14.12 Game Theory Lecture Notes
Reputation and Signaling

In these notes, we discuss the issues of reputation from an incomplete information point of view, using the centipede game. We also introduce the signaling games and illustrate the separating, pooling, and partial-pooling equilibria.

1 Reputation

Consider a game in which a player i has two types, say A and B. Imagine that if the other players believe that i is of type A, then i's equilibrium payoff will be much higher than his equilibrium payoff when the other players believe that he is of type B. If there's a long future in the game and i is patient, then he will act as if he is of type A even when his type is B, in order to convince the other players that he is of type A. In other words, he will try to form a reputation for being of type A. This will change the equilibrium behavior dramatically when the other players assign positive probability to each type. For example, if a seller thinks that the buyer does not value the good that much, then he will be willing to sell the good at a low price. Then, even if the buyer values the good highly, he will pretend that he does not value the good that much and will not buy the object at higher prices—although he could have bought at those prices if it were common knowledge that he values the object highly. (If the players are sufficiently patient, then in equilibrium the price will be very low.) Likewise, in the entry deterrence game, if it is possible that the incumbent gains from a fight in case of an entry, if this is the incumbent's private information, and if there is a long future in the game, then he will fight whenever the entrant enters, in order to form a reputation for being a fighter and deter the future entries. In that case, the entrants will avoid entering even if they are confident that the incumbent is not a fighter. We will now illustrate this notion of reputation formation on the centipede game.

Consider the centipede game in figure 1. In this game, a player prefers to exit (or

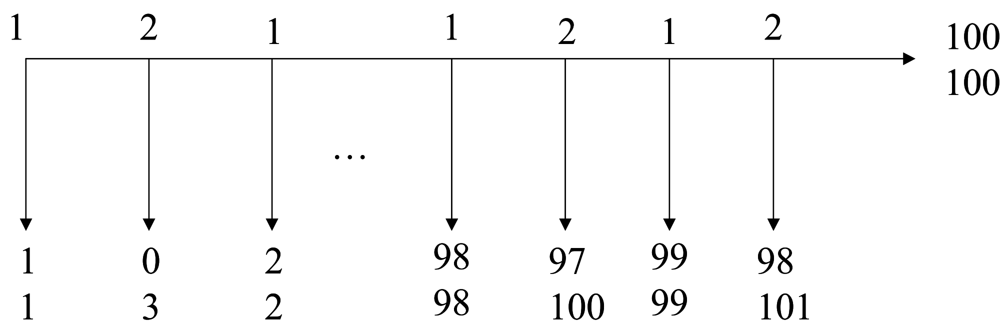

Figure 1:

to go down) if he believes that the other player will exit at the next node. Moreover, player 2 prefers exiting at the last node. Therefore, the unique backward induction outcome in this game is that each player goes down at each node. In particular, player 1 goes down at the first node and the game ends. This outcome is considered to be very counterintuitive, as the players forego very high payoffs. We will see that it is not robust to asymmetric information, in the sense that the outcome would change dramatically if there were a slight probability that a player is of a certain "irrational" type. In figure 2, we consider such a case. Here, player 2 assigns probability .999 to the event that player 1 is a regular rational type, but she also assigns probability .001 to the event that player 1 is a "nice" irrational type who would not want to exit the game. We index the nodes by n starting from the end. Sequential rationality requires that player 1 goes across at each information set on the lower branch. Moreover at the last information set ($n = 1$), player 2 goes down with probability 1. These facts are indicated in figure 3. We will now prove further facts. We need some notation. Let's write μ_n for the probability player 2 assigns to the lower branch at information set n. Moreover, let p_n be the probability that player 1 goes down at node n if he is rational.

Facts about the perfect Bayesian Nash equilibrium

1. *For any $n > 1$, player 2 goes across with positive probability.* Suppose that player

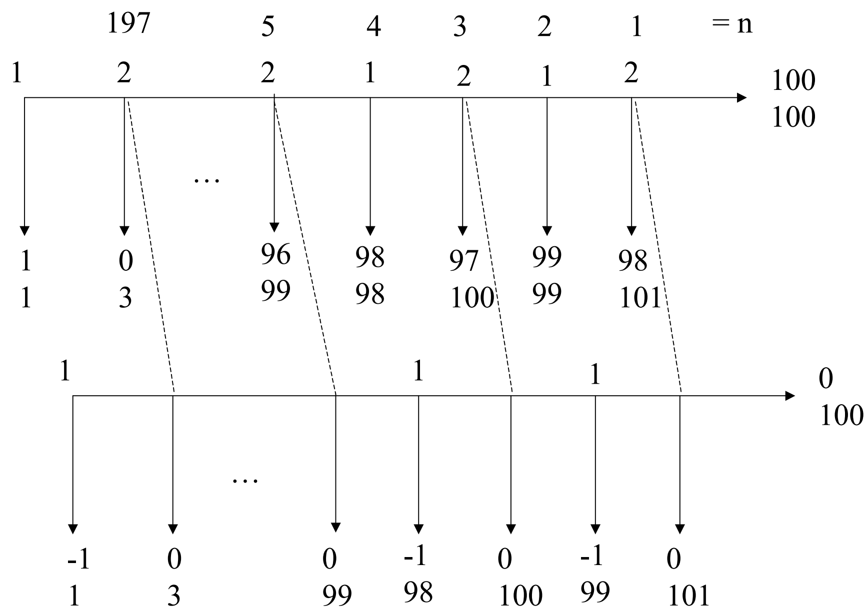

Figure 2:

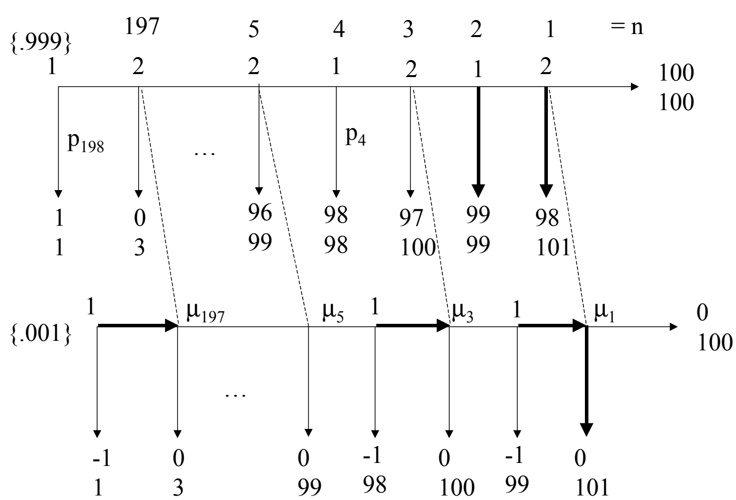

Figure 3:

2 goes down with probability 1 at $n > 1$. Then, if rational, player 1 must go down with probability 1 at $n+1$. Hence, we must have $\mu_n = 1$. That is, player 2 is sure that player 1 is irrational and will go across until the end. She must then go across with probability 1 at n.

2. *Every information set of player 2 is reached.* (This is because irrational player 1 and player 2 go across with positive probability.) Hence, the beliefs are determined by the Bayes' rule.

3. *For any $n > 2$, rational player 1 goes across with positive probability.* Suppose that rational player 1 goes down with probability 1 at $n > 2$. Then, $\mu_{n-1} = 1$, and thus player 2 must go across with probability 1 at $n-1$. Then, rational player 1 must go across at n with probability 1.

4. *If player 2 strictly prefers to go across at n, then*

   (a) *1 goes across with probability 1 at n+1,*

   (b) *2 must strictly prefer to go across at n+2,*

   (c) *2's posterior at n is her prior.*

   Parts a and b must be clear by now. Inductive application of parts a and b imply that players go across with probability 1 until n. Then, c follows from the Bayes' rule.

5. *If rational player 1 goes across with probability 1 at n, then 2's posterior at $n-1$ is her prior.* (This follows from 4.c.)

6. *There exists n^* such that each player go across with probability 1 before n^* (i.e., when $n > n^*$), and rational player 1 and player 2 mix after n^* (i.e., when $n < n^*$).* (This fallows from 1,3, and 4.)

We will now compute the equilibrium. Firstly, note that player 2 mixes at $n = 3$ (i.e., $n^* > 3$). [Otherwise, we would have $\mu_3 = .001$ by 4.c, when player 2 would go down with probability 1 at $n = 3$, contradicting Fact 1.] That is, she is indifferent between going across and going down at $n = 3$. Hence,

$$100 = 101\mu_3 + 99\,(1-\mu_3) = 99 + 2\mu_3,$$

i.e.,

$$\mu_3 = 1/2.$$

Now consider any odd n with $3 < n < n^*$; player 2 moves at n. Write x for the payoff of player 2 at n (if she goes down). Since player 2 is indifferent being going across and going down, we have

$$x = \mu_n(x+1) + (1-\mu_n)[(x-1)p_{n-1} + (1-p_{n-1})(x+1)].$$

The left hand side of this equation is the payoff if she goes down. Let us look at the right hand side. If she goes across, with probability μ_n, player 1 is irrational and will go across at $n-1$ with probability 1, reaching the information set at $n-2$. Since player 2 is indifferent between going down and going across at the information set $n-2$, the expected payoff from reaching this information set for player 2 is $x+1$. This gives the first term. With probability $1-\mu_n$, he is rational and will go down at $n-1$ with probability p_{n-1}, yielding payoff $x-1$ for player 2, and will go across with probability $1-p_{n-1}$, reaching the information set at $n-2$, which yields $x+1$ for player 2. This gives the second term. After some algebraic manipulations, this equation simplifies to

$$(1-\mu_n)p_{n-1} = 1/2. \tag{1}$$

But, by the Bayes' rule, we have

$$\begin{aligned}
\mu_{n-2} &= \frac{\mu_n}{\mu_n + (1-\mu_n)(1-p_{n-1})} \\
&= \frac{\mu_n}{\mu_n + 1 - \mu_n - (1-\mu_n)p_{n-1}} = \frac{\mu_n}{1-(1-\mu_n)p_{n-1}} \\
&= 2\mu_n, \qquad (2)
\end{aligned}$$

where the last equality is due to (1). Therefore,

$$\mu_n = \frac{\mu_{n-2}}{2}.$$

By applying the last equality iteratively, we obtain

$$\begin{aligned}
\mu_3 &= 1/2 \\
\mu_5 &= \mu_3/2 = 1/4 \\
\mu_7 &= \mu_5/2 = 1/8
\end{aligned}$$

$$\begin{aligned}
\mu_9 &= \mu_7/2 = 1/16 \\
\mu_{11} &= \mu_9/2 = 1/32 \\
\mu_{13} &= \mu_{11}/2 = 1/64 \\
\mu_{15} &= \mu_{13}/2 = 1/128 \\
\mu_{17} &= \mu_{15}/2 = 1/256 \\
\mu_{19} &= \mu_{17}/2 = 1/512 \\
\mu_{21} &= \mu_{19}/2 = 1/1024 < .001.
\end{aligned}$$

Therefore, $n^* = 20$. At any even $n < n^*$, player 1 goes across with probability

$$p_n = \frac{1}{2(1-\mu_{n+1})},$$

and at $n-1$ player 2 mixes so that player 1 is indifferent between going across and going down. If we write q_{n-1} for the probability that 2 goes down at $n-1$ and y for the payoff of rational 1 at n, then we have

$$y = q_{n-1}(y-1) + (1-q_{n-1})(y+1) = y + 1 - 2q_{n-1},$$

i.e.,

$$q_{n-1} = 1/2.$$

At $n^*+1 = 21$, $\mu_{n^*+1} = .001$, and at $n^*-1 = 19$, $\mu_{n^*-1} = 1/512$. Hence, by Bayes' rule,

$$1/512 = \mu_{n^*-1} = \frac{\mu_{n^*+1}}{1-\left(1-\mu_{n^*+1}\right)p_{n^*}} = \frac{.001}{1-.999p_{n^*}},$$

yielding

$$p_{n^*} = \frac{1-.512}{.999} \cong .488.$$

At any $n > n^*$, each player goes across with probability 1.

2 Signaling Games

In a signaling game, there are two players: Sender (denoted by S) and Receiver (denoted by R). First, Nature selects a type t_i from a type space $T = \{t_1, \ldots, t_I\}$ with probability $p(t_i)$. Sender observes t_i, and then chooses a message m_j from a message space $M =$

$\{m_1, \ldots, m_J\}$ to be sent to the receiver. Receiver observes m_j (but not t_i), and then chooses an action a_k from an action space $A = \{a_1, \ldots, a_K\}$. The payoffs for the Sender and the Receiver are $U_S(t_i, m_j, a_k)$ and $U_R(t_i, m_j, a_k)$, respectively. A good example for a signaling game is the Beer-Quiche game (depicted in figure 4).

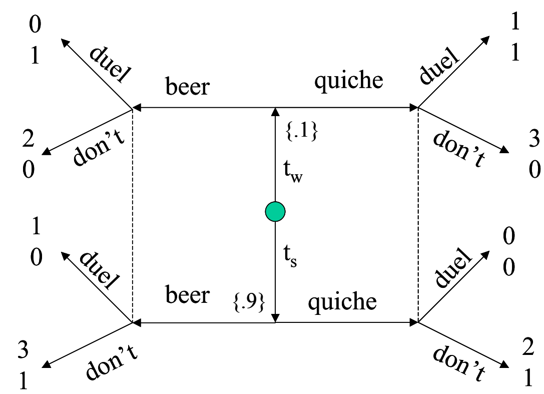

Figure 4:

In this game, Sender is player 1; Receiver is player 2. Type space is $T = \{t_s, t_w\}$, messages are beer and quiche (i.e., $M = \{\text{beer, quiche}\}$), and the actions are "duel" and "don't". In this game there were two equilibrium outcome, either both types have beer or both types have quiche. (The former equilibrium is depicted in figure 5.)

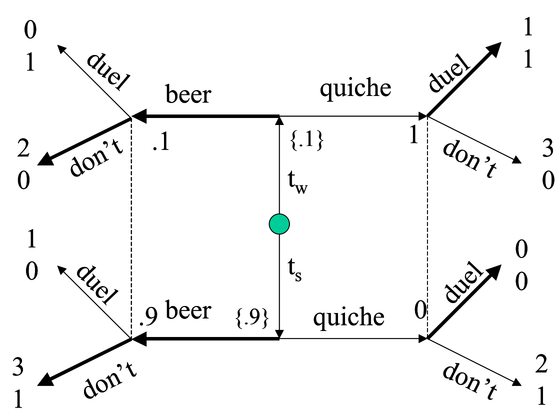

Figure 5:

Such equilibria are called pooling equilibrium, because the types are all send the same

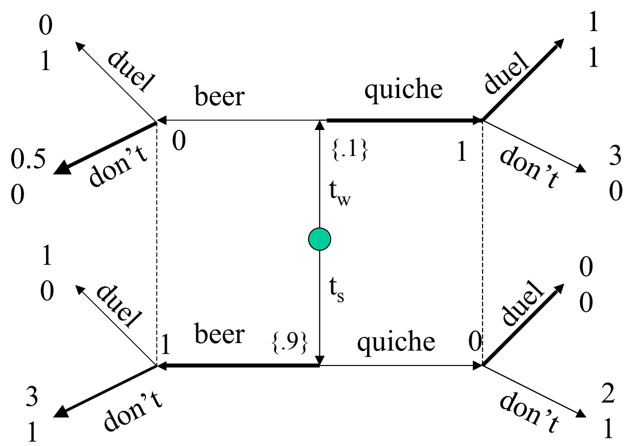

Figure 6:

message, and thus no information is conveyed to the receiver; the receiver's posterior beliefs on the path of equilibrium is the same as his priors. Formally,

Definition 1 *A* pooling equilibrium *is an equilibrium in which all types of sender send the same message.*

Another type of perfect Bayesian Nash equilibrium is separating equilibrium:

Definition 2 *A* separating equilibrium *is an equilibrium in which all types of sender send different messages.*

Since each type sends a different message on the path of a separating equilibrium receiver learns the sender's type (the truth). Such an equilibrium is depicted in Figure 6. Notice that player 2 assigns probability 1 to the strong type after beer and to the weak type after quiche.

Most equilibria in many games will be partially pooling and partially separating:

Definition 3 *A partially separating/pooling equilibrium is an equilibrium in which some types of sender send the same message, while some others sends some other messages.*

Such an equilibrium is depicted in figure 7 for a beer-quiche game in which player 1 is very likely to be weak. In this equilibrium, strong player 1 orders beer, while the weak

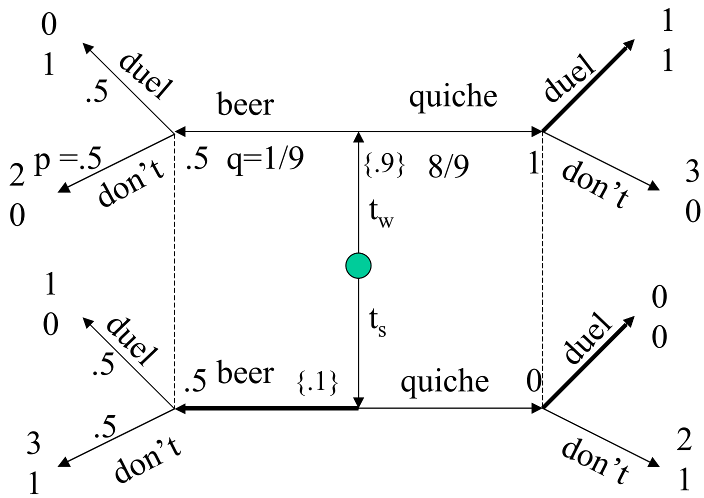

Figure 7:

type mixes between beer and quiche so that after beer, player 2 finds the types equally likely, making him indifferent between the duel and not duel. After beer, player 2 mixes between duel and not duel with equal probabilities so that the weak type of player 1 is indifferent between beer and quiche, allowing him to mix. Note that the beliefs are derived via the Bayes' rule. Note also that, after quiche, player 2 knows that player 1 is weak, while he is not sure about player 1's type after beer — although he updates his belief and find player 1 more likely to be strong.