Signaling Games

Niccolò Lomys[*]

April 5, 2019

Preamble

- These notes heavily draw upon Mailath (2019), Mishra (2010), Kartik (2009), Hörner (2008), Mas-Colell, Whinston and Green (1995), and Fudenberg and Tirole (1991a). All errors are my own. Please bring any error, including typos, to my attention.
- Other excellent surveys on signaling games include Kreps and Sobel (1994) and Riley (2001).

1 General Theory

Signaling games refer narrowly to a class of two-player games of incomplete information in which one player is informed and the other is not. The informed player's strategy set consists of messages contingent on information and the uninformed player's strategy set consists of actions contingent on messages. More generally, a signaling game includes any strategic setting in which players can use the actions of their opponents to make inferences about hidden information.

1.1 Basic Signaling Game

A *finite signaling game* is a dynamic game of incomplete information with two players: a sender, S, and a receiver, R. The timing of the game is the following:

1. Nature draws a *type* t for the sender from the finite set of possible types T according to a probability distribution with full support $p \in \Delta_{++}(T)$.
2. The sender (informed player) observes his type t and chooses a *message* m from the finite set of possible messages M.
3. The receiver (uninformed player) observes the message m, but not the sender's type t, and chooses an *action* a from the finite set of possible actions A. This ends the game and payoffs realize.

[*]Toulouse School of Economics, University of Toulouse Capitole; niccolo.lomys@tse-fr.eu.

4. *Payoffs* are given by $u_S(t, m, a)$ for the sender and $u_R(t, m, a)$ for the receiver.

The implicit assumption here is that, for some reason, the informed party's private information cannot be verifiably disclosed.

Remark 1. In general signaling games, the sets T, M, and A need not be finite. Moreover, the set of possible messages can be a function of the sender's type and the set of possible actions can be a function of the sender's message.

A *strategy for the sender* is a function $\sigma_S \colon T \to \Delta(M)$. A *strategy for the receiver* is a function $\sigma_R \colon M \to \Delta(A)$. With some abuse of notation, denote by $\sigma_S(m \mid t)$ the probability that $\sigma_S(t)$ assigns to m, and with $\sigma_R(a \mid m)$ the probability that $\sigma_R(m)$ assigns to a. A sender's strategy is:

- *Separating* if different sender's types choose different messages; that is, M can be partitioned into sets $\{M_t\}_{t\in T}$ such that, for each t, $\sum_{m\in M_t} \sigma_S(m \mid t) = 1$.
- *Pooling* if all sender's types choose the same message; that is, there is a single message $m \in M$ such that that $\sigma_S(m \mid t) = 1$ for all $t \in T$.
- *Partially separating/pooling* if some sender's types choose the same message, while some other different sender types choose different messages.
- *Hybrid* if some sender's types randomize, whereas others do not.
- *Completely mixed* if all sender's types randomize.

The payoff of the sender of type t to strategy σ_S when the receiver plays strategy σ_R is defined pointwise as

$$U_S(t, \sigma_S(t), \sigma_R) \coloneqq \sum_{m\in M} \left(\sum_{a\in A} u_S(t, m, a)\sigma_R(a \mid m) \right) \sigma_S(m \mid t).$$

The receiver's *ex ante* payoff to strategy σ_R when the sender plays strategy σ_S is defined pointwise as

$$U_R(\sigma_S, \sigma_R) \coloneqq \sum_{t\in T} \left(\sum_{m\in M} \left(\sum_{a\in A} u_R(t, m, a)\sigma_R(a \mid m) \right) \sigma_S(m \mid t) \right) p(t).$$

The receiver, who observes the sender's move before choosing his own action, should update his beliefs about t and base his choice of a on the posterior distribution $\mu(\cdot \mid m) \in \Delta(T)$. How is this posterior formed? In a Bayesian equilibrium, the sender's action can depend on his type. Let $\sigma_S^*(t)$ denote this strategy. Knowing σ_S^* and observing m, the receiver can use Bayes' rule to update $p(\cdot)$ into $\mu(\cdot \mid m)$. The natural extension of subgame perfect Nash equilibrium to the signaling game is *perfect Bayesian equilibrium*, which requires that the receiver maximizes his payoff conditional on m for all $m \in M$, where the receiver's *conditional* payoff to strategy σ_R is defined pointwise as

$$U_R(m, \sigma_R(m)) \coloneqq \sum_{t\in T} \left(\sum_{a\in A} u_R(t, m, a)\sigma_R(a \mid m) \right) \mu(t \mid m).$$

Definition 1 (Perfect Bayesian Equilibrium)**.** *A perfect Bayesian equilibrium (hereafter, PBE) of the signaling game is a strategy profile* (σ_S^*, σ_R^*) *and a beliefs system* $\{\mu(\cdot \mid m)\}_{m \in M}$*, where* $\mu(\cdot \mid m) \in \Delta(T)$ *for all* $m \in M$*, such that:*

1. *For all* $t \in T$*,*
$$\sigma_S^*(t) \in \underset{\sigma_S(t) \in \Delta(M)}{\arg\max}\; U_S(t, \sigma_S(t), \sigma_R^*);$$

2. *For all* $m \in M$*,*
$$\sigma_R^*(m) \in \underset{\sigma_R(m) \in \Delta(A)}{\arg\max}\; U_R(m, \sigma_R(m));$$

3. *For all* $m \in M$*,* $\mu(\cdot \mid m)$ *is given by*
$$\mu(t \mid m) = \frac{p(t)\sigma_S^*(m \mid t)}{\sum_{t' \in T} p(t')\sigma_S^*(m \mid t')}$$
if $\sum_{t' \in T} p(t')\sigma_S^*(m \mid t') > 0$*, and* $\mu(\cdot \mid m)$ *is any element of* $\Delta(T)$ *if* $\sum_{t' \in T} p(t')\sigma_S^*(m \mid t') = 0$*.*

If the sender's strategy σ_S^* *is separating, pooling, partially separating/pooling, hybrid, or completely mixed, then we say that the PBE is separating, pooling, partially separating/pooling, hybrid, or completely mixed.*

The first two PBE requirements are the perfection conditions (sequential rationality): the first one states that the sender takes into account the effect of his message on the receiver's action; the second one states that the receiver reacts optimally to the sender's message given his posterior belief about t. The third PBE requirement corresponds to the application of Bayes' rule. In particular, the requirement states that the beliefs at information sets on the equilibrium path must be computed using Bayes' rule. Note that if message m is not part of the sender's optimal strategy for some type—in which case the information set following message m is off the equilibrium path—observing m is a probability zero event, and Bayes' rule does not pin down posterior beliefs. Any posterior beliefs are then admissible, and so the receiver can play any action that is a best response for some arbitrary belief. This means that the only action excluded are those which are dominated given that m is played. As we will see, the purpose of the refinements of the PBE concept is to put some restriction of these posterior beliefs.

To sum up, a PBE is a set of strategies and beliefs such that, at any stage of the game, strategies are optimal given the beliefs, and the beliefs are obtained from equilibrium strategies and observed actions using Bayes' rule whenever possible.[1]

Remark 2. If type t chooses message m in a separating PBE, then, by the third equilibrium requirement, the receiver assigns probability 1 to t after observing m. Therefore, after the sender takes his action, the receiver learns his type (putting probability 1 on the correct type). In contrast, in a pooling PBE, the receiver learns nothing from the sender's signal.

[1]For the notion of PBE as used in these notes, see Fudenberg and Tirole (1991b).

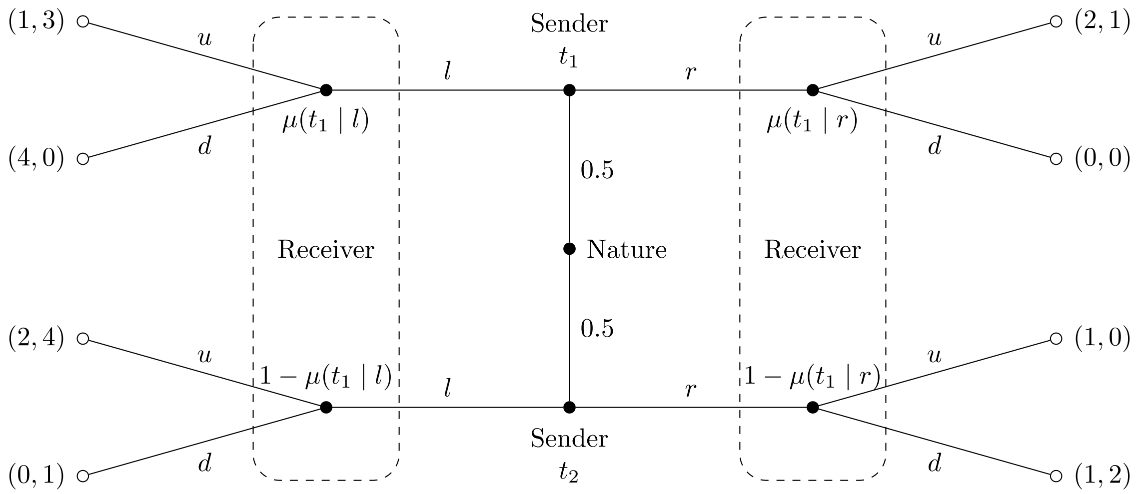

Figure 1: An Example of a Signaling Game

Remark 3. For the finite signaling games we discuss in this section and for the job market signaling game we discuss in Section 2, the notion of PBE is equivalent to the sequential equilibrium notion.

Example 1. Consider the signaling game in Figure 1. Here: $T = \{t_1, t_2\}$, with $p(t_1) = p(t_2) = 1/2$; $M = \{l, r\}$; $A = \{u, d\}$. There are four possible PBEs in pure strategies.

1. *Pooling on* l. Suppose there is a PBE in which both sender's types play l. By Bayes' rule, the receiver's beliefs must satisfy $\mu(t_1 \mid l) = 1/2$. Thus, after observing message l, the receiver's best response is to play u. Hence, by playing l, the sender of type t_1 receives payoff 1, and the sender of type t_2 receives payoff 2. What does the receiver do if he observes message r? He plays u if and only if $\mu(t_1 \mid r) \geq 2/3$. But if $\mu(t_1 \mid r) \geq 2/3$, the sender of type t_1 receives payoff $2 > 1$ by playing r. So, in equilibrium, we must have $\mu(t_1 \mid r) \leq 2/3$ and the receiver must play d when he observes message r. In this case, by playing r, the sender of type t_1 receives payoff 0, and the sender of type t_2 receives payoff 1, which are smaller than the respective payoffs by playing l. To sum up, the followings describe pooling PBEs of this signaling game:

$$\begin{aligned} \sigma_S^*(l \mid t_1) = 1 \quad &\text{and} \quad \sigma_S^*(l \mid t_2) = 1 \\ \sigma_R^*(u \mid l) = 1 \quad &\text{and} \quad \sigma_R^*(d \mid r) = 1 \\ \mu(t_1 \mid l) = \frac{1}{2} \quad &\text{and} \quad \mu(t_1 \mid r) \leq \frac{2}{3}. \end{aligned}$$

2. *Pooling on* r. Suppose there is a PBE in which both sender's types play r. By Bayes' rule, the receiver's beliefs must satisfy $\mu(t_1 \mid r) = 1/2$. Hence, the receiver's best response is to play d. This gives payoff 0 to the sender of type t_1 and 1 to the sender of type t_2. For any value of $\mu(t_1 \mid l)$, the receiver plays u after observing message l. This gives payoff 1 to t_1 and 2 to t_2, which are greater than the respective payoffs by playing r. Hence, no PBE with pooling on r exists.

3. *Separating with t_1 choosing l.* Suppose there is a PBE in which the sender of type t_1 chooses l and the sender of type t_2 chooses r. Now, both information sets are reached in equilibrium. So, $\mu(t_1 \mid l) = 1$ and $\mu(t_1 \mid r) = 0$. The receiver's best response is to play u after observing l and d after observing r. This gives payoffs 1 to t_1 and 1 to t_2. If type t_2 deviates and plays l, then he gets a payoff of 2 (since the receiver responds with u). Hence, no separating PBE with t_1 choosing l exists.

4. *Separating with t_1 choosing r.* Suppose there is a PBE in which the sender of type t_1 chooses r and the sender of type t_2 chooses l. Now, $\mu(t_1 \mid l) = 0$ and $\mu(t_1 \mid r) = 1$. The receiver's best response is to play u for both l and r. This gives payoffs 2 to t_1 and 2 to t_2. If t_1 deviates and plays l, then he gets a payoff of 1; so, it is not optimal to deviate for t_1. If t_2 deviates and plays r, then he gets a payoff of 1; so, it is not optimal to deviate for t_2. To sum up, the followings describe the separating PBE of the signaling game:

$$\begin{aligned} \sigma_S^*(r \mid t_1) = 1 \quad &\text{and} \quad \sigma_S^*(l \mid t_2) = 1 \\ \sigma_R^*(u \mid l) = 1 \quad &\text{and} \quad \sigma_R^*(u \mid r) = 1 \\ \mu(t_1 \mid l) = 0 \quad &\text{and} \quad \mu(t_1 \mid r) = 1. \end{aligned}$$

1.2 Equilibrium Refinements

Multiple equilibria arise in signaling games because PBE does not constrain the receiver's response to messages sent with zero probability in equilibrium. The equilibrium set shrinks if one imposes restrictions on the beliefs the receiver can have at unreached information sets (or, in other words, if one restricts the meaning of unsent messages). We now discuss techniques that refine the set of equilibria in signaling games. In particular, the ideas we discuss in these notes are due to Cho and Kreps (1987) and Banks and Sobel (1987).

Dominance and Requirement 0. Consider the signaling game in Figure 2. Consider a pooling PBE where both sender's types play l. By Bayes' rule, $\mu(t_1 \mid l) = 1/2$, and the receiver's best response to message l is to play u. Hence, by playing l, the sender of type t_1 receives payoff 3 and the sender of type t_2 receives payoff 1. The receiver's best response to message r is d if $\mu(t_1 \mid r) \geq 1/2$ and u if $\mu(t_1 \mid r) \leq 1/2$. Suppose the receiver sets $\mu(t_1 \mid r) \geq 1/2$ and plays d in response to r. Then, both t_1 and t_2 get 0 by deviating, which is smaller than the respective payoffs by playing l. To sum up, the followings describe pooling PBEs of this signaling game:

$$\sigma_S^*(l \mid t_1) = 1 \quad \text{and} \quad \sigma_S^*(l \mid t_2) = 1 \tag{1}$$

$$\sigma_R^*(u \mid l) = 1 \quad \text{and} \quad \sigma_R^*(d \mid r) = 1 \tag{2}$$

$$\mu(t_1 \mid l) = \frac{1}{2} \quad \text{and} \quad \mu(t_1 \mid r) \geq \frac{1}{2}. \tag{3}$$

Are these PBEs "reasonable"? The key feature here is that it makes little sense for the sender of type t_1 to play r; this is so because the smallest payoff the sender of type t_1 can receive by playing l (2) is greater than the largest payoff he can receive by playing r (1). So, if the Receiver

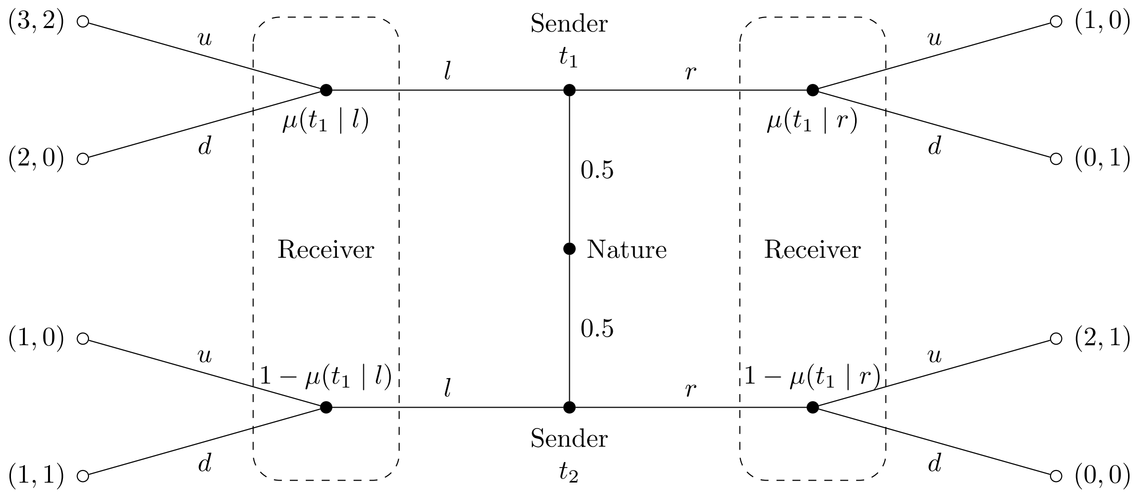

Figure 2: Domination in a Signaling Game

observes r, he should infer the sender's type to be t_2, and thus have belief $\mu(t_1 \mid r) = 0$. This motivates the next definition and the first equilibrium refinement.

Definition 2 (Dominance)**.** *The message $m \in M$ is* dominated for the sender of type $t \in T$ *if there exists another message $m' \in M$ such that the smallest payoff the sender of type t can receive from m' is greater than the largest payoff he can receive from m, that is if*

$$\min_{a \in A} u_S(t, m', a) > \max_{a \in A} u_S(t, m, a).$$

Definition 3 (Dominance Criterion)**.** *The PBE $\left\{(\sigma_S^*, \sigma_R^*), \{\mu(\cdot \mid m)\}_{m \in M}\right\}$ of the signaling game satisfies the* Dominance Criterion *if the following condition holds:*

- *If the information set following message $m \in M$ is off the equilibrium path and m is dominated for some type t, then $\mu(t \mid m) = 0$ if possible (i.e. provided m is not dominated for all $t \in T$).*

Clearly, the pooling PBEs of the signaling game in Figure 2 described by (1)-(3) do not satisfy the Dominance Criterion. The game game in Figure 2 also has the separating PBE described by

$$\sigma_S^*(l \mid t_1) = 1 \quad \text{and} \quad \sigma_S^*(r \mid t_2) = 1$$
$$\sigma_R^*(u \mid l) = 1 \quad \text{and} \quad \sigma_R^*(u \mid r) = 1$$
$$\mu(t_1 \mid l) = 1 \quad \text{and} \quad \mu(t_1 \mid r) = 0.$$

This separating PBE trivially satisfies the Dominance Criterion since all information sets are on the equilibrium path. For a non-trivial example, reverse the receiver's payoff from type t_2 when he plays r: 1 from playing d and 0 from playing u in Figure 2. Now, the followings describe pooling PBEs of the modified signaling game:

$$\sigma_S^*(l \mid t_1) = 1 \quad \text{and} \quad \sigma_S^*(l \mid t_2) = 1$$

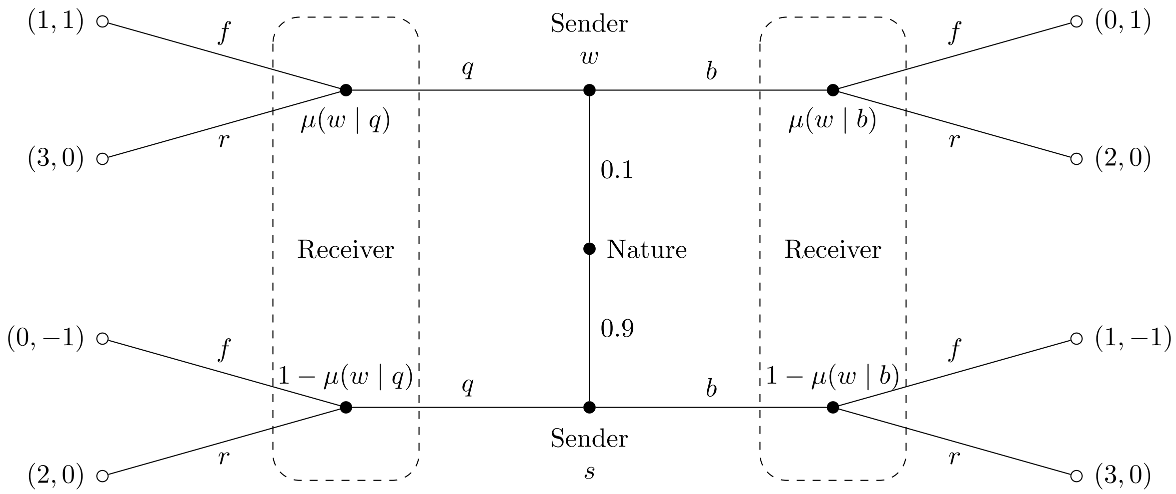

Figure 3: Beer or Quiche

$$\sigma^*_R(u \mid l) = 1 \quad \text{and} \quad \sigma^*_R(d \mid r) = 1$$
$$\mu(t_1 \mid l) = \frac{1}{2} \quad \text{and} \quad \mu(t_1 \mid r) \in [0, 1].$$

The pooling PBE with $\mu(t_1 \mid r) = 0$ satisfies the Dominance Criterion (make sure you understand this).

Equilibrium Dominance and The Intuitive Criterion. In many games, however, there are PBE that still seem unreasonable despite they satisfy the Dominance Criterion. For instance, consider the famous *beer or quiche* game in Figure 3, where:

- The types of sender are weak, w, and strong, s, $T = \{w, s\}$, with $p(w) = 0.1$ and $p(s) = 0.9$;
- The messages available to the sender are quiche, q, and beer, b, $M = \{q, b\}$;
- The receiver's actions are fight, f, and runaway, r, $A = \{f, r\}$.

The narrative is as follows. The strong type likes beer for breakfast, while the weak type likes quiche. The sender is ordering his breakfast, while the receiver, who is a bully, is watching and contemplating whether to pick a fight with the sender. The receiver would like to pick a fight if the sender is weak but runaway if he is strong. His payoffs are such that if he assign probability greater than 1/2 to weak, he prefers a fight, and if he assigns probability greater than 1/2 to strong, he prefers to runaway. The sender would like to avoid a fight: he gets 1 utile from the preferred breakfast and 2 utiles from avoiding the fight. Before observing the sender's breakfast choice, the receiver finds it more likely that the sender is strong.[2]

In the beer or quiche signaling game there cannot be a separating equilibrium. By way of contradiction, suppose there is one. Then: (i) the two sender's types choose different messages; (ii) the receiver infers the sender's type correctly and fights type w. But then, the w type can profitably deviate to the s type's message in the putative equilibrium, which is a contradiction.

[2]The labels are an allusion to a popular 1982 book, "Real Men Don't Eat Quiche" by Bruce Feirstein. The Beer or Quiche game is proposed by Cho and Kreps (1987).

Consider a pooling PBE where both sender's types play q. By Bayes' rule, $\mu(w \mid 1) = 1/10$, and the receiver's best response to message q is to play r. Hence, by playing q, the sender of type w receives payoff 3 and the sender of type s receives payoff 2. The receiver's best response to message b is f if $\mu(w \mid b) \geq 1/2$ and r if $\mu(w \mid b) \leq 1/2$. Suppose the receiver sets $\mu(w \mid b) \geq 1/2$ and plays f in response to b. Then, by playing b, the sender of type w receives payoff 0 and the sender of type s receives payoff 1, which are smaller than the respective payoffs by playing b. To sum up, the followings describe pooling PBEs the beer or quiche signaling game:

$$\sigma_S^*(q \mid w) = 1 \quad \text{and} \quad \sigma_S^*(q \mid s) = 1 \tag{4}$$

$$\sigma_R^*(r \mid q) = 1 \quad \text{and} \quad \sigma_R^*(f \mid b) = 1 \tag{5}$$

$$\mu(w \mid q) = \frac{1}{10} \quad \text{and} \quad \mu(w \mid b) \geq \frac{1}{2}. \tag{6}$$

Since message b is not dominated for any type, the Dominance Criterion is trivially satisfied here. In particular, the sender of type w is not guaranteed to do better by choosing b over q (w's smallest payoff from q is 1, whereas w's largest payoff from b is 2). Despite the Dominance Criterion is satisfied, these pooling equilibria are often argued to be unintuitive. Would the sender of type w ever "reasonably" deviate to play b? In this pooling PBEs, the sender of type w receives payoff 3, and this is strictly greater than his payoff from b no matter how the receiver responds. On the other hand, if by deviating to b the sender of type s can "signal" that he is indeed s, he is strictly better off, since the receiver's best response is to play r, yielding payoff of $3 > 2$. Given this, one would expect the s type to deviate, and not the w type. That is, if the receiver see b, he should believe that it is the s type who has deviated. In that case, $\mu(w \mid b) = 0$, and the receiver chooses r. This means, the s type will deviate to choose b. This motivates the next definition and the next equilibrium refinement.

Definition 4 (Equilibrium Dominance). *Let $\{(\sigma_S^*, \sigma_R^*), \{\mu(\cdot \mid m)\}_{m \in M}\}$ be a PBE of the signaling game. The message $m \in M$ is* equilibrium dominated *for the sender of type $t \in T$ if t's equilibrium payoff, denoted $U_S^*(t)$, is greater than the largest payoff he can receive from m, that is if*

$$U_S^*(t) > \max_{a \in A} u_S(t, m, a).$$

Definition 5 (Intuitive Criterion). *The PBE $\{(\sigma_S^*, \sigma_R^*), \{\mu(\cdot \mid m)\}_{m \in M}\}$ of the signaling game satisfies the* Intuitive Criterion *if the following condition holds:*

- *If the information set following message $m \in M$ is off the equilibrium path and m is equilibrium dominated for some type t, then $\mu(t \mid m) = 0$ if possible (i.e. provided m is not equilibrium dominated for all $t \in T$).*

Clearly, the pooling PBEs of the beer or quiche signaling game in Figure 3 described by (4)-(6) do not satisfy the intuitive criterion. The beer or quiche signaling game also has the pooling PBEs described by the following:

$$\sigma_S^*(b \mid w) = 1 \quad \text{and} \quad \sigma_S^*(b \mid s) = 1 \tag{7}$$

$$\sigma_R^*(f \mid q) = 1 \quad \text{and} \quad \sigma_R^*(r \mid b) = 1 \tag{8}$$

$$\mu(w \mid q) \geq \frac{1}{2} \quad \text{and} \quad \mu(w \mid b) = \frac{1}{10}. \tag{9}$$

It is easy to check that these pooling PBEs satisfy the Intuitive Criterion (make sure you understand this).

Dominance arguments use only the fact that the action m is dominated and thus is independent of the equilibrium that is eliminated. By contrast, equilibrium dominance is relative to the equilibrium that is eliminated. Any equilibrium satisfying the Intuitive Criterion must also satisfy the Dominance Criterion. Thus, if a PBE satisfies the Intuitive Criterion, it is redundant to check whether it also satisfies the Dominance Criterion. It can be shown that a PBE satisfying the Intuitive Criterion always exists in signaling games (see Cho and Kreps (1987)).

This kind of argument is often referred to as *forward induction* in game theory. The basic idea is that an agent tries to form beliefs on his information set based on what could have happened in the past. Here, the receiver is forming beliefs at the off equilibrium information set based on which type could have "rationally" sent the message corresponding to this information set. In particular, the Intuitive Criterion is implied by Kohlberg and Mertens (1986) implementation of forward induction and strategic stability.

1.3 Exercises

Exercise 1. Consider the signaling game illustrated in Figure 1, where $p \in (0, 1)$.

(a) Show that the game has no Nash equilibrium in pure strategies (note: this implies that there is no PBE in pure strategies).

(b) Suppose $p = 1/2$. Describe a PBE of the game.

(c) How does your answer to part (b) change if $p = 1/10$?

Solution.

(a) To begin, note that by choosing l both sender types can get at least -2 and at most 2. We proceed by contradiction.

First, suppose there is an equilibrium in which the receiver chooses t_r after observing message r. Then, the sender of type t_1 sends message r because $3 > 2$ and the sender of type t_2 sends message l because $-2 > -3$. Then, the receiver's beliefs are $\mu(t_1 \mid r) = 1$ and $\mu(t_2 \mid l) = 1$. However, given these beliefs, choosing b_r after observing message r is a profitable deviation for the receiver.

Next, suppose there is an equilibrium in which the receiver chooses b_r after observing message r. Then, the sender of type t_1 sends message l because $-2 > -3$ and the sender of type t_2 sends message l because $3 > 2$. Then, the receiver's beliefs are $\mu(t_1 \mid l) = 1$ and $\mu(t_2 \mid 2) = 1$. However, given these beliefs, choosing t_r after observing message r is a profitable deviation for the receiver.

To sum up, the game in Figure 1 has no Nash equilibrium in pure strategies.

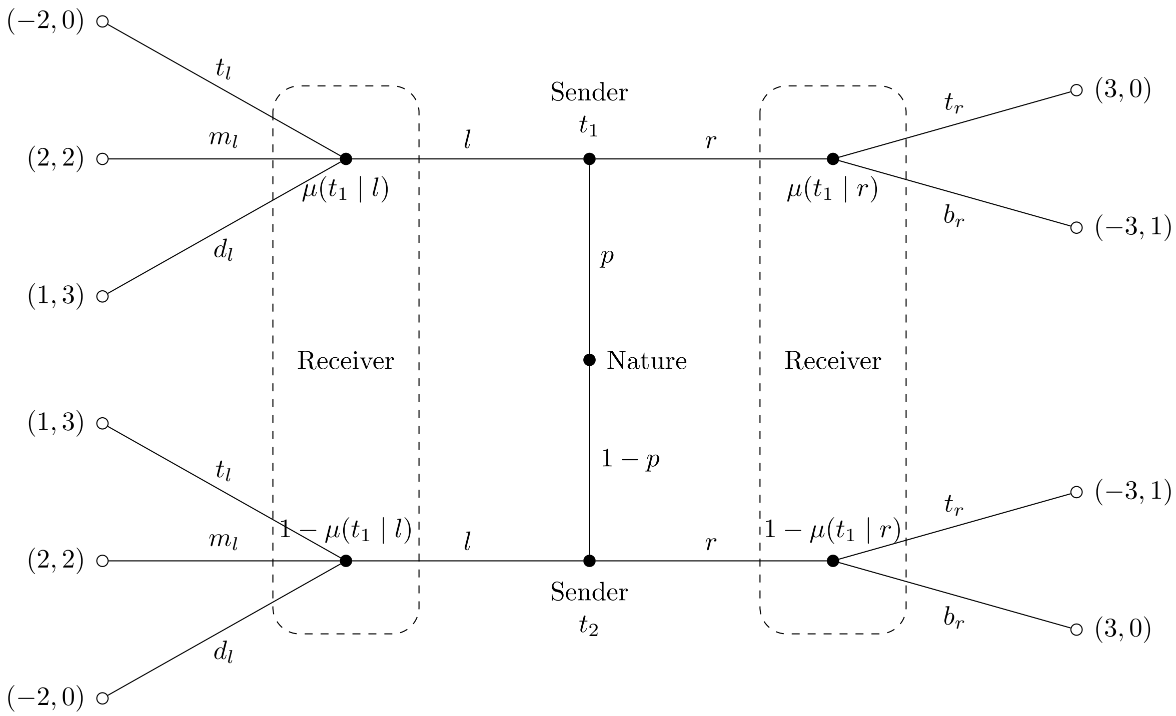

Figure 4: Figure for Exercise 1

(b) The following strategies describe a PBE of the signaling game in Figure 1 with $p = 1/2$:

$$\sigma_S^*(l \mid t_1) = 1 \text{ and } \sigma_S^*(l \mid t_2) = 1$$
$$\sigma_R^*(m_l \mid l) = 1 \text{ and } \sigma_R^*(t_r \mid r) = \frac{1}{2}.$$

for an appropriate belief system, as we will now show.

Since under σ_S^* no type sends r with positive probability, we can arbitrarily specify $\mu(t_1 \mid r)$ and $\mu(t_2 \mid r)$ to any value. In order for the receiver to be indifferent between t_r and b_r after observing r, it must be that

$$\mu(t_1 \mid r) = \frac{1}{2}.$$

Since $\sigma_S^*(l \mid t_1) = \sigma_S^*(l \mid t_2) = 1$, by Bayes' rule we must have

$$\mu(t_1 \mid l) = p = 1/2.$$

Hence, after observing l, the receiver's expected payoff from choosing t_l or d_l is $3/2$, whereas that from choosing m_l is 2. Thus, it is optimal for the receiver to choose m_l after observing l.

Now consider the sender. Given σ_R^*, both types of the sender get 0 from choosing message r and 2 from choosing message l. Thus, it is optimal to chose message l with probability 1.

(c) If $p = 1/10$, then the strategy profile and the belief system described in part (b) is no longer a PBE of the signaling game. To see this, note that when $p = 1/10$, after observing l, the receiver's expected payoff is $3/10$ from choosing b_l, 2 from choosing m_l and $27/10$ from choosing t_l. Hence, it is optimal for the receiver to choose t_l after observing l. Moreover, if we set $\sigma_R^*(t_l \mid l) = 1$, then choosing r: (i) is a profitable deviation for the sender of type t_1 for any $\mu(t_1 \mid r) \leq 1/2$; (i) is a profitable deviation for the sender of type t_2 for any $\mu(t_1 \mid r) < 1/2$ (make sure you understand this). Hence, there cannot be a PBE in which both sender types pool on message l.

The followings describe a PBE of the signaling game in Figure 1 with $p = 1/10$:

$$\sigma_S^*(l \mid t_1) = 0 \ \text{ and } \ \sigma_S^*(l \mid t_2) = \frac{8}{9}$$
$$\sigma_R^*(t_l \mid l) = 1 \ \text{ and } \ \sigma_R^*(t_r \mid r) = \frac{1}{3}$$
$$\mu(t_1 \mid l) = 0 \ \text{ and } \ \mu(t_1 \mid r) = \frac{1}{2}.$$

Verify this by yourself. ■

Exercise 2 (Cheap Talk)**.** Gilligan and Krehbiel (1989) depict the open rule in Congress as a cheap talk game, that is, a signaling game in which signals are costless (see Crawford and Sobel (1982) for the first example of a cheap talk game). As a rough approximation, the committee proposes a policy, but the floor can introduce amendments and choose the policy it likes. The open rule is depicted as a two-player game, with a single member in the committee (player 1) and a single representative on the floor (player 2, who stands for the median voter). The object of the final decision is a policy $a_2 \in \mathbb{R}$. The outcome given policy a_2 is $x = a_2 + \omega$, where ω is a random variable uniformly distributed on $[0, 1]$. The committee knows ω; the floor does not. The preferences of both players are quadratic with bliss point $x = 0$ for the floor and $x = x_c \in (0, 1/4)$ for the committee: $u_1(x) := -(x - x_c)^2$ and $u_2(x) := -x^2$.

(a) Show that there is always a "babbling" PBE in which a_1 is uninformative and $a_2 = 1/2$.

(b) Look for informative PBEs. In particular, find an equilibrium in which the committee "reports low" when $\omega \in [0, \omega^*]$ and "reports high" when $\omega \in (\omega^*, 1]$.

Solution.

(a) Consider the following strategy profile and beliefs:

- Player 1 chooses $a_1 = 0$ regardless of ω.
- Player 2 believes that ω is uniformly distributed on $[0, 1]$ after any message $a_1 \in \mathbb{R}$.
- Player 2 chooses $a_2 = 1/2$ after all $a_1 \in \mathbb{R}$.

Beliefs are clearly such that they satisfy Bayes' rule whenever possible. Given his beliefs, player 2 chooses $a_2(a_1)$ to maximize

$$-\mathbb{E}_\omega\left[(a_2 + \omega)^2 \mid a_1\right].$$

This expression is maximized at $a_2 = -\mathbb{E}(\omega \mid a_1)$. Conditional on any a_1, player 2 believes that ω is uniformly distributed on $[0,1]$, and so $\mathbb{E}(\omega \mid a_1) = 1/2$. Thus, $a_2(a_1) = 1/2$ for all a_1 is a best response for player 2. Finally, since player 1's action has no effect on player 2's action, player 1's strategy is also a best response.

(b) Consider the following strategy profile and beliefs:

- Player 1 chooses a_1 at random from some distribution with full support on $(-\infty, 0)$ if $\omega \in [0, \omega^*]$, where $\omega^* := 2x_c + 1/2$; Player 1 chooses a_1 at random from some distribution with full support on $[0, +\infty)$ if $\omega \in [\omega^*, 0]$.
- Player 2 believes that ω is uniformly distributed on $[0, \omega^*]$ after observing message $a_1 < 0$ and that ω is uniformly distributed on $[\omega^*, 1]$ after observing message $a_1 \geq 0$.
- Player 2 chooses $a_2 = -\omega^*/2$ if $a_1 < 0$ and $a_2 = -(\omega^* + 1)/2$ if $a_2 \geq O$.

Suppose player 1 plays the strategy above. Then, player 2's best response is to choose $a_2(a_1)$ to maximize

$$-\mathbb{E}_\omega\left[(a_2 + \omega)^2 \mid a_1\right].$$

Again, this expression is maximized at $a_2 = -\mathbb{E}(\omega \mid a_1)$. Conditional on $a_1 < 0$, player 2 believes that ω is uniformly distributed on $[0, \omega^*]$, and so

$$\mathbb{E}(\omega \mid a_1 < 0) = \frac{\omega^*}{2}.$$

Conditional on $a_1 \geq 0$, player 2 believes that ω is uniformly distributed on $[\omega^*, 1]$, and so

$$\mathbb{E}(\omega \mid a_1 \geq 0) = \frac{\omega^* + 1}{2}.$$

Thus, player 2's strategy is optimal.

Next, consider player 1. If player 1 observes $\omega \in [0, \omega^*]$, his payoffs to sending messages a_1 drawn from the distribution with full support on $(-\infty, 0)$ is

$$-\left[x_c - \left(\frac{\omega^*}{2} + \omega\right)\right]^2,$$

whereas his payoffs to sending messages a_1 drawn from the distribution with full support on $[0, +\infty)$ is

$$-\left[x_c - \left(\frac{\omega^* + 1}{2} + \omega\right)\right]^2$$

The first expression is larger and thus player 1 is playing a best response if $\omega \leq \omega^*$ (verify this by yourself). The same argument shows that sending messages a_1 drawn from the distribution with full support on $[0, +\infty)$ is a best response for player 1 if $\omega \in (\omega^*, 1]$. ■

2 Job Market Signaling

Signaling often works as a market mechanism through which adverse selection can be solved or mitigated (for an introduction to adverse selection, see Rey (2018)). The idea is that "high types" may have actions they can take to distinguish themselves from "low types". In this section, we present the job market signaling model, whose study originated with Spence (1973, 1974).

2.1 Setting

We consider the simplest version of Spence's model. There is worker and two firms. We denote a generic firm by $i \in \{1, 2\}$. The worker has private type (interpreted as his ability or productivity) $\theta \in \Theta := \{\theta_L, \theta_H\}$, with $\theta_H > \theta_L > 0$. Let $\lambda \in (0, 1)$ be the probability that the worker is of type θ_H. Before entering the labor market, the worker can obtain education at some cost. In particular, a worker of type θ can obtain education level $e \in \mathbb{R}_+$ at cost $c(e, \theta) \geq 0$. Firms observe the worker's education level after he has acquired it (but they do not observe his type) and each simultaneously offers a wage $w_i \in \mathbb{R}_+$. Finally, the worker accepts one or neither job and payoffs realize. We normalize the reservation wage for both types of worker to zero. Payoffs are as follows: the worker gets

$$u(w, e, \theta) := w - c(e, \theta) \tag{10}$$

if he accepts an offer, and $-c(e, \theta)$ if he does not; the firm that employs the worker gets $\theta - w$; a firm that does not employ the worker gets 0. To sum up, the timing of the game is the following:

1. Nature draws the worker's type θ according to λ.

2. The worker observes θ and chooses an education level e.

3. After observing e, but not θ, each firm i simultaneously offers a wage, w_i.

4. The worker accepts one or neither job and payoffs realize.

It is sufficient to only consider two firms. This is because when firms compete for the worker through their wage offerings, two firms are sufficient for Bertrand competition to drive wages up to the marginal product of labor, θ, or, more generally, $\mathbb{E}(\theta) := \lambda\theta_H + (1 - \lambda)\theta_L$.

As specified, education is absolutely worthless in terms of increasing productivity—it is solely an instrument to potentially signal some private information. This is known as *purely dissipative* signaling. In practice, education also has a productivity-enhancing purpose. Even in that case, it can serve as a signaling instrument (see Exercise 4). As we will see, what is important is the difference in marginal costs of acquiring education for the different worker types.

We assume the cost function $c(\cdot, \cdot)$ satisfies the following conditions (subscripts denote partial derivatives):

- $c(\cdot, \cdot)$ is twice continuously differentiable.

- $c(0, \theta) = 0$ for all θ; that is, the cost of no education is zero (normalization).

- $c_e(e, \theta) > 0$ and $c_{ee}(e, \theta) > 0$ for all e and for all θ; that is, acquiring more education is always costly on the margin, and this marginal cost is increasing.

- $c_\theta(e, \theta) < 0$ for all $e > 0$ and all θ, and $c_{e\theta}(e, \theta) < 0$ for all e and for all θ; that is, the cost and the marginal cost of education is lower for the high type of worker.

Remark 4. Suppose acquiring education is not possible. Under the assumption that the reservation wage for both types of worker is zero, in a market equilibrium both types are employed at wage $w = \mathbb{E}(\theta)$. This is efficient and Pareto efficient. In other words, under the assumption that the reservation wage for both types of worker is zero, there is no adverse selection effect in a market equilibrium. Our goal here is to study how signaling works and its potential inefficiencies, and this easiest done with this simplifying assumption. It can, however, be weakened (see Exercise 5).

2.2 Basic Properties

The goal is to study whether and how education can be used by workers to signal their type to the firms. Formally, a strategy for the worker is a function $e\colon \Theta \to \mathbb{R}_+$, and a strategy for firm i is a function $w_i\colon \mathbb{R}_+ \to \mathbb{R}_+$. In what follows, we are going to analyze pure strategy PBEs that satisfy the following additional property: for any level of education chosen by the worker, e, both firms have the same belief over the worker's type; that is, if $\mu_i(e)$ represents the belief of firm $i \in \{1, 2\}$ that the worker is type θ_H given that he has chosen education e, then we require that $\mu_1(e) = \mu_2(e)$ for all e. Denote $\mu(e)$ such common belief. For short, hereafter we will just refer to any of these pure PBEs as an equilibrium.

Using generalized backward induction, we start analyzing play at the end of the game.

- *Stage 4.* In the last stage, sequential rationality requires that the worker accepts a job offer from the firm that offers the higher wage.[3] If the two firms offer the same wage, the worker accepts one of them uniformly at random.

- *Stage 3.* Given the worker's behavior in stage 4, it is straightforward that for any belief $\mu(e)$, there is a unique (pure strategy) Nash Equilibrium in the subgame at stage 3 when both firms simultaneously propose wages: they must both offer

$$w(e) = \mathbb{E}[\theta \mid \mu(e)] := \mu(e)\theta_H + (1 - \mu(e))\theta_L. \tag{11}$$

  To see this, note that a firm expects strictly negative payoffs if it hires the worker at a wage larger than $\mathbb{E}[\theta \mid \mu(e)]$, and strictly positive payoffs if it hires him at less than $\mathbb{E}[\theta \mid \mu(e)]$. Since firms are competing in Bertrand price competition, the unique mutual best responses are $w(e)$ as defined above. What can be said about $w(e)$? At this stage, not much, except that in any equilibrium, for all e, $w(e) \in [\theta_L, \theta_H]$ because $\mu(e) \in [0, 1]$. Note that in particular, we cannot say that $w(\cdot)$ even need be increasing. Another point to

[3] Strictly speaking, he can reject if the wage is exactly 0, but we resolve indifference in favor of acceptance, for simplicity. This is not important.

note is that there is a one-to-one and onto mapping from $\mu(e) \in [0,1]$ to $w(e) \in [\theta_L, \theta_H]$. Remember that equilibrium requires $\mu(e)$ to be derived in accordance with Bayes' rule applied to the worker's education choices, whenever possible.

- *Stage 2.* To study the worker's choice of education, we must consider his preferences over wage-education pairs. To this end, consider the utility from acquiring education e and then receiving a wage w, for type θ in (10). To find indifference curves, we set $u(w, e, \theta) = \overline{u}$ for some $\overline{u} \in \mathbb{R}$ and implicitly differentiate, obtaining the marginal rate of substitution between wages and education at any given (w, e) pair:

$$\left.\frac{dw}{de}\right|_{u=\overline{u}} = c_e(e, \theta) > 0.$$

Thus, indifference curves are upward sloping in $e-w$ plane, and moreover, at any particular (e, w), they are steeper for θ_L than for θ_H by the assumption that $c_{e\theta} < 0$. Therefore, an indifference curve for type θ_H crosses an indifference curve for type θ_L at most once. This is known as the *(Spence-Mirrlees) single crossing property*, and it plays a key role in the analysis of many signaling models. [Figure A in class! or Figure 13.C.2 page 453 in Mas-Colell et al. (1995)] shows a graphical representation.[4]

Obviously, the choice of e for a worker of either type will depend on the wage function $w(e)$ from stage 3. But in turn, the function $w(e)$ (or equivalently, $\mu(e)$) must be derived from Bayes' rule for any e that is chosen by the worker (of either type). For any e not chosen by either type, any $w(e) \in [\theta_L, \theta_H]$ is permissible since Bayes' rule does not apply. As usual in signaling games, this flexibility in specifying off-the-equilibrium-path wages yields a multiplicity of equilibria.

2.3 Separating Equilibria

Here, the two types of worker choose different education levels, thus "separating" themselves. Let $e^*(\theta)$ denote the equilibrium education choice of the worker of type $\theta \in \{\theta_L, \theta_H\}$, and $w^*(e)$ denote an equilibrium wage offer given the equilibrium beliefs $\mu(e)$.

Claim 1. *In any separating equilibrium, $w^*(e^*(\theta)) = \theta$ for $\theta \in \{\theta_L, \theta_H\}$, i.e. each worker type receives a wage equal to his marginal productivity.*

Proof. By definition, in a separating equilibrium, the two types choose different education levels, call them $e^*(\theta_L) \neq e^*(\theta_H)$. Bayes' rule applies on the equilibrium path, and implies that $\mu(e^*(\theta_L)) = 0$ and $\mu(e^*(\theta_H)) = 1$. By substituting each of these beliefs into equation (11), we have that the resulting wages are θ_L and θ_H, respectively. ■

Claim 2. *In any separating equilibrium, $e^*(\theta_L) = 0$, i.e. the low productivity worker chooses not to get education.*

[4] In Figure A, $IC(\theta) := \{(e, w) \in \mathbb{R}^2_+ : u(w, e, \theta) = u(\hat{w}, \hat{e}, \theta)\}$ for $\theta \in \{\theta_L, \theta_H\}$.

Proof. By way of contradiction, suppose that $e^*(\theta_L) > 0$. By Claim 1, type θ_L's equilibrium utility is $\theta_L - c(e^*(\theta_L), \theta_L)$. If instead θ_L chose education level 0, he would receive a utility of at least $\theta_L - c(0, \theta_L)$, because $w^*(0) \geq \theta_L$. Since $c_e(e, \theta_L) > 0$ for all e, it follows that the worker of type θ_L gets a strictly higher utility by deviating to an education level of 0, a contradiction with equilibrium play. ■

Claims 1 and 2 combined imply that the equilibrium utility for a low type is $u(\theta_L, 0, \theta_L) = \theta_L$. This puts the low type on the indifference curve passing through the point $(0, \theta_L)$ in the e-w plane, as drawn in [Figure B in class! or Figure 13.C.4 page 454 in Mas-Colell et al. (1995)].

We can use this picture to construct a separating equilibrium. By Claim 1, the θ_H worker must receive a wage of θ_H, hence an allocation somewhere on the horizontal dotted line at θ_H. If the allocation were to the left of where IC(θ_L) crosses that dotted line at θ_H, then type θ_L would prefer to mimic the θ_H worker rather than separate (i.e. it would prefer to choose the education level that θ_H is supposed to, rather than 0); this follows from the fact that allocations to the left of a given indifference curve are more desirable. So, a candidate allocation for the high type is education level $\hat{e}$ with wage θ_H in [Figure C in class! or Figure 13.C.5 page 454 in Mas-Colell et al. (1995), where $\tilde{e}$ is our $\hat{e}$]. That is, we set $e^*(\theta_H) = \hat{e}$ and $w^*(\hat{e}) = \theta_H$ where $\hat{e}$ is formally the solution to

$$u(\theta_H, \hat{e}, \theta_L) = u(\theta_L, 0, \theta_L).$$

In words, $\hat{e}$ is the education level that makes type θ_L indifferent between acquiring $\hat{e}$ with wage θ_H and acquiring 0 with wage θ_L.

It remains only to specify the wage schedule $w^*(e)$ at all points $e \notin \{0, \hat{e}\}$. Since we are free to specify any $w^*(e) \in [\theta_L, \theta_H]$, consider the one that is drawn in [Figure C in class! or Figure 13.C.5 page 454 in Mas-Colell et al. (1995)].

Given this wage schedule, it is clear that both types are playing optimally by choosing 0 and $\hat{e}$ respectively, i.e. neither type strictly prefers choosing any other education level and receiving the associated wage over its prescribed education and associated wage. To see this, note that, for each type that he may be, the worker's indifference curve is at its highest possible level along the schedule $w^*(e)$. Moreover, beliefs (or wages) are correct on the equilibrium path, and thus firms are playing optimally. Thus, e^* and w^* as defined is in fact a separating equilibrium. It is obvious that there are various wage schedules that can support the same equilibrium education choices (since firms can have a variety of beliefs off the equilibrium path): an alternate schedule that works, for example, is $w(e) = \theta_L$ for all $e \in [0, \hat{e})$ and $w(e) = \theta_H$ for all $e \geq \hat{e}$ (for an illustration, see Figure 13.C.6 page 455 in Mas-Colell et al. (1995), where $\tilde{e}$ is our $\hat{e}$).

The more interesting question is whether there are other education levels that can be sustained in a separating equilibrium. Claim 2 says that the low type must always play $e^*(\theta_L) = 0$, but could we vary $e^*(\theta_H)$? Yes. Let $\bar{e}$ be the education level that solves

$$u(\theta_H, \bar{e}, \theta_H) = u(\theta_L, 0, \theta_H).$$

In words, $\bar{e}$ is the education level that makes type θ_H indifferent between acquiring $\bar{e}$ with wage θ_H and acquiring 0 with wage θ_L. The single-crossing property stemming from $c_{e\theta}(e, \theta) < 0$ for all e and all θ ensures that $\bar{e} > \hat{e}$. A separating equilibrium where $e^*(\theta_H) = \bar{e}$ is illustrated in [Figure D in class! or Figure 13.C.7 page 455 in Mas-Colell et al. (1995), where e_1 is our $\bar{e}$].

It follows from the construction logic that for every $e \in [\hat{e}, \bar{e}]$, there is a separating equilibrium with $e^*(\theta_H) = e$ (make sure you understand this). Moreover, there is no separating equilibrium with $e^*(\theta_H) \notin [\hat{e}, \bar{e}]$. As explained above, if $e < \hat{e}$ is an equilibrium education level for high type, then the low type is better off by deviating and pretending to be of high type by getting education e. Similarly, any education level $e > \bar{e}$ for the high type cannot be sustained in equilibrium. This is because at any education level greater than $\bar{e}$, the high type can deviate and get zero education level and pretend to be of low type. That will push his indifference curve "up", and hence will lead to a profitable deviation.

Remark 5. The fundamental reason that education can serve as a signal here is that the marginal cost of education depends on the worker's type. Because the marginal cost of education is higher for a low-ability worker (since $c_{e\theta}(e, \theta) < 0$), a worker of type θ_H may find it worthwhile to get some positive level of education $e > 0$ to raise his wage by some amount $\Delta w > 0$, whereas a worker of type θ_L may be unwilling to get this same level of education in return for the same wage increase. As a result, firms can reasonably come to regard the education level as a signal of worker quality.

It is easy to see that we can Pareto-rank separating equilibrium allocations.

Proposition 1. *A separating equilibrium with $e^*(\theta_H) = e_1$ Pareto-dominates a separating equilibrium with $e^*(\theta_H) = e_2$ if and only if $e_1 < e_2$.*

Proof. Straightforward, since firms are making 0 expected profit in any separating equilibrium, the low type receives a payoff equal to θ_L in any separating equilibrium, and the high type of worker prefers acquiring less education to more at the same wage θ_H. ■

Therefore, the separating equilibria with $e^*(\theta_H) = \hat{e}$ Pareto-dominate all other separating equilibria, whereas the separating equilibria with $e^*(\theta_H) = \bar{e}$ are Pareto-dominated by all other separating equilibria.

Next, compare the welfare of these equilibria to the case where signaling is not possible. In particular, comparing the separating equilibria to the situation with no signaling opportunity:

- The two firms are equally well off, as in both cases they earn zero expected profit.
- The worker of type θ_L is worse off with signaling: in both cases θ_L incurs no education costs, but when signaling is possible, he receives a wage of θ_L rather than $\mathbb{E}(\theta)$.
- The payoffs of the worker of type θ_H may go either way. Consider two scenarios. First, suppose that $\mathbb{E}(\theta)$ is low compared to θ_H; then a high type worker gets in equilibrium a wage, θ_H, which offsets his cost of education by an amount sufficient to increase the payoff from no signaling. On the other hand, consider the case when $\mathbb{E}(\theta)$ is close to θ_H. Then, signaling does not lead to a major wage increase. Further, there is the extra cost of education, and this brings the payoff down for the high type worker. The problem here is

that when signaling is present, if the high type chooses zero education level, he is treated as low type and given θ_L wage. So, he is forced to get costly education even though it reduces his welfare from no signaling. Note that high value of $\mathbb{E}(\theta)$ means high value of λ (greater probability of high type worker). Hence, a separating equilibrium such that the worker of type θ_H is better off than without signaling exists if and only if λ is sufficiently small. For an illustration, see Figure 13.C.8 page 456 in Mas-Colell et al. (1995).

None of the separating equilibria yields an efficient outcome in the Spence model with unproductive education. Hence, an opportunity to signal can create inefficiencies in an otherwise efficient market (recall Remark 4). In a more general model where the worker's reservation wage is not zero but depends on his type, the welfare effect of a signaling opportunity may become ambiguous (see Exercise 5).

2.4 Pooling Equilibria

Here, the two types of worker choose the same education level, thereby "pooling" together. That is, we are looking for equilibria where $e^*(\theta_L) = e^*(\theta_H) = e^P$ for some $e^P \in \mathbb{R}_+$. In any such equilibrium, Bayes' rule implies that $\mu(e^P) = \lambda$; hence, from equation (11), $w^*(e^P) = \mathbb{E}(\theta)$. It only remains to determine which e^P are feasible in a pooling equilibrium. Define $\underline{e}$ as the education level that makes type θ_L indifferent between acquiring education $\underline{e}$ with wage $\mathbb{E}(\theta)$ and acquiring education 0 with wage θ_L. Formally, $\underline{e}$ is the solution to

$$u(\mathbb{E}(\theta), \underline{e}, \theta_L) = u(\theta_L, 0, \theta_L).$$

Since $\mathbb{E}(\theta) \in (\theta_L, \theta_H)$, it follows that $\underline{e} \in (0, \hat{e})$. [Figure E in class! or Figure 13.C.10 page 457 in Mas-Colell et al. (1995), where e' is our $\underline{e}$] shows the construction of a pooling equilibrium with $e^P = \underline{e}$. Given this wage schedule, both worker types are playing optimally by choosing $e^P = \underline{e}$ as their indifference curve is at its highest possible level along the schedule. Moreover, the wage schedule is consistent with Bayes' rule on the equilibrium path. Of course, there are multiple wage schedules that can support this pooling choice of $e^P = \underline{e}$.

The construction logic implies that there is a pooling equilibrium for any $e^P \in [0, \underline{e}]$, but not for $e^P > \underline{e}$. The reason for the latter is that a worker of type θ_L would strictly prefer to choose education 0 and get and $w^*(0) \geq \theta_L$ rather than choose $e^P > \underline{e}$ and get wage $\mathbb{E}(\theta)$. [Figure F in class! illustrates a pooling equilibrium with $e^P = 0$.]

We can also Pareto-rank the pooling equilibria.

Proposition 2. *A pooling equilibrium with education level e^P Pareto-dominates a pooling equilibrium with education level $\tilde{e}^P$ if and only if $e^P < \tilde{e}^P$.*

Proof. Straightforward, since firms are making 0 expected profit in any pooling equilibrium, both types of of worker receive the same wage in any pooling equilibrium, and both type of workers strictly prefer acquiring lower education levels for a given wage. ■

Therefore, the pooling equilibria with $e^P = 0$ Pareto-dominates all other pooling equilibria, and the pooling equilibria with $e^P = \underline{e}$ are Pareto-dominated by all other pooling equilibria.

Note also that any pooling equilibrium with $e^P > 0$ is completely wasteful in the sense that both types of worker would be better off if the ability to signal had been absent altogether, and the market just functioned with no education acquisition and a wage rate of $\mathbb{E}(\theta)$. Thus, pooling equilibria are (weakly) Pareto dominated by the no-signaling outcome.

2.5 Applying the Intuitive Criterion

The latitude in selecting $w^*(e)$ for all e that are not chosen in equilibrium is what leads to the multiplicity of equilibria. To refine the set of equilibria, we can apply the Intuitive Criterion. The reasoning is exactly the same as in Section 1. In the job market signaling game, a pure PBE $\left\{(e^*(\cdot), w^*(\cdot)), \{\mu(e)\}_{e\in\mathbb{R}_+}\right\}$ satisfies the Intuitive Criterion if $\mu(e) = 1$ for any $e \in \mathbb{R}_+$ such that:

1. e is not chosen by either type in the equilibrium;

2. For all $w \in [\theta_L, \theta_H]$, $u(w^*(e^*(\theta_L)), e^*(\theta_L), \theta_L) > u(w, e, \theta_L)$;

3. For some $w \in [\theta_L, \theta_H]$, $u(w^*(e^*(\theta_H)), e^*(\theta_H), \theta_H) < u(w, e, \theta_H)$.

Condition 2 is the key and says that e is equilibrium dominated for θ_L: that is, type θ_L gets strictly higher utility in the equilibrium than any $w \in [\theta_L, \theta_H]$ it could get in return for choosing the out-of-equilibrium education e. Condition 3 says that e is not equilibrium dominated for θ_H; that is, there is some $w \in [\theta_L, \theta_H]$ that would make type θ_H prefer acquiring e if it received w in return, relative to what it gets in equilibrium. Note that it is sufficient to check this condition using the most attractive wage, i.e. $w = \theta_H$.

It turns out that the Intuitive Criterion is very powerful in the job market signaling game with 2 types of worker, as the following proposition shows.

Proposition 3. *The only signaling equilibria (amongst both pooling and separating) that satisfy the Intuitive Criterion are the separating equilibria with* $e^*(\theta_H) = \hat{e}$.

Proof. See Exercise 7. ■

Thus, the application of the Intuitive Criterion yields a unique equilibrium outcome (i.e. the equilibria that survive can only differ in off-the-equilibrium path wages), which is the Pareto-efficient separating equilibrium.

2.6 Exercises

Exercise 3 (Verifiable Disclosure)**.** Consider a game in which, first, nature draws a worker's type (ability) from some continuous distribution on $[\underline{\theta}, \overline{\theta}]$, where $\overline{\theta} > \underline{\theta} > 0$. Once the worker observes his type, he can choose whether to submit to a costless test that perfectly reveals his ability. Finally, after observing whether the worker has taken the test and its outcome if he has, two firms bid for the worker's services. Prove that in ant subgame perfect Nash equilibrium of this model all worker types submit to the test, and firms offer a wage no greater than $\underline{\theta}$ to any worker not doing so.

Solution. This is Exercise 13.C.1 in Mas-Colell et al. (1995).

Exercise 4 (Productive Education)**.** Consider the job market signaling model under the same assumptions as above, except that now education is productive. Specifically, if the worker's type is θ and his education level is e, then his productivity when employed in a firm is $\theta(1 + \alpha e)$ for some $\alpha > 0$.

(a) Find the perfect information competitive outcome.

(b) Find all pure-strategy PBEs and relate them to the perfect information competitive outcome.

Solution. This is Exercise 13.C.2 in Mas-Colell et al. (1995).

Exercise 5 (Adverse Section and Signaling)**.** Consider the job market signaling model under the same assumptions as above, except that now the reservation wage for both types of worker is $r > 0$. Assume $\theta_L < r < \theta_H$ and $\mathbb{E}[\theta] < r$.

(a) What is the efficient outcome?

(b) Suppose acquiring education is not possible. What is the competitive market equilibrium.

(c) Find all pure-strategy PBEs. Pareto-rank these equilibria and relate them to the outcome arising without signaling opportunities.

Solution. See Mas-Colell et al. (1995) pages 459–460.

Exercise 6 (Three Types)**.** Consider an extension of the job market signaling model to an environment with three types. Provide an example in which more than one PBE satisfies the Intuitive Criterion.

Solution. See Cho and Kreps (1987). In the paper, the authors use the notion of "Riley outcome" to refer to the separating equilibria with $e^*(\theta_H) = \hat{e}$ (i.e. to the separating equilibria that Pareto-dominate all other separating equilibria). The example itself appears under the subtitle "Case B. More Than Two Types" on pages 212–214.

Exercise 7. Prove Proposition 3. [You are encouraged to use graphical arguments, provided that you explain your figures.]

Solution. We briefly discussed it in class. See also Mas-Colell et al. (1995) pages 457–458 and 470–471.

References

Banks, Jeffrey S. and Joel Sobel (1987), "Equilibrium Selection in Signaling Games." *Econometrica*, 55, 647–661.

Cho, In-Koo and David M. Kreps (1987), "Signaling Games and Stable Equilibria." *The Quarterly Journal of Economics*, 102, 179–221.

Crawford, Vincent P. and Joel Sobel (1982), "Strategic Information Transmission." *Econometrica*, 50, 1431–1451.

Fudenberg, Drew and Jean Tirole (1991a), "Game Theory." *MIT Press.*

——— (1991b), "Perfect Bayesian Equilibrium and Sequential Equilibrium." *Journal of Economic Theory*, 53, 236–260.

Gilligan, Thomas W. and Keith Krehbiel (1989), "Collective Choice without Procedural Commitment." Peter C. Ordeshook ed. *Models of Strategic Choice in Politics.* University of Michigan Press.

Hörner, Johannes (2008), "Signalling And Screening." *The New Palgrave Dictionary of Economics.*

Kartik, Navin (2009), "Lecture Notes for 1st Year Ph.D. Game Theory." *Lecture Notes.*

Kohlberg, Elon and Jean-François Mertens (1986), "On the Strategic Stability of Equilibria." *Econometrica*, 54, 1003–1037.

Kreps, David M. and Joel Sobel (1994), "Signaling." Robert J. Aumann and Sergiu Hart eds., *Handbook of Game Theory with Economic Applications*, 849–867.

Mailath, George J. (2019), "Modeling Strategic Behavior: A Graduate Introduction to Game Theory and Mechanism Design." *World Scientific.*

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green (1995), "Microeconomic Theory." *Oxford University Press.*

Mishra, Debasis (2010), "Lecture Notes on Adverse Selection and Signaling." *Lecture Notes.*

Rey, Patrick (2018), "Microeconomics 2." *Lecture Notes.*

Riley, John G. (2001), "Silver Signals: Twenty-Five Years of Screening and Signaling." *Journal of Economic Literature*, 39, 432–478.

Spence, Michael (1973), "Job Market Signaling." *The Quarterly Journal of Economics*, 87, 355–373.

——— (1974), "Market Signaling." *Harvard University Press.*