they have the same expected revenue, but each agent has the same expected payment in each mechanism. We state this result as a corollary of Theorem 2.6 which is intuitively clear. The payment identity means that the payment rule is precisely determined by the allocation rule, up to $p_i(0)$ term.

Corollary 2.9 *Consider any two mechanisms where 0-valued agents pay nothing, if the mechanisms have the same BNE outcome then they have same expected revenue.*

We can now quantitatively compare the second-price and first-price auctions from a revenue standpoint. Consider the case where the agent's values are distributed independently and identically. What is the equilibrium outcome of the second-price auction? The agent with the highest valuation wins. What is the equilibrium outcome of the first-price auction? This question requires a little more thought. Since the distributions are identical, it is reasonable to expect that there is a symmetric equilibrium, i.e., one where $s_i = s_{i'}$ for all i and i'. Furthermore, it is reasonable to expect that the strategies are monotone, i.e., an agent with a higher value will out bid an agent with a lower value. Under these assumptions, the agent with the highest value wins. Of course, in both auctions a 0-valued agent will pay nothing. Therefore, we can conclude that their expected revenues are identical.

Recall from Chapter 1 where we calculated the expected revenue of the second-price auction for two bidders with values $U[0, 1]$. We saw that the revenue was equal to the expected minimum of the two values, i.e., 1/3. Earlier in this chapter we calculated the BNE of the first-price auction in the same setting and we concluded that the strategies satisfy $s_i(z) = z/2$. So what is the expected revenue of the first-price auction? The expected value of the highest bidder is 2/3 so their expected bid, which is exactly our expected revenue, is 1/3. The revenues are indeed equivalent.

Corollary 2.10 *When agent's values are independent and identically distributed, the second-price and first-price auction have the same expected revenue.*

Of course, much more bizarre auctions are governed by revenue equivalence. As an exercise the reader is encourage to verify that the *all-pay auction*; where agents submit bids, the highest bidder wins, and all agents pay their bid; is revenue equivalent to the first- and second-price auctions.

2.8 Solving for Bayes-Nash Equilibrium

While it is quite important to know what outcomes are possible in BNE, it is also often important to be able to solve for the BNE strategies. For instance, suppose we are actually a bidder bidding in an auction. How shall we bid? In this section we give a elegant technique for calculating BNE strategies in symmetric settings using revenue equivalence. Actually, we use something a little stronger than revenue equivalence: *interim payment equivalence*.[4] This is the fact that if two mechanisms have the same allocation rule, they must have the same payment rule (because the payment rules satisfy the payment identity).

Suppose we are to solve for the BNE strategies of mechanism M. The approach is to express an agent's payment in M as a function of the agent's action, then to calculate the agents expected payment in a strategically-simple mechanism M' that is revenue equivalent to M (usually a "second-price implementation" of M). Setting these terms equal and solving for the agents action gives the equilibrium strategy.

We give the high level the procedure below. As a running example we will calculate the equilibrium strategies in the first-price auction with two $U[0,1]$ bidders, in doing so we will use a calculation of expected payments in the strategically-simple second-price auction in the same setting.

1. *Guess* what the outcome might be in Bayes-Nash equilibrium.

   E.g., for the first price auction with two bidders with values $U[0,1]$, in BNE, we expect the bidder with the highest value to win. Thus, guess that the highest bidder always wins.

2. *Calculate* the expected payment of a bidder with fixed value in a simple auction with the same equilibrium outcome.

   E.g., Recall that it is a dominant strategy equilibrium (a special case of Bayes-Nash equilibrium) in the second-price auction for each bidder to bid their value. I.e., $b_1 = v_1$ and $b_2 = v_2$. Thus, the second-price auction also selects the bidder with the highest value to win. So, let us calculate the expected payment of player 1 when their value is v_1.

   $$\mathbf{E}[p_1(v_1)] = \mathbf{E}[p_1(v_1) \mid 1 \text{ wins}]\,\mathbf{Pr}[1 \text{ wins}] + \mathbf{E}[p_1(v_1) \mid 1 \text{ loses}]\,\mathbf{Pr}[1 \text{ loses}]$$

   Let us calculate each of these components for the second-price auction:

   $$\begin{aligned}\mathbf{E}[p_1(v_1) \mid 1 \text{ wins}] &= \mathbf{E}[v_2 \mid v_2 < v_1] \\ &= v_1/2.\end{aligned}$$

   The first step follows by the definition of the second-price auction and its dominant strategy equilibrium (i.e., $b_2 = v_2$). The second step follows because in expectation a uniform random variable evenly divides the interval it is over, and once we condition on $v_2 < v_1$, v_2 is $U[0, v_1]$.

   $$\begin{aligned}\mathbf{Pr}[1 \text{ wins}] &= \mathbf{Pr}[v_2 < v_1] \\ &= v_1\end{aligned}$$

[4] In Bayesian games, time is broken into three stages *ex ante*, *interim*, and *ex post* with respect to what is known by whom. In the ex ante stage, only the distribution of valuations, $\mathbf{F}$, is known. In the interim stage, each agent i knows their own value v_i but the other agents' values are known only has being distributed according to $\mathbf{F}_{-i}|_{v_i}$. In the ex post stage, all values of all agents are known to everyone. Therefore the interim payment of an agent is given simply by the payment rule: $p_i(v_i)$.

The first step follows from the definition of the second-price auction and its dominant strategy equilibrium. The second step is because v_1 is uniform on $[0, 1]$.

$$\mathbf{E}[p_1(v_1) \mid 1 \text{ loses}] = 0$$

This is because a loser pays nothing in the second-price auction. This means that we do not need to calculate $\mathbf{Pr}[1 \text{ loses}]$. Plug these into the equation above to obtain:

$$\mathbf{E}[p_1(v_1)] = v_1^2/2.$$

3. *Solve* for bidding strategies from expected payments.

   E.g., by revenue equivalence the expected payment of player 1 with value v_1 is $v_1^2/2$ in both the first-price and second-price auction. We can recalculate this expected payment in the first-price auction using the bid of the player as a variable and then solve for what that bid must be.

   Again,

$$\mathbf{E}[p_1(v_1)] = \mathbf{E}[p_1(v_1) \mid 1 \text{ wins}]\,\mathbf{Pr}[1 \text{ wins}] + \mathbf{E}[p_1(v_1) \mid 1 \text{ loses}]\,\mathbf{Pr}[1 \text{ loses}]$$

   Let's calculate each of these components:

$$\mathbf{E}[p_i(v_i) \mid 1 \text{ wins}] = s_1(v_1)$$

   This by the definition of the first-price auction: if you win you pay your bid.

$$\mathbf{Pr}[1 \text{ wins}] = \mathbf{Pr}[v_2 < v_1] = v_1$$

   This is the same as above.

$$\mathbf{E}[p_1(v_1) \mid 1 \text{ loses}] = 0$$

   This is because a loser pays nothing in the first-price auction. This means that we do not need to calculate $\mathbf{Pr}[1 \text{ loses}]$. Plug these into the equation above to obtain:

$$\mathbf{E}[p_1(v_1)] = s_1(v_1) \cdot v_1.$$

   This must equal the expected payment calculated in the previous step for the second-price auction, implying:

$$v_1^2/2 = s_1(v_1) \cdot v_1$$

   We can solve for $s_1(v_1)$ and get

$$s_1(v_1) = v_1/2.$$

4. Verify initial guess was correct.

   Indeed, if agents follow symmetric strategies $s_1(z) = s_2(z) = z/2$ then the agent with the highest value wins.

In the above first-price auction example it should be readily apparent that we did slightly more work than we had to. In this case it would have been enough to note that in both the first- and second-price auction a loser pays nothing. We could therefore simply equate the expected payments conditioned on winning:

$$\mathbf{E}[p_1(v_1) \mid 1 \text{ wins}] = \underbrace{v_1/2}_{\text{second-price}} = \underbrace{s_1(v_i)}_{\text{first-price}}.$$

We can also work through the above framework for the *all-pay* auction where the agents submit bids, the highest bid wins, but all bidders pay their bid. Again this is revenue equivalent to the second-price auction. However, now we compare the total expected payment (regardless of winning) to conclude:

$$\mathbf{E}[p_1(v_1)] = \underbrace{v_1^2/2}_{\text{second-price}} = \underbrace{s_1(v_i)}_{\text{all-pay}}.$$

I.e., the BNE strategies for the all-pay auction are $s_i(z) = z^2/2$.

2.9 The Revelation Principle

We are interested in designing mechanisms and, while the characterization of Bayes-Nash equilibrium is elegant, solving for equilibrium is still generally quite challenging. The final piece of the puzzle, and the one that has enabled much of modern mechanism design is the *revelation principle*. The revelation principle states, informally, that if we are searching among mechanisms for one with a desirable equilibrium we may restrict our search to single-round, sealed-bid mechanisms in which truthtelling is an equilibrium.

Definition 2.11 A *direct revelation* mechanism is single-round, sealed bid, and has action space equal to the type space, (i.e., an agent can bid any value they might have)

Definition 2.12 A direct relevation mechanism is *Bayesian incentive compatible* (BIC) if truthtelling is a Bayes-Nash equilibrium.

Definition 2.13 A direct revelation mechanism is *(ex post) incentive compatible* (IC) if truthtelling is a dominant strategy equilibrium.

Theorem 2.14 *Any mechanism M with good BNE (resp. DSE) can be converted into a BIC (resp. IC) mechanism M' with the same BNE (resp. DSE) outcome.*

Proof: We will prove the BNE variant of the theorem. Let $\mathbf{s}$, $\mathbf{F}$, and M be in BNE. Define single-round, sealed-bid mechanism M' as follows:

1. Accept sealed bids $\mathbf{b}$.
2. Simulate $\mathbf{s}(\mathbf{b})$ in M.
3. Output the outcome of the simulation.

We now claim that $\mathbf{s}$ is a BNE of M implies truthtelling is a BNE of M' (for distribution $\mathbf{F}$). Let $\mathbf{s}'$ denote the truthtelling strategy. In M', consider agent i and suppose all other agents are truthtelling. This means that the actions of the other players in M are distributed as $\mathbf{s}_{-i}(\mathbf{s}'_{-i}(\mathbf{v}_{-i})) = \mathbf{s}_{-i}(\mathbf{v}_{-i})$ for $\mathbf{v}_{-i} \sim \mathbf{F}_{-i}|_{v_i}$. Of course, in M if other players are playing $\mathbf{s}_{-i}(\mathbf{v}_{-i})$ then since $\mathbf{s}$ is a BNE, i's best response is to play $s_i(v_i)$ as well. The only way i can play this action in the simulation of M is by playing the truthtelling strategy $s'_i(v_i) = v_i$ in M'. ■

Notice that we already seen the revelation principle in action. The second-price auction is the revelation principle applied to the English auction.

Because of the revelation principle, for many of the mechanism design problems we consider, we will look first for Bayesian or ex post incentive compatible mechanisms. The revelation principle guarantees that, in our search for optimal BNE mechanisms, it suffices to search only those that are BIC (and likewise for DSE and IC). The following are corollaries of our BNE and DSE characterization theorems.

Corollary 2.15 *M is BIC for $\mathbf{F}$ if and only if for all i,*

1. *(monotonicity) $x_i(v_i)$ is monotone non-decreasing, and*
2. *(payment identity) $p_i(v_i) = v_i x_i(v_i) - \int_0^{v_i} x_i(z)dz + p_i(0)$,*

where often $p_i(0) = 0$.

Corollary 2.16 *M is IC if and only if for all i,*

1. *(monotonicity) $x_i(\mathbf{v}_{-i}, v_i)$ is monotone non-decreasing in v_i, and*
2. *(payment identity) $p_i(\mathbf{v}_{-i}, v_i) = v_i x_i(\mathbf{v}_{-i}, v_i) - \int_0^{v_i} x_i(\mathbf{v}_{-i}, z)dz + p_i(\mathbf{v}_{-i}, 0)$,*

where often $p_i(0) = 0$.

When we construct mechanisms we will use the "if" directions of these theorms. When discussing incentive compatible mechanisms we will assume that agents follow their equilibrium strategies and, therefore, each agent's bid is equal to their valuation.

Between IC and BIC clearly IC is a stronger incentive constraint and we should prefer it over BIC if possible. Importantly, IC requires fewer assumptions on the agents. For an IC mechanisms, agents must only know their values; while for a BIC mechanism, agents must also know the distribution of other agent values. Unfortunately, there will be some settings where we derive BIC mechanisms where no analogous IC result is known.

The revelation principle fails to hold in some settings of interest. We will take special care to point these out. Two such settings, for instance, are where agents only learn their values over time, or where the designer does not know the prior distribution (and hence cannot simulate the agent strategies).

2.10 Notes

The characterization of Bayes-Nash equilibrium, revenue equivalence, and the revelation principle come from Myerson [18]. The BNE characterization proof presented here comes from Archer and Tardos [2].