LECTURE NOTES ON ADVERSE SELECTION AND SIGNALING

Debasis Mishra[*]

April 5, 2010

1 INTRODUCTION

In general competitive equilibrium theory, it is assumed that the characteristics of the commodities are observable to the firms and consumers. The objective of this section is to relax this *complete markets* assumption. In practice, there are many scenarios where the information is asymmetrically distributed in a market. We give some examples to illustrate this.

1. When a firm hires a worker (a University hires a doctoral student etc.), the firm may know less than the worker about his innate ability.
2. When an insurance firm offers a health insurance to an individual, the individual knows about his health and exercising habits more than the firm.
3. In the used-car market, the seller of a car may have more information about the car than the buyer.

A number of questions immediately arise in such settings.

1. How do we characterize market equilibria in markets with asymmetric information?
2. What are the properties of these equilibria?
3. Are there possibilities for market to intervene and improve welfare?

---

[*] Planning Unit, Indian Statistical Institute, 7 Shahid Jit Singh Marg, New Delhi 110016, India, E-mail: `dmishra@isid.ac.in`

2 COMPETITIVE EQUILIBRIUM WITH INFORMATIONAL ASYMMETRIES

We introduce the following model, similar to Akerlof's "market for lemons" model. There are two types of agents in the markets.

- FIRMS: There are many identical firms that can hire workers. Each produces the same output using identical **constant returns to scale** technology in which labor is the only input. Each firm is risk neutral, seeks to maximize its profit, and acts as a price-taker. For simplicity, assume that price of every firm's output is 1.
- WORKERS: There are N workers. Workers differ in the number of units of output they can produce if hired by a firm. This is called the **type** of a worker, and is denoted by $\theta \in \mathbb{R}_+$. Further, the set of all possible types lies in an interval $\Theta = [\underline{\theta}, \bar{\theta}]$, where $0 \leq \underline{\theta} < \bar{\theta} < \infty$. The proportion of workers with type θ or less is given by the distribution function $F(\theta)$, which is assumed to be non-degenerate (i.e., $F(\theta) < 1$ for all $\theta < \bar{\theta}$).

  Each worker wants to maximize the amount he earns (wage) from labor. Define the home production function of each worker as $r : \Theta \rightarrow \mathbb{R}$. If a worker of type θ decides to stay home, he earns $r(\theta)$ (unit price of 1 for simplicity). So, $r(\theta)$ is the opportunity cost of worker of type θ of accepting employment. Hence, a worker accepts employment in a firm if and only if his wage is at least $r(\theta)$.

In a competitive market, each type of worker is a commodity. So, in competitive equilibrium, there is an equilibrium wage $w^*(\theta)$ for worker of type θ. Given the competitive and constant returns to scale nature of firms, in a competitive equilibrium, we have $w^*(\theta) = \theta$ for all $\theta \in \Theta$, and the set of workers who accept employment is given by $\{\theta : r(\theta) \leq \theta\}$.

We verify that such a competitive equilibrium is Pareto efficient. This follows from the first welfare theorem, but can be verified directly. Recall that Pareto optimality maximizes the aggregate surplus. Here the surplus is revenue generated by workers' labor. A type θ worker gets a revenue of θ if he gets employed and gets $r(\theta)$ from home production. For all $\theta \in \Theta$, let $x(\theta) \in \{0, 1\}$ be a binary variable denoting if the worker is employed (value 1) or not (value zero). So, the aggregate surplus can be maximized by maximizing the following expression (expected total revenue).

$$\int_{\underline{\theta}}^{\bar{\theta}} N[\theta x(\theta) + r(\theta)(1 - x(\theta))]dF(\theta).$$

Clearly, this is maximized by setting $x(\theta) = 1$ for all $\theta \geq r(\theta)$ and setting $x(\theta) = 0$ for all $\theta < r(\theta)$. Hence, in any Pareto optimal allocation the set of workers that are employed by firms must be $\{\theta : \theta \geq r(\theta)\}$.

2.1 Unobservable Types of Workers

We now develop a definition of competitive equilibrium, when workers' type is not observable. Since workers' type is not observable, the wage offered to all the workers must be same - say w. So, the set of types workers who will get employment at this wage is given by

$$\Theta(w) = \{\theta : r(\theta) \leq w\}.$$

Now, we need to determine the demand function. If the average type of workers who accept employment is μ, its demand for labor is given by: $z(w) = 0$ if $\mu < w$; $z(w) = [0, \infty]$ if $\mu = w$; and $z(w) = \infty$ if $\mu > w$.

So, if workers of type Θ^* accept employment, then firm's belief about the average type of these workers must be correctly reflected in equilibrium. Hence, we must have $\mu = E[\theta) : \theta \in \Theta^*]$. This way, demand for labor can equal supply if and only if $w = E[\theta : \theta \in \Theta^*]$. Note however, that the expectation is not well defined if $\Theta^* = \emptyset$. We ignore this case and focus on equilibria where trade takes place.

Definition 1 *In the competitive labor market model with unobservable worker types, a* **competitive equilibrium** *is a wage w^* and a set Θ^* of worker types who accept employment such that*

$$\begin{aligned} \Theta^* &= \{\theta : r(\theta) \leq w^*\} \\ w^* &= E[\theta : \theta \in \Theta^*] \text{ if } \Theta^* \neq \emptyset. \end{aligned}$$

This involves *rational expectations* on the part of the firm, i.e., a firm must correctly anticipate the average type of workers accepting employment.

This type of competitive equilibrium will fail to be Pareto optimal. We first consider a simple setting where $r(\theta) = r$ for all $\theta \in \Theta$ and $F(r) \in (0, 1)$. The Pareto optimal allocation is that workers with type $\theta \geq r$ accept employment and those with type $\theta < r$ not accepting employment.

Now, consider the competitive equilibrium. If $r(\theta) = r$, the set of workers who accept employment at a given wage w is given by $\Theta(w) = \Theta$ if $r \leq w$ and $\Theta(w) = \emptyset$ if $r > w$. In either case, $E[\theta : \theta \in \Theta(w)] = E[\theta]$. So, equilibrium wage should be $E[\theta]$. If $E[\theta] \geq r$, all workers accept employment and otherwise none do so. We consider both cases:

Case 1: Suppose $E[\theta] \geq r$. Then, every worker accepts employment. If $F(r)$ is low, then Pareto optimality will require that many workers do not accept employment. So, if the average type of the workers is high (because of large enough set of good workers), then everyone gets to work and bad workers are also selected.

Case 2: No worker receives employment. However, if $F(r)$ is high, then Pareto optimality will require that many workers accept employment. So, if the average type of the workers is low (because of large enough set of bad workers), then nobody gets to work and good workers are left out.

2.2 Adverse Selection

If $r(\cdot)$ is no longer a constant, then this may exaggerate to a phenomenon known as *adverse selection* occurs. **Adverse selection** is said to occur when an informed individual's trading decision depends on her unobservable characteristics in a manner that adversely affects the uninformed agents in the market. In the labor market context, adverse selection arises only relatively less capable workers accept a firm's employment offer at *any* given wage.

From our illustration in last section, it seems that adverse selection may happen if there are some workers who should be employed and some who should not be. As, we illustrate now, a market may collapse when everyone should be working. Suppose $r(\theta) \leq \theta$ for all θ and $r(\cdot)$ is a strictly increasing function. The first assumption implies that at a Pareto optimal allocation, every worker must be employed in some firm. The second assumption implies that workers who are more productive at firm are also more productive at home. It is this assumption that drives adverse selection: at a given wage w, since the payoff to a more capable worker is greater at home, he prefers staying at home whereas the less capable worker joins the firm.

By the equilibrium condition the equilibrium wage can be determined by the following equation:

$$w^* = E[\theta : r(\theta) \leq w^*].$$

Figure 1 illustrates adverse selection. We have assumed $\underline{\theta} = 1$ and $\bar{\theta} = 4$. The left graph depicts $E[\theta : r(\theta) \leq w]$ as a function of w and the right graph depicts $r(\theta)$ as a function of θ. Focus on the left graph. Note that $E[\theta : r(\theta) \leq r(\underline{\theta})] = \underline{\theta}$ and $E[\theta : r(\theta) \leq w]$ for any $w \geq r(\bar{\theta})$ is $E[\theta]$. The equilibrium wage w^* is obtained from this graph. Taking the corresponding point in the right graph, we obtain the cut-off for the worker to be employed, and see that a large portion of workers may be unemployed, even though Pareto efficiency requires that everyone must be employed.

We now give an example to illustrate the collapse of the market. Let $\Theta = [0, 2]$ and $r(\theta) = \alpha\theta$ for all $\theta \in [0, 2]$, for some $\alpha < 1$. Suppose θ is distributed uniformly in $[0, 2]$. Then, $E[\theta : \alpha\theta \leq w] = \frac{w}{2\alpha}$. So, the equilibrium wage is $w^* = 0$ and only workers of type 0 accept employment but nobody else. However, in a Pareto optimal allocation, everyone should be employed since $\theta \geq r(\theta)$.

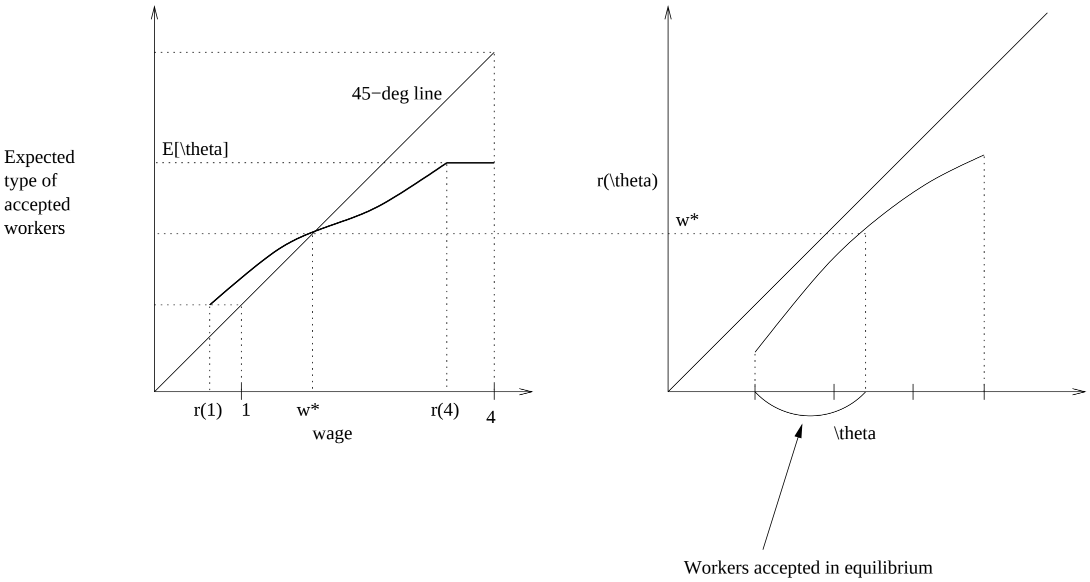

Figure 1: Adverse Selction

The competitive equilibrium need not not be unique. This is because of the fact that the curve $E[\theta : r(\theta) \leq w]$ may have any shape. However, in each equilibrium, the firms must earn zero profit. However, wages in each equilibrium is different, implying that these equilibria can be *Pareto ranked* - firms prefer higher wage equilibria. The low wage Pareto dominated equilibria exist because of coordination failure. Firms expect worker type to be low and offer low wage, and as a result low type workers only get selected. If firms knew that good type workers can be attracted by offering high wage rates then they would have done so.

3 Game Theoretic Analysis of Adverse Selection

In this section, we ask the question if the type of competitive equilibria achieved in the adverse selection model can be viewed as an outcome of a richer model in which firms engage in strategic wage offerings.

The situation with multiple equilibria may signal some concerns in this regard. Consider the situation in Figure 2. For example, if firms were strategic, then they can increase payoff at w_2^* equilibrium by making a slightly larger wage offer.

We now analyze a simple 2-stage game. For simplicity, let there be two firms, say 1 and 2. The functions $F(\cdot)$ and $r(\cdot)$ are common knowledge. The game proceeds as follows:

- Stage 1: Firms announce their wages w_1 and w_2.
- Stage 2: Each worker decides either (a) to stay at home or (b) to join firm 1 or (c)

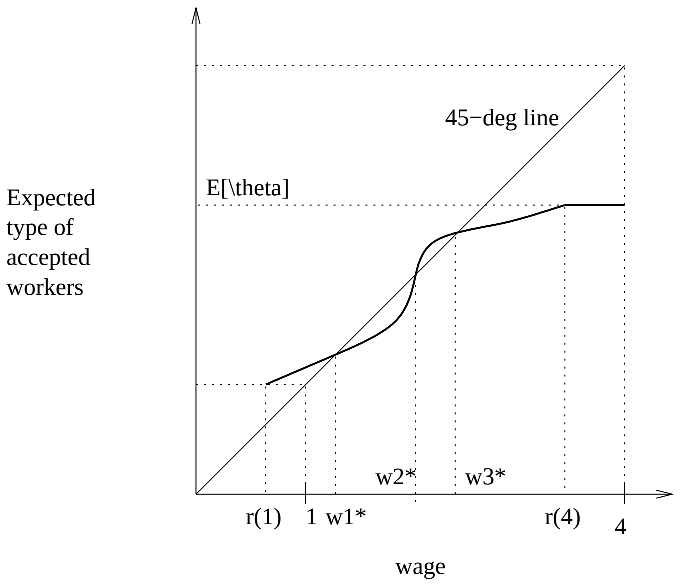

Figure 2: Multiple Equilibria in Adverse Selction

to join firm 2 (if a worker is indifferent about joining either firm, then it joins each of them with probability $\frac{1}{2}$).

The following result characterizes the subgame perfect Nash equilibria (SPNEs) of this game for the adverse selection model when $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all $\theta \in \Theta$ and $F(\cdot)$ has an associated continuous density function $f(\cdot)$ with $f(\theta) > 0$ for all $\theta \in \Theta$.

Theorem 1 *Let W^* denote the set of competitive equilibrium wages for the adverse selection model, and let $w^* = \max\{w : w \in W^*\}$. Suppose $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all $\theta \in \Theta$ and $E[\theta : r(\theta) \leq w]$ is continuous in all w.*

1. *If $w^* > r(\underline{\theta})$ and there is an $\epsilon > 0$ such that $E[\theta : r(\theta) \leq w'] > w'$ for all $w' \in (w' - \epsilon, w^*)$, then there is a unique pure strategy SPNE of the two-stage game theoretic model. In this SPNE, each employed worker receives a wage of w^*, and workers with types in the set $\Theta(w^*) = \{\theta : r(\theta) \leq w^*\}$ accept employment in the firms.*
2. *If $w^* = r(\underline{\theta})$, there are multiple pure strategy SPNEs. However, in every pure strategy SPNE each agent's payoff exactly equals her payoff in the highest-wage competitive equilibrium.*

Proof: Note that in any SPNE, a worker of type θ must accept an employment offer if and only if it is at least $r(\theta)$ and he must accept the firm's offer which is the highest above $r(\theta)$ (breaking ties symmetrically). Also, note that the point w^* is well-defined. This is because, for $w \geq r(\bar{\theta})$, $E[\theta : r(\theta) \leq w] = E[\theta]$. By continuity of $E[\theta : r(\theta) \leq w]$ with respect to w, we conclude that w^* is well-defined.

Now, we need to determine the equilibrium behavior of the firm. We consider two possible cases.

Case 1: Suppose $w^* > r(\underline{\theta})$. We derive a firm's equilibrium behavior in several steps.

Step 1: We first show that the wage offered by both firms must be the same and both firms must be hiring in any SPNE. To this end, note that if both firms are not attracting any worker (i.e, $\max(w_1, w_2) < r(\underline{\theta})$) then any firm can make positive payoff by offering an wage $w^* - \epsilon$ for some $\epsilon > 0$ since $E[\theta : r(\theta) \leq (w^* - \epsilon)] - (w^* - \epsilon) > 0$ and since $w^* > r(\underline{\theta})$. If one of the firms, say firm 1, is offering wage w_1 and not attracting any worker and w_2 is a wage where firm 2 is attracting workers, then there are two subcases: (a) firm 2 is making zero payoff - in this case it can deviate to $w^* - \epsilon$ and make positive payoff, and (b) firm 2 is making positive payoff - in this case firm 1 can deviate to this wage and enjoy positive payoff. So, in any SPNE, both firms must be hiring workers. Then, if wages offered by both firms are different, then all workers will choose the firm with the higher wage. Hence, it cannot be a SPNE that both firms offer different wages.

Step 2: Next, we show that in any SPNE, both firms must earn exactly zero. To see this, suppose there is an SPNE in which a total of M workers are hired at a wage w. Let the aggregate earning of both firms be

$$\Pi = M(E[\theta : r(\theta) \leq w] - w).$$

Assume for contradiction $\Pi > 0$, which implies that $M > 0$. This further implies that $w \geq r(\underline{\theta})$. In this case, the weakly less profitable firm, say firm 1, must be earning no more than $\frac{\Pi}{2}$. But firm 1 can earn profits of at least $M(E[\theta : r(\theta) \leq (w + \alpha)] - (w + \alpha))$ by increasing the wage to $(w + \alpha)$ for $\alpha > 0$. Since $E[\theta : r(\theta) \leq w]$ is continuous in w, we can choose α small enough such that this profit is made arbitrarily close to Π. Thus, firm 1 will be better off deviating. So, in any SPNE the wage w chosen by firms must belong to W^*.

Step 3: We conclude by arguing that in any SPNE the wage chosen must be w^*. If both firms offer wage w^*, then no firm has an incentive to offer a lower wage since he will not be able to attract any workers. Also, if a firm offers any higher wage he gets a payoff of $E[\theta : r(\theta) \leq w] - w$, where $w > w^*$. But note that $E[\theta : r(\theta) \leq \bar{\theta}] = E[\theta]$, which is assumed to be finite. Since w^* is the maximum point in W^*, the sign of the expression $E[\theta : r(\theta) \leq w] - w$ is the same for all $w > w^*$. If this sign is positive then the curve $E[\theta : r(\theta) \leq w]$ cannot cross the 45 degree line for all values of $w > w^*$. This is a contradiction to the fact that the slope of the curve is zero after $w \geq \bar{\theta}$.

Finally, suppose there is some $w \neq w^*$ which is a SPNE. By our assumption, $w \in W^*$. Hence, $w < w^*$. In that case, one of the firms is better of by choosing a wage arbitrarily close to w^* and making strictly positive payoff. This gives us the desired contradiction.

This completes the argument for Case 1.

CASE 2: Suppose $w^* = r(\underline{\theta})$. As argued previously, any wage offer $w > w^*$ gives a firm negative payoff since $E[\theta : r(\theta) \leq w] - w < 0$ for all $w > w^*$. Further a firm earns exactly zero by announcing any wage $w \leq w^*$. So any wage pair (w_1, w_2) such that $\max(w_1, w_2) \leq w^*$ is a SPNE. In every such SPNE, every worker of type θ earns $r(\theta)$ and firms earn zero. ■

A key difference between the game theoretic model and the competitive equilibrium model is the information that firms require to have. In the competitive equilibrium model, firms only need to know the average productivity of employed workers. However, the game theoretic model they need to know the underlying market mechanism, in particular the relationship between wage offered and quality of employed workers. The game theoretic model tells us that if such sophistication is possible for the firms, then the coordination problem that can arise in the competitive equilibrium model may disappear.

4 MARKET INTERVENTIONS

We saw that asymmetric information between agents can lead to competitive equilibria that are not Pareto optimal. If a central authority knew all the information of the agents, then it can intervene and improve these outcomes by defining some lump-sum transfers among agents. In practice, however, it is difficult to believe why the central authority will have more information than the market participants. In that case, such transfers between agents is difficult to execute - for example, if the types of workers are not known, it is difficult to define who transfers to whom.

DEFINITION **2** *An allocation that cannot be Pareto improved by an authority who is unable to observe agents' private information is known as* **constrained Pareto optimum**.

Clearly, a constrained Pareto optimum allocation need not be Pareto optimum. But a Pareto optimum allocation is necessarily a constrained Pareto optimum allocation. We now study if a constrained Pareto optimum competitive equilibrium exists in the adverse selection model.

4.1 WHAT KIND OF MARKET INTERVENTIONS?

A market intervention involves lump-sum transfers amongst agents. We want to verify if we can improve a competitive equilibrium allocation. If a Pareto improvement is possible where the firm receives positive payoff, then a Pareto improvement is also possible by increasing the wage to drive the firm's payoff to zero and increasing the payoff of workers. Hence, without loss of generality we will verify if a Pareto improvement is possible where transfers are between workers. To this end, we can assume that the central authority takes charge of firms and introduces transfers. In this case, note that the firms get zero profit.

Next, since the authority does not know the type of workers, any transfer must depend only on the fact whether the worker is employed or not. Thus, there should be no loss of generality if we assume that the market intervention is by the central authority taking charge of firms and offering wages w_e to employed workers and w_u to unemployed workers.

4.2 COMPETITIVE EQUILIBRIUM AND CONSTRAINED PARETO OPTIMA

We first consider any competitive equilibrium which is not Pareto dominated, i.e., all competitive equilibrium wages which is not equal to w^* (the maximum one). In this case, the authority can set a $w_e = w^*$ and $w_u = 0$ to improve the payoff of workers while maintaining the expected payoff of firms at zero. Hence, a Pareto improvement is possible. So, any competitive equilibrium which is not the maximum competitive equilibrium is not constrained Pareto optimum. However, the maximum competitive equilibrium is constrained Pareto optimum.

THEOREM **2** *In the adverse selection labor market model (where $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and $F(\cdot)$ has an associated density $f(\cdot)$ such that $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$), the highest-wage competitive equilibrium is a constrained Pareto optimum.*

Proof: If all workers are employed in the highest-wage competitive equilibrium, then the outcome is fully Pareto optimal, and hence, constrained Pareto optimal. So, suppose some are not employed. Let w_e and w_u be the wages offered to employed and unemployed workers respectively. Then, the set of workers who accept employment are

$$\{\theta : w_u + r(\theta) \leq w_e\}.$$

Hence, the highest type that accepts employment by this intervention is given by $r^{-1}(w_e - w_u)$ (if some workers accept employment). So, a particular intervention is aimed at setting an acceptable type of workers. So, suppose $\hat{\theta}$ is the highest type of workers accepting employment because of an intervention, where $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$. To do so, we must have

$$w_u + r(\hat{\theta}) = w_e. \tag{1}$$

Further, the authority must balance budget.

$$w_u(1 - F(\hat{\theta})) + w_e F(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta) d\theta. \tag{2}$$

We can restrict attention to the case where the budget is balanced since any Pareto improvement which results in budget surplus can be converted to a Pareto improvement where budget is balanced.

Substituting, and using the fact that w_u and w_e are functions of $\hat{\theta}$, we get that

$$\begin{aligned} w_u(\hat{\theta}) &= \int_{\underline{\theta}}^{\bar{\theta}} \theta f(\theta) d\theta - r(\hat{\theta}) F(\hat{\theta}) \qquad (3) \\ &= F(\hat{\theta})(E[\theta : \theta \leq \hat{\theta}] - r(\hat{\theta})). \qquad (4) \end{aligned}$$

and

$$\begin{aligned} w_e(\hat{\theta}) &= \int_{\underline{\theta}}^{\bar{\theta}} \theta f(\theta) d\theta - r(\hat{\theta})(1 - F(\hat{\theta}) \qquad (5) \\ &= F(\hat{\theta})(E[\theta : \theta \leq \hat{\theta}] - r(\hat{\theta})) + r(\hat{\theta}). \qquad (6) \end{aligned}$$

Now, let θ^* denote the highest type worker who accepts employment in the highest-wage competitive equilibrium. We know that $r(\theta^*) = w^* = E[\theta : r(\theta) \leq w^*] = E[\theta : \theta \leq \theta^*]$. From Equations 3 and 5, we get that $w_u(\theta^*) = 0$ and $w_e(\theta^*) = r(\theta^*)$.

This shows that if an intervention is aimed at setting $\hat{\theta} = \theta^*$, then no Pareto improvement can be achieved over the highest-wage competitive equilibrium. We now examine if a Pareto improvement can be achieved by targeting $\hat{\theta} \neq \theta^*$.

If $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ with $\hat{\theta} \neq \theta^*$, then new payoff of type $\underline{\theta}$ workers is $w_e(\hat{\theta})$ and old wage is $r(\theta^*)$. Hence, if $w_e(\hat{\theta}) < r(\theta^*)$, then workers of type $\underline{\theta}$ will be worse off. Also, type $\bar{\theta}$ workers are unemployed in highest wage competitive equilibrium. So, their payoff is $r(\bar{\theta})$. With intervention, their payoff is $w_u(\hat{\theta}) + r(\bar{\theta})$. Hence, if $w_u(\hat{\theta}) < 0$, then type $\bar{\theta}$ workers are worse off.

This implies that for any Pareto improvement we must have $w_e(\hat{\theta}) \geq r(\theta^*)$ and $w_u(\hat{\theta}) \geq 0$.

Now, consider $\hat{\theta} < \theta^*$. Since $r(\theta^*) > r(\hat{\theta})$, we get by Equation 5

$$w_e(\hat{\theta}) \leq \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta) d\theta + r(\theta^*)(1 - F(\hat{\theta})). \tag{7}$$

Hence, we can write

$$\begin{aligned} w_e(\hat{\theta}) - r(\theta^*) &\leq F(\hat{\theta})(E[\theta : \theta \leq \hat{\theta}] - r(\theta^*)) \\ &= F(\hat{\theta})(E[\theta : \theta \leq \hat{\theta}] - E[\theta : \theta \leq (\theta^*)]) \\ &< 0. \end{aligned}$$

But this leads to Pareto inferior outcomes for workers of type $\underline{\theta}$.

Now, consider $\hat{\theta} > \theta^*$. We know that $E[\theta : r(\theta) \leq w] < w$ for all $w > w^*$. Since $r(\theta^*) = w^*$ and $r(\cdot)$ is strictly increasing, we get that $r(\hat{\theta}) > w^*$. Hence, $E[\theta : r(\theta) \leq r(\hat{\theta})] < r(\hat{\theta})$. But, we know that $E[\theta : r(\theta) \leq r(\hat{\theta})] = E[\theta : \theta \leq \hat{\theta}]$. So, $E[\theta : \theta \leq \hat{\theta}] < r(\hat{\theta})$. Equation 3 implies that $w_u(\hat{\theta}) < 0$. So, type $\bar{\theta}$ workers are worse off. ■

5 Weak Perfect Bayesian Equilibrium

5.1 An Example

Consider the extensive form game in Figure 3. There are two firms Firm I (Incumbent) and Firm E (Entrant). Firm E has three strategies: out, In1, and In2, where In1 and In2 refers to two types of entries and out refers to not entering. Firm I has two strategies if Firm E enters: F (fight) and A (accommodate). The payoffs are as shown in the Figure 3.

One of the Nash equilibria of this game is (out,fight if entry occurs). Since the only subgame of this game is the game itself, this is also a subgame perfect Nash equilibrium (SPNE). The problem with this SPNE is that if Firm I believes that entry has occurred then he should prefer strategy A over F. However, this SPNE recommends F.

To eliminate equilibria of this nature, which cannot be eliminated by SPNE, several refinements have been proposed. We consider one of them.

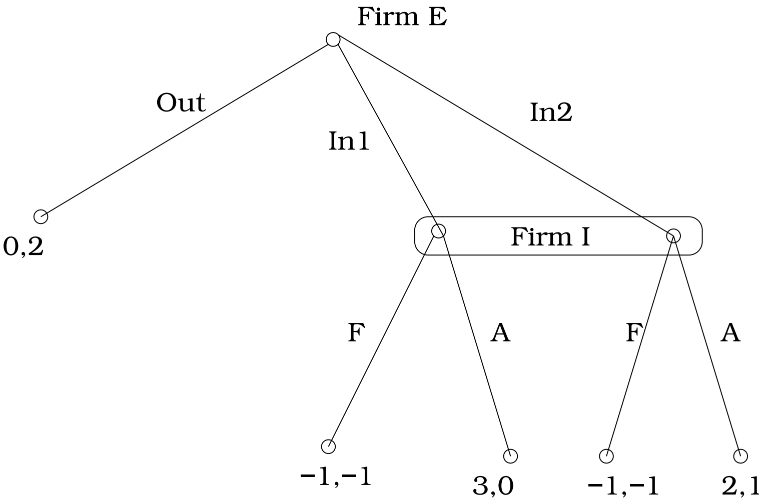

Figure 3: An Example

5.2 System of Beliefs and Sequential Rationality

Definition **3** *A* **system of beliefs** μ *in an extensive game* Γ *is a specification of a probability* $\mu(x) \in [0,1]$ *for each decision node* x *in* Γ *such that*

$$\sum_{x \in H} \mu(x) = 1,$$

for all information sets H.

A system of beliefs can be thought of as specifying, for each information set, a probabilistic assessment by the player who moves at that set of the relative likelihoods of being at each of the information set's various nodes, conditional on play having reached that information set.

In the game in Figure 3, such a specification will involve for example $\mu(E) = 1, \mu(I_1) = 0.3, \mu(I_2) = 0.7$, where E is the node corresponding to Firm E, I_1 is the left node of Firm I and I_2 is the right node of Firm I.

To define sequential rationality, let $E[u_i|H, \mu, \sigma_i, \sigma_{-i}]$ denote player i's expected utility starting at his information set H if his beliefs regarding the conditional probabilities of being at various nodes in H are given by μ, if he follows strategy σ_i, and if his rivals follow σ_{-i}. For example, in Figure 3, if Firm I has belief that with probability λ he will be in the decision node corresponding to In1, then his expected payoff by playing A is $(1 - \lambda)$.

Definition 4 *A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_n)$ in an extensive form game Γ is* **sequentially rational** *at information set H given a system of beliefs μ if, denoting $i(H)$ the player who moves at information set H, we have*

$$E[u_{i(H)}|H, \mu, \sigma_{i(H)}, \sigma_{-i(H)}] \geq E[u_{i(H)}|H, \mu, \sigma'_{i(H)}, \sigma_{-i(H)}]$$

for all $\sigma'_{i(H)}$. If strategy profile σ satisfies this condition at all the information sets H, then we say that σ is **sequentially rational given belief system** *μ.*

In words, a strategy profile σ is sequentially rational if no player finds it worthwhile, once his information sets has been reached, to revise his strategy given his beliefs about what has already occured (via μ) and his rivals' strategies.

As an example, consider the information set of Firm I in Figure 3. Suppose the beliefs of Firm I being at left node and right node are $\frac{1}{3}$ and $\frac{2}{3}$ respectively. Let us verify that strategy profile (entry, accommodate if entry occurs) is sequentially rational at this information set given this belief system. The expected payoff of the Firm I by playing A is $\frac{2}{3}$ and by playing F is -1. Hence, it is sequentially rational. Indeed, for any belief system, a sequential rational strategy profile must have Firm I choosing A. Thus, the earlier SPNE (out, fight if entry occurs) is clearly not sequentially rational for information set of Firm I given any belief system.

5.3 WEAK PERFECT BAYESIAN EQUILIBRIUM

With these definitions, the notion of *perfect Bayesian equilibrium* can be defined. First, strategies must be sequentially rational given beliefs. Second, beliefs must be *consistent* with the strategies.

To explain consistency, consider how one might compute beliefs in the special case when each player plays completely mixed strategies, i.e., every action is played with strictly positive probability at every information set. In this case, every information set is reached with positive probability. The notion of consistency when strategy profile is σ is: For each node x in a given player's information set H, the player should compute the probability of reaching that node given strategies σ, $\text{Prob}(x|\sigma)$, and he should assign conditional probabilities of being at each of these nodes given that play has reached information set H by using Bayes' rule:

$$\text{Prob}(x|H,\sigma) = \frac{\text{Prob}(x|\sigma)}{\sum_{x' \in H} \text{Prob}(x'|\sigma)}.$$

As an example, consider the game in Figure 3. Suppose Firm E uses the completely mixed strategy $\frac{1}{4}$Out, $\frac{1}{2}$In1, and $\frac{1}{4}$In2. Then, probability of reaching Firm I's information set is $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$. Using Bayes' rule, the probability of being at the left node, given that this information set is reached, is $\frac{2}{3}$, and that of right node is $\frac{1}{3}$. For Firm I, his beliefs should be consistent with this strategy, and he should assign exactly these beliefs.

The more difficult issue arises when a particular information set is not reached due to the strategy played. In that case, we cannot use Bayes' rule to compute the node probabilities. For example, if Firm E uses the pure strategy (out), then Firm I can use *any* belief on his information set and still it will be consistent [1]. The weak perfect Bayesian equilibrium takes a *weak* view towards such situations. In particular, it allows us to place any system of beliefs at these information sets. In that sense, the modifier *weak* is attached to this solution concept.

DEFINITION **5** *A profile of strategies and system of beliefs* (σ, μ) *is a* **weak perfect Bayesian Nash equilibrium** *(WPBE) in extensive game* Γ *if it has the following properties:*

1. *The strategy profile* σ *is sequentially rational given belief system* μ*.*
2. *The belief system* μ *should be consistent with strategy profile* σ*, i.e., the system of beliefs* μ *should be derived from strategy profile* σ *through Bayes' rule whenever possible. That is, for any information set* H *such that* $\sum_{x' \in H} Prob(x'|\sigma) > 0$*, we must have for all* $x \in H$*,*

$$\mu(x) = \frac{Prob(x|\sigma)}{\sum_{x' \in H} Prob(x'|\sigma)}.$$

The connection between WPBE and Nash equilibrium is established in the following theorem.

---

[1] This leads to problems in many games - but not in the signaling games we discuss. To rectify this, a refinement of SPNE has been proposed. It is known as the *sequential equilibrium*. We will skip discussions on sequential equilibrium.

THEOREM **3** *A strategy profile σ is a Nash equilibrium of an extensive form game Γ if and only if there exists a belief system μ such that*

- *The strategy profile σ is sequentially rational given belief system μ at* **all information sets H such that** $Prob(H|\sigma) > 0$.
- *The system of beliefs μ is derived from strategy profile σ through Bayes' rule whenever possible.*

We skip the proof. The main difference here is that Nash equilibrium requires sequential rationality at nodes that are **in equilibrium path** but not otherwise. Hence, for the game in Figure 3, the strategy profile (out, fight if entry occurs) is a Nash equilibrium (because the strategy profile puts zero belief on reaching the information set of Firm I, and hence, Firm I need not be sequentially rational at that information set).

Of course the WPBE requires that the strategy profile be sequentially rational at **all** information sets (i.e., information sets reached in equilibrium and off equilibrium path). Hence, a WPBE is also a Nash equilibrium. But as we just saw, not every Nash equilibrium is a WPBE.

5.4 MORE EXAMPLES

In the example in Figure 3, Firm I must play A if In1 or In2 is played, irrespective of his beliefs. Thus, the Nash equilibrium (Out, F) cannot be part of any WPBE.

How about (In1,A)? For this, we have to construct the beliefs given these strategies. Since Firm E plays In1, Firm I should belief that the left node of the information set is reached with probability 1. His best response A is sequentially rational to this belief. Hence this is a WPBE.

Now, consider a modification to this game in Figure 4. Assume $k > -1$.

Let σ_F be the probability with which Firm I fights if entry occurs. Let $\sigma_0, \sigma_1, \sigma_2$ denote the probabilities with which Firm E plays Out, In1, and In2 respectively. Let μ_1 be the belief of Firm I that Firm E played In1 when his information set is reached.

Note that Firm I fights if $-1 \geq -2\mu_1 + (1-\mu_1)(1)$ or $\mu_1 \geq \frac{2}{3}$. Suppose $\mu_1 > \frac{2}{3}$, sequential rationality dictates that Firm I must play F. Given this, Firm E must play In2. But consistency will then require that $\mu_1 = 0$. This is a contradiction.

Now, suppose $\mu_1 < \frac{2}{3}$. Then Firm I plays A. Then Firm E plays In1. Consistency requires then $\mu_1 = 1$. This is a contradiction.

Hence, the only possibility is $\mu_1 = \frac{2}{3}$. In such an equilibrium, Firm E must be randomizing with positive probabilities on In1 and In2 (he cannot be playing Out since In2 dominates Out and he cannot be playing pure In1 or In2 since that will imply $\mu_1 \neq \frac{2}{3}$). In particular, he should put twice as much probability on In1 than on In2. This means, Firm E must be

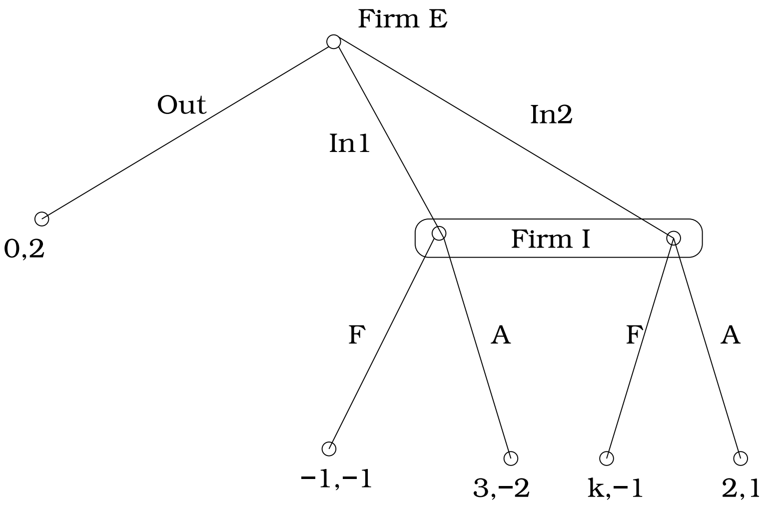

Figure 4: An Example - $k > 0$

indifferent between In1 and In2, i.e., $-1\sigma_F + 3(1-\sigma_F) = k\sigma_F + 2(1-\sigma_F)$. This gives us $\sigma_F = \frac{1}{k+2}$. Hence, Firm E's payoff in playing In1 or In2 is then $\frac{3k+2}{k+2} > 0$, and hence he must play Out with zero probability. Therefore, the unique WPBE in this game ($k > 0$) has $(\sigma_0, \sigma_1, \sigma_2) = (0, \frac{2}{3}, \frac{1}{3})$, $\sigma_F = \frac{1}{k+2}$, and $\mu_1 = \frac{2}{3}$.

6 SEQUENTIAL EQUILIBRIUM

6.1 AN EXAMPLE

The requirement that a weak perfect Bayesian Nash equilibrium (WPBE) puts on beliefs is mild: beliefs should be probability distributions over nodes in an information set, and they should be computed using Bayes' rule in information sets that are reached with positive probability in equilibrium. No restriction is imposed on beliefs *off equilibrium path* (except the restriction imposed by sequential rationality). This has led to several refinements of WPBE in the literature.

To understand the problem of not specifying beliefs off equilibrium path, consider the extensive form game in Figure 5. The WPBE beliefs are given in square brackets and the WPBE strategies are given with dotted lines. Because of Bayes' rule, the beliefs in the information set of Player 1 has to be (0.5,0.5). Suppose Player 2 has a belief of μ on his left node. Then, his expected payoff by playing l is 5 and expected payoff by playing r is $2\mu + 10(1-\mu) = 10 - 8\mu$. So, Player 2 plays l if $\mu > \frac{5}{8}$, plays r if $\mu < \frac{5}{8}$, and is indifferent at $\mu = \frac{5}{8}$. At $\mu = 0.9$, Player 2 plays l. Player 1, being sequentially rational, plays x. Now, the information set of Player 2 is not reached in equilibrium. Hence, the beliefs specified in the WPBE are still fine. But, they are certainly not reasonable. Because, if Player 1 ever deviates, and plays y with positive probability, then both the nodes of Player 2's information set will be reached with equal probability. Hence, the only reasonable belief for Player 2 is to place equal probability on both the nodes.

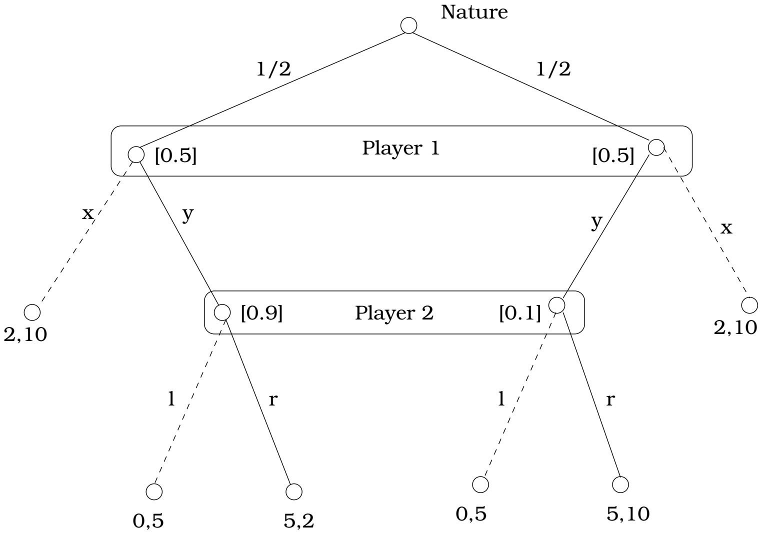

Figure 5: An Example

Consider another example in Figure 6. Clearly, the specified strategy and beliefs constitute a WPBE. Again, the beliefs are clearly fine according to WPBE since the information set is not reached in equilibrium. But this WPBE is not a subgame perfect Nash equilibrium. The subgame after Firm E's entry does not have a Nash equilibrium in which Firm E plays A and Firm I plays f. This illustrates that the WPBE is not a good generalization of subgame perfect Nash equilibrium.

6.2 Sequential Equilibrium

The refinement *sequential equilibrium* introduces consistency indirectly through limiting sequence of strategies and beliefs.

Definition **6** *A strategy profile and system of beliefs* (σ, μ) *is a* **sequential equilibrium** *of extensive form game* Γ *if it has the following properties:*

1. *Strategy profile* σ *is sequentially rational given belief system* μ*.*
2. *There exists a sequence of completely mixed strategies* $\{\sigma^k\}_{k=1}^{\infty}$*, with* $\lim_{k\to\infty} \sigma^k = \sigma$*, such that* $\mu = \lim_{k\to\infty} \mu^k$*, where* μ^k *denotes the beliefs derived from strategy profile* σ^k *using Bayes' rule. Beliefs derived in this manner are called* consistent *beliefs.*

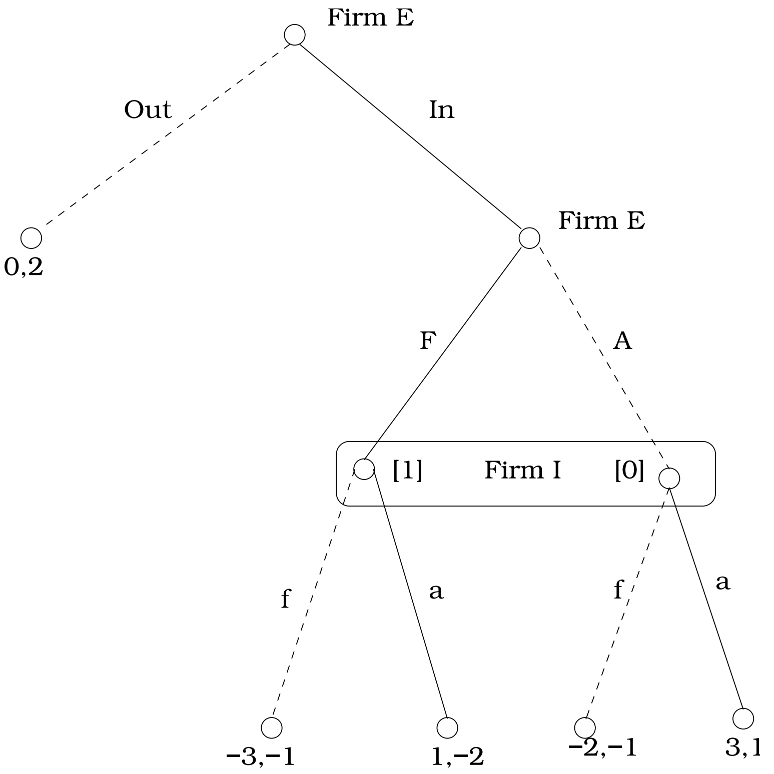

Figure 6: An Example

By requiring completely mixed strategies, which are arbitrarily close to the prescribed strategies, we can use Bayes' rule to derive beliefs in information sets not reached in equilibrium also. In essence, the sequential equilibrium requires that beliefs be justifiable as coming from some set of totally mixed strategies that close to the equilibrium strategies, i.e., small perturbations of the equilibrium strategies. This can be believed as if players making small mistakes in playing equilibrium strategies. Every sequential equilibrium is a WPBE, but in general the reverse is not true.

Theorem 4 *In every sequential equilibrium* (σ, μ) *of an extensive form game* Γ*, the equilibrium strategy profile* σ *constitutes a subgame perfect Nash equilibrium of* Γ*.*

In Figure 5, if Player 1 plays y with some probability, then Player 2 should put 0.5 belief on both nodes of his information set. By sequential rationality, he should then play r. Player 1 will then choose y.

7 More Examples

Consider the example in Figure 7. Let μ be the belief of Player 2 on node $2x$. Then, he plays L if $3\mu > 1 - \mu$ or $\mu > \frac{1}{4}$. He plays R if $\mu < \frac{1}{4}$, and mixes at $\mu = \frac{1}{4}$. If Player 2 plays

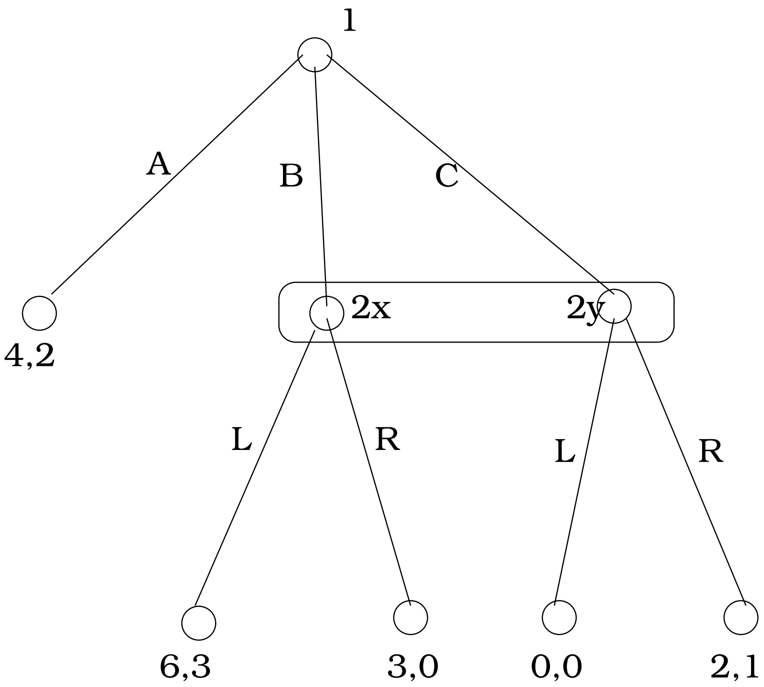

Figure 7: An Example

L, then Player 1 will play B, which is consistent with beliefs. So, one sequential equilibrium is (B, L) with $\mu = 1$.

If Player 2 plays R, then Player 1 plays A. Then, any belief with $\mu \leq \frac{1}{4}$ is consistent by WPBE. One can choose limiting strategies that are consistent with these beliefs. So, these are also sequential equilibrium.

At $\mu = \frac{1}{4}$, Player 2 mixes, say $\alpha L + (1-\alpha)R$. Player 1 cannot play C in equilibrium (dominated by B). It cannot play B also since that will imply $\mu = 1$. For it to play A, $4 \geq 6\alpha + 3(1-\alpha)$ or $\alpha \leq \frac{1}{3}$.

Consider another example in Figure 8. Consider an equilibrium strategy in which Player 1 plays A, Player 2 plays E. We want to find out what beliefs are consistent for Player 3. Let us suppose $\epsilon_B, \epsilon_C, \epsilon_D$ are the weights on strategies B, C, D. So, $\mu(x) = \frac{\epsilon_B}{\epsilon_B + \epsilon_C \epsilon_D}$. If $\epsilon_B = \epsilon_C = \epsilon_D = \epsilon$, then $\mu(x) = \frac{1}{1+\epsilon} \rightarrow 1$. So, $\mu(x) = 1$ is consistent. If $\alpha \in (0, 1)$, then $\epsilon_C = \epsilon_D = \epsilon$ and $\epsilon_B = \frac{\epsilon^2 \alpha}{1-\alpha}$ gives $\mu(x) = \alpha$. So, $\mu(x) = \alpha$ is consistent. Finally, $\epsilon_B = \epsilon^3$ gives $\mu(x) = \frac{\epsilon}{1+\epsilon} \rightarrow 0$. So, $\mu(x) = 0$ is also consistent.

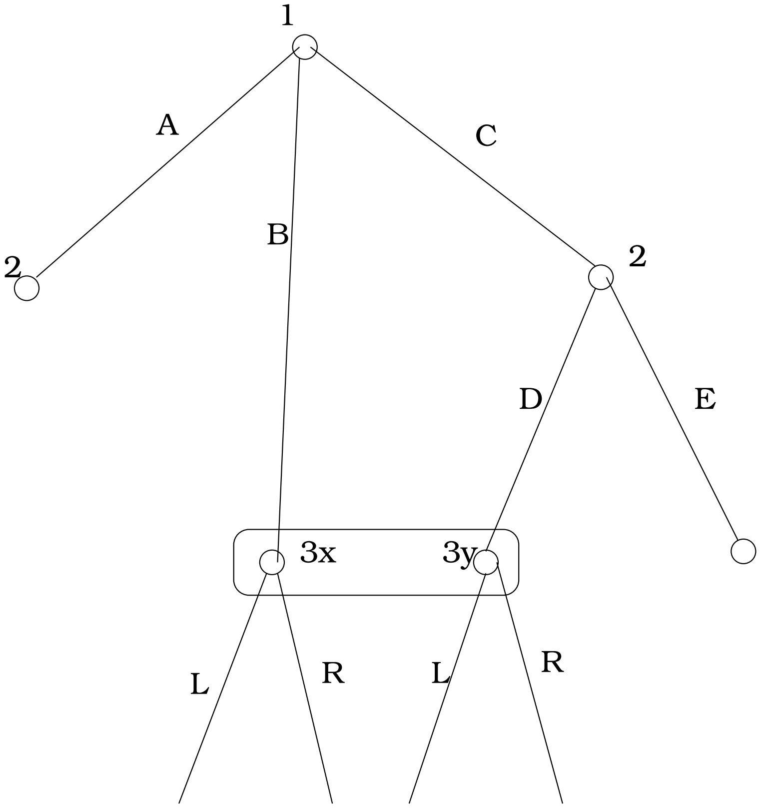

Figure 8: An Example

8 SIGNALING - A BASIC INTRODUCTION

8.1 PERFECT BAYESIAN EQUILIBRIUM

We introduce a basic model of signaling game with two players. In the next section, we give a full scale example of signaling in job market setting.

A signaling game is a dynamic game of incomplete information involving two players: a sender (S) and a receiver (R). The timing of the game is as follows:

1. Nature draws a type t_i for the sender from the set of possible types $T = \{t_1, \ldots, t_n\}$ according to a probability distribution $p(\cdot)$ with $p(t_i) > 0$ for all $t_i \in T$ and $\sum_{t_i \in T} p(t_i) = 1$.

2. The sender S observes his type t_i and chooses a **message** m_j from a set of possible messages $M = \{m_1, \ldots, m_l\}$.

3. The receiver R observes message m_j (but not the type of sender) and chooses an **action** a_k from a set of actions $A = \{a_1, \ldots, a_q\}$.

4. Payoffs are given by $u_S(t_i, m_j, a_k)$ and $u_R(t_i, m_j, a_k)$.

In many cases, the sets T, M, and A are intervals on real lines. Also, in many cases, the set of messages is a function of type and the set of actions is also a function of message.

There are many applications of signaling games. One of the applications, which we will study later, is the job market game. In the job market signaling game, the sender is a worker and the receiver is a firm. The type of a worker is his productivity and the message of a worker is his education level. The action of a firm is the wage offered to the worker.

Figure 9 gives an extensive form representation of an abstract signaling game with $T = \{t_1, t_2\}$, $M = \{m_1, m_2\}$, $A = \{a_1, a_2\}$, and probability of realizing type t_1 being p.

Strategy of sender: In a signaling game, the (pure) strategy of a sender is a function of his type. It is denoted as $m(t_i)$, where $m : T \to M$. In the example in Figure 9, we have four possible strategies of the sender:

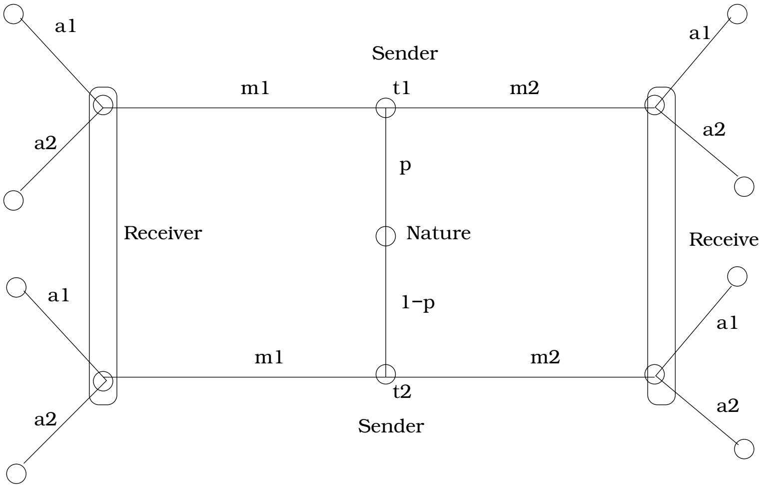

Figure 9: An Example of a Signaling Game

P1 Play m_1 if Nature draws t_1 and play m_1 if Nature draws t_2.

P2 Play m_2 if Nature draws t_1 and play m_2 if Nature draws t_2.

S1 Play m_1 if Nature draws t_1 and play m_2 if Nature draws t_2.

S2 Play m_2 if Nature draws t_1 and play m_1 if Nature draws t_2.

Strategies P_1 and P_2 are called **pooling strategies**, i.e., strategies where both types play the same message. Strategies S_1 and S_2 are called **separating strategies**, i.e., strategies where both types play different messages. In a model with more than two types, one can think of *partial pooling* strategies, where types in a given set play the same strategy. Similarly, one can think of *hybrid* strategies where some types randomize.

Strategy of receiver: The (pure) strategy of a receiver is a function of the message he receives. It is denoted as $a(m_j)$, where $a : M \to A$. In the example in Figure 9, we have four possible strategies of the receiver:

A1 Play a_1 if message is m_1 and play a_1 if message is m_2.

A2 Play a_1 if message is m_1 and play a_2 if message is m_2.

A3 Play a_2 if message is m_1 and play a_2 if message is m_2.

A4 Play a_2 if message is m_1 and play a_1 if message is m_2.

Now, we put a series of requirements for any equilibrium of the signaling game. Denote by $\mu(t_i|m_j)$ the belief of the receiver about the type of the sender when he receives a message m_j. The first requirement is that $\mu(\cdot)$ should be a belief system.

Requirement 1: After observing any message m_j, the Receiver should have beliefs about which types could have sent this message. Denote this belief by $\mu(t_i|m_j)$ where $\mu(t_i|m_j) \geq 0$ for each $t_i \in T$ and

$$\sum_{t_i \in T} \mu(t_i|m_j) = 1 \qquad \forall\ m_j \in M.$$

Given the Receiver's beliefs and Sender's message, the Receiver chooses an optimal action by maximizing his expected payoff.

Requirement 2R: For each message $m_j \in M$, the Receiver's optimal action $a^*(m_j)$ solves

$$\max_{a_k \in A} \sum_{t_i \in T} u_R(t_i, m_j, a_k)\mu(t_i|m_j)$$

The Sender has full information. Given the optimal strategies of the Receiver, he chooses an optimal message.

Requirement 2S: Given his type t_i, the Sender's optimal message $m^*(t_i)$ solves

$$\max_{m_j \in M} u_S(t_i, m_j, a^*(m_j))$$

Finally, given Sender's strategy $m^*(t_i)$, let T_j be the set of types that could send message m_j, i.e., $m^*(t_i) = m_j$ for all $t_i \in T_j$. If T_j is non-empty, then the information set corresponding to the message m_j is on the equilibrium path. If T_j is empty, then the information set corresponding to the message m_j is not in the equilibrium path. The final requirement says that the beliefs on information sets in the equilibrium path must be computed using Bayes' rule.

Requirement 3: For each $m_j \in M$, if there exists $t_i \in T$ with $m^*(t_i) = m_j$, then

$$\mu(t_i|m_j) = \frac{p(t_i)}{\sum_{t'_i \in T_j} p(t'_i)},$$

where $p(t_i)$ is the probability of realizing type t_i.

DEFINITION **7** *A pure strategy* **perfect Bayesian equilibrium** *in a signaling game is a pair of strategies $m^*(\cdot)$ and $a^*(\cdot)$ and a belief system $\mu(\cdot)$ satisfying Requirements 1, 2R, 2S, and 3.*

If the Sender's strategy is a pooling or a separating strategy in equilibrium, then we call such an equilibrium a pooling equilibrium or a separating equilibrium respectively.

9 AN EXAMPLE

In the rest of the discussion, we will analyze the equilibria in the signaling game of Figure 10. We denote the belief on the left information set as $(p, 1-p)$ and the belief on the right information set as $(q, 1-q)$.

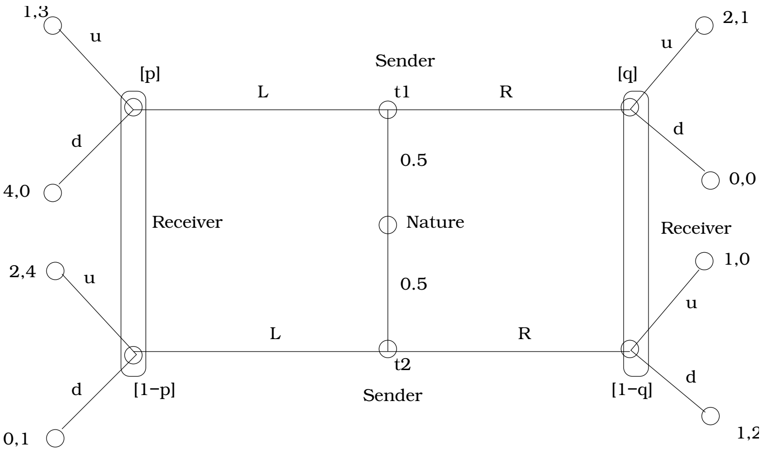

Figure 10: An Example of a Signaling Game

There are four types of equilibria that are possible.

1. *Pooling on L:* Suppose there is an equilibrium in which both types send message L. Then, by Bayes' rule, $p = 0.5$. Irrespective of the value of p, if the Receiver sees a message L, then it is optimal for him to play u. What does the Receiver do if he sees

R? He plays u if and only if $q \geq 2(1-q)$ or $q \geq \frac{2}{3}$. Hence, by playing L, the Sender types receive payoffs of 1 and 2 respectively. But if $q > \frac{2}{3}$, then type t_1 gets a payoff of 2 by playing R. So, in equilibrium $q \leq \frac{2}{3}$ and Receiver must play d when he receives message R. In that case, the Sender gets payoffs 0 and 1 respectively, which is less than the respective payoffs by playing L. Hence, an equilibrium is $(L, L), (u, d), p = 0.5$, and $q \leq \frac{2}{3}$.

2. *Pooling on R:* Clearly, $q = 0.5$. Hence, the best response of the Receiver is d. This gives a payoff of $(0, 1)$ for the Sender. For any value of p, the Receiver plays u for message L. Hence, the payoff of the Sender is $(1, 2)$, which is greater than $(0, 1)$. Hence, no equilibrium is possible with pooling on R.

3. *Separation, with t_1 playing L:* Now, both the information sets are reached in equilibrium. So, $p = 1$ and $q = 0$. Receiver's best response is u for L and d for R. This gives payoff 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, this is not an equilibrium.

4. *Separation, with t_1 playing R:* Now, $p = 0$ and $q = 1$. So, the best response of the Receiver is (u, u), and both Sender types earn payoff of 2. If t_1 deviates and plays L, then he gets a payoff 1. So, it is not optimal to deviate for t_1. If t_2 deviates and plays R, then t_2 gets a payoff of 1 (always). Thus, $(R, L), (u, u), p = 0, q = 1)$ is a separating perfect Bayesian equilibrium.

10 Refinements for Signaling Games

In this section, we look at some refinements for signaling games. The idea is to form beliefs off equilibrium path. One may think that sequential equilibrium may help, but unfortunately anything that can be sustained in perfect Bayesian equilibrium can also be sustained in sequential equilibrium in signaling games (verify this). Hence, there is a need for further refinements.

Consider the example in Figure 11. Consider the pooling equilibrium when both types play L. Then $p = 0.5$, and the best response of Receiver to L is u. The best response of Receiver to R is d for $q \geq \frac{1}{2}$ and u for $q \leq \frac{1}{2}$. Suppose Receiver sets $q \geq \frac{1}{2}$ and plays d in response to R. Then, both t_1 and t_2 get zero by deviating. Hence, $[(L, L), (u, d), p = 0.5, q \geq \frac{1}{2}]$ is a pooling equilibrium.

But does this equilibrium make sense? The key feature here is that it makes no sense for t_1 to play R. So, if the Receiver observes R, he should infer the type to be t_2, and thus have $q = 0$.

Definition 8 *In a signaling game, the message $m_j \in M$ is* **dominated for type t_i** *from T if there exists another message $m'_j \in M$ such that the lowest payoff of t_i from m'_j is greater*

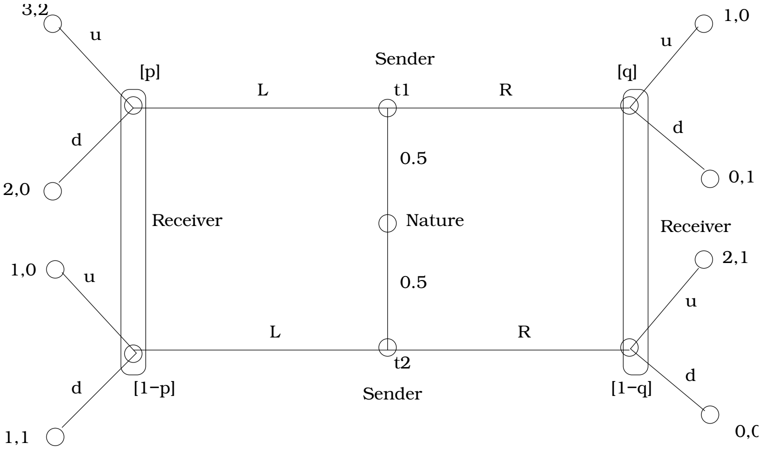

Figure 11: Domination in a Signaling Game

than the highest payoff of t_i *from* m_j*, i.e.,*

$$\min_{a_k \in A} u_S(t_i, m'_j, a_k) > \max_{a_k \in A} u_S(t_i, m_j, a_k).$$

Requirement 4: If the information set following m_j is off the equilibrium path and m_j is dominated for type t_i, then (if possible) the Receiver's belief $\mu(t_i|m_j)$ should equal zero. (This is possible provided m_j is not dominated for all types in T.)

In the game in Figure 11, the separating perfect Bayesian equilibrium $[(L, R), (u, u), p = 1, q = 0]$ trivially satisfies Requirement 4 since all information sets are on equilibrium path. To get a non-trivial example, reverse the payoff of Receiver from type t_2 when he plays R: 1 from playing d and 0 from playing u in Figure 11. Now, $[(L, L), (u, d), p = 0.5, q]$ is a pooling perfect Bayesian equilibrium for any value of q. But it satisfies Requirement 4 when $q = 0$.

In many games, there are perfect Bayesian equilibria that seem unreasonable but still satisfy Requirement 4. To get rid of some unreasonable equilibria, we need to impose additional reasonable requirements.

Consider the famous "Beer and Quiche" game in Figure 12. Notice that the Wimpy type prefers Quiche and the Surly type prefers Beer. Also, both types would prefer not to duel with the Receiver. Now, consider the pooling equilibrium when both types choose Quiche. Then, $p = 0.1$. Then, the Receiver's response to Quiche is not duel. Also, for $q \geq \frac{1}{2}$, the Receiver must duel in response to Beer. Since both types prefer not dueling to dueling, $[(Quiche, Quiche), (not, duel), p = 0.1, q \geq \frac{1}{2}]$ is a perfect Bayesian equilibrium. Since message Beer is not dominated for any type, Requirement 4 is trivially satisfied here. In particular, the Wimpy type is not guaranteed to do better by choosing Quiche over Beer (minimum payoff from Quiche is 1 whereas maximum payoff from Beer is 2).

But still this equilibrium is not reasonable. The main question is if the Receiver sees Beer, what should he conclude the type be? Consider the Wimpy type. He gets 3 in equilibrium.

If he deviates to Beer, then the maximum he can get is 2. On the other hand, the Surly type can improve his equilibrium payoff by deviating to Beer (if the Receiver plays "not"). Given this, one would expect the Surly type to deviate, and not the wimpy type. Hence, if the Receiver see Beer, he should believe that it is the Surly type who has deviated. In that case, $q = 0$, and the Receiver chooses not to duel. This means, the Surly type will deviate to choose Beer. This leads to the following refinement.

DEFINITION **9** *Given a perfect Bayesian equilibrium in a signaling game, the message $m_j \in M$ is* **equilibrium dominated for type** $t_i \in T$ *if t_i's equilibrium payoff, denoted by $u^*(t_i)$, is greater than t_i's highest possible payoff from m_j, i.e.,*

$$u^*(t_i) > \max_{a_k \in A} u_S(t_i, m_j, a_k).$$

Requirement 5 - The "Intuitive Criterion": If the information set following m_j is off equilibrium path and m_j is equilibrium dominated for type t_i, then the Receiver's belief $\mu(t_i|m_j)$ should equal zero (if possible). (This is possible if m_j is not equilibrium dominated for all types.)

Notice that any equilibrium satisfying Requirement 5 must also satisfy Requirement 4. So, Requirement 4 is redundant in the presence of Requirement 5. It can be shown (Cho and Kreps, 1987) that perfect Bayesian equilibrium satisfying Requirement 5 always exists in signaling games.

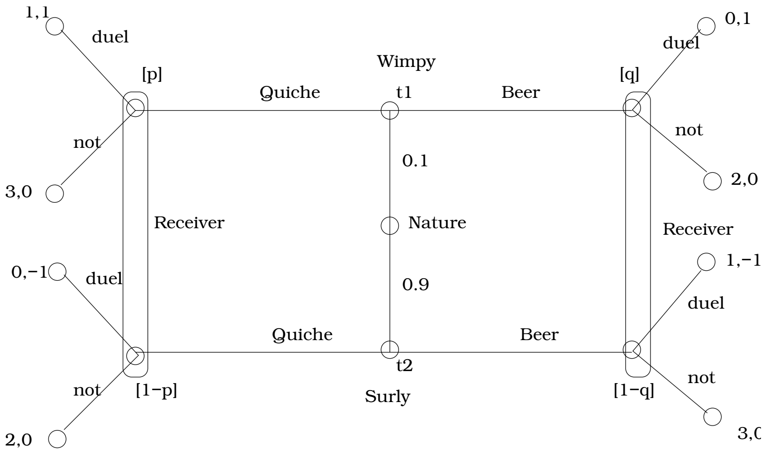

Figure 12: Equilibrium Domination in Beer and Quiche Signaling Game

This kind of argument is often referred to as *forward induction* in game theory. The basic idea is 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.

11 JOB MARKET SIGNALING

We now investigate our job market adverse selection model in the presence of signaling. For simplicity, we focus on the case where there are two firms and one worker who can be of two types $\{\theta_L, \theta_H\}$ where $\theta_H > \theta_L > 0$. The probability that a worker is of type θ_H is $\lambda \in (0, 1)$. The important extension from the previous model is that before entering the market, a worker may get education which is visible to firms. This education serves as a signal. We analyze an extreme case where we assume that the education does *nothing* to the productivity of the worker (you can think of the education as taking an aptitude test, which signals the type of the worker). The education is however costly. Education level $e \in \mathbb{R}_+$. The cost of education level e for a worker of type θ is given by $c(e, \theta)$ (we can assume this cost to be financial cost or mental cost of appearing in a test). We assume the following about the cost function (subscripts denote partial derivatives):

- $c(\cdot, \cdot)$ is a twice continuously differentiable function.
- $c(0, \theta) = 0$ for all θ - cost of no education is zero.
- $c_e(e, \theta) > 0$ and $c_{ee}(e, \theta) > 0$ for all θ and for all e.
- $c_\theta(e, \theta) < 0$ for all $e > 0$ and $c_{e,\theta}(e, \theta) < 0$ for all e and for all θ - both cost and marginal cost of education is lower for high types.

An implication of these assumptions is that **indifference curves** of two types of workers cross **exactly once** and when they do the **slope of the indifference curve of the high type is smaller than that of the low type**. This property is called the **single crossing** property. This happens because worker's marginal rate of substitution between wages and education at any given (w, e) pair is $c_e(e, \theta)$, which is decreasing in θ because $c_{e\theta}(e, \theta) < 0$.

If w is the wage offered to a worker of type θ who has an education level e, then his utility is given by

$$u(w, e|\theta) = w - c(e, \theta).$$

A worker can earn $r(\theta)$ by not accepting employment if his type is θ. We will show that equilibria of this game may involve high type workers getting higher level of education and low type workers getting no education to distinguish themselves in the market. As a result, the firms correctly recognize the workers. Though this may lead to correct allocation of labor, it need not result in Pareto improvements since education is costly.

First, we focus on the case where $r(\theta_L) = r(\theta_H) = 0$. Note that under this assumption the wage offered in the no signaling model is $E[\theta] = \lambda\theta_H + (1 - \lambda)\theta_L$ and is Pareto efficient. Thus, we can see now the effect of signaling clearly.

The extensive form game goes as follows. The worker realizes its type. Then, he choses to get an education level, which is observed by the firms (but not the type of the worker). Then, firms simultaneously offer a wage to the worker. Finally, worker has three strategies, to accept firm 1, to accept firm 2, and to not accept anything.

First, we focus on perfect Bayesian equilibrium (PBE) - note that it is equivalent to sequential equilibrium in this model. We start from the end of the game tree. Clearly, the worker must accept the firm's offer which is higher (with equal probability of joining either in case of ties). First, both firms must assign the same belief since they get the same information. Suppose they observe an education level e. Then, let $\mu(e)$ be the belief that a worker of type θ_H has this education level e. If so, then the expected productivity of the worker is $\mu(e)\theta_H + (1-\mu(e))\theta_L$. Setting wage equal to this expected productivity is a Nash equilibrium of this simultaneous move game. To see this, note that any wage higher than this will give a firm negative expected payoff and any wage lower than his will not attract the worker (zero payoff). So, in any pure strategy PBE, both firms must be offering a wage equal to his expected productivity $\mu(e)\theta_H + (1-\mu(e))\theta_L$. This further implies that equilibrium wage lies in $[\theta_L, \theta_H]$.

11.1 SEPARATING EQUILIBRIA

Now, we need to determine the equilibrium education levels of worker. We first analyze separating equilibria. Let $e^*(\theta)$ be the worker's equilibrium education level when his type is θ and $w^*(e)$ be the firms' wage offer in equilibrium after seeing education level e.

LEMMA **1** *In any separating PBE,* $w^*(e^*(\theta_H)) = \theta_H$ *and* $w^*(e^*(\theta_L)) = \theta_L$ *(so, each worker type receives a wage equal to his type).*

Proof: In the separating PBE, the information set corresponding to $e^*(\theta_H)$ and $e^*(\theta_L)$ is reached. So, firms must place correct beliefs according to Bayes' rule. According to the strategies prescribed, θ_H type worker plays $e^*(\theta_H)$ and θ_L type worker plays $e^*(\theta_L)$. Hence, $\mu(e^*(\theta_H)) = 1$ and $\mu(e^*(\theta_L)) = 0$. The resulting wage is then θ_H for $e^*(\theta_H)$ and θ_L for $e^*(\theta_L)$. ■

LEMMA **2** *In any separating PBE,* $e^*(\theta_L) = 0$ *(low type chooses not to get education).*

Proof: Assume for contradiction that in some separating PBE $e^*(\theta_L) = e > 0$. By Lemma 1, $w^*(e) = \theta_L$. Now, by deviating to 0 education level, the worker will at least get a wage of θ_L (remember wage lies in $[\theta_L, \theta_H]$). But he will save on education cost, and thus get a higher payoff. This is a contradiction to the fact that this is a PBE. ■

Lemma 2 has some implications. First, it says that in any separating PBE, low type workers get no education and high type workers get some positive level of education. Second, equilibrium utility of low type worker is θ_L. Let $\bar{e}$ be the education level at which if offered a wage θ_H, the low type worker gets utility θ_L, i.e., $\theta_L = \theta_H - c(\bar{e}, \theta_L)$ or $c(\bar{e}, \theta_L) = (\theta_H - \theta_L)$.

Now, we describe one separating PBE: $e^*(\theta_H) = \bar{e}$, $e^*(\theta_L) = 0$ and the wage schedule $w^*(e)$ is as shown in Figure 13. The firms' belief following education level e is

$$\mu(e) = \frac{w^*(e) - \theta_L}{\theta_H - \theta_L}.$$

Note that $\mu(e) \in [0, 1]$ for all e.

To verify that this is indeed a PBE, we note that $\mu(0) = 0$ and $\mu(\bar{e}) = 1$. For other e, we are free to choose μ according to PBE. The wage offers are exactly expected type: $\mu(e)\theta_H + (1 - \mu(e))\theta_L = w^*(e)$.

We now verify if worker's strategy is optimal. If a worker is of low type, by sending any other signal, he gets a higher wage but not high enough to offset his cost (as shown in Figure 13). Similarly, the high type worker maximizes his utility by choosing the specified education level (as shown in Figure 13).

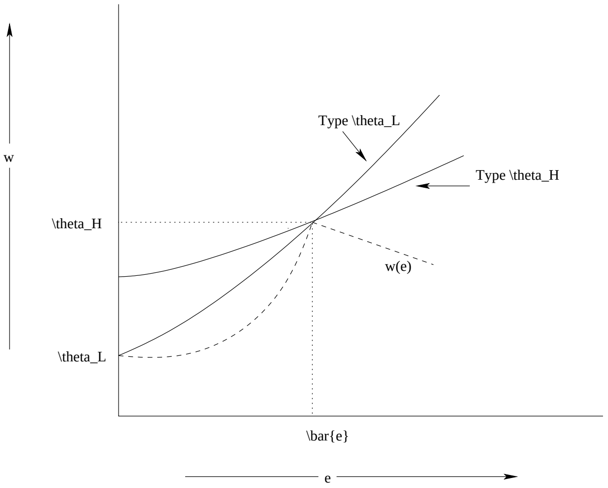

Figure 13: Wage schedule in a separating PBE

The separating equilibria in Figure 13 is not the only separating equilibrium we have. For the same education level chosen by the high type, the firms can have different beliefs off equilibrium path (and hence wage levels). Figure 14 shows another separating equilibrium.

Further, there are separating equilibrium where the high type may choose education level higher than $\bar{e}$. Figure 15 shows another education level e_1 for high type in a separating equilibrium and a corresponding wage schedule. We argue that in any separating equilibrium the education level chosen by the high type lies between $\bar{e}$ and e_1. Note that if $e < \bar{e}$ is an

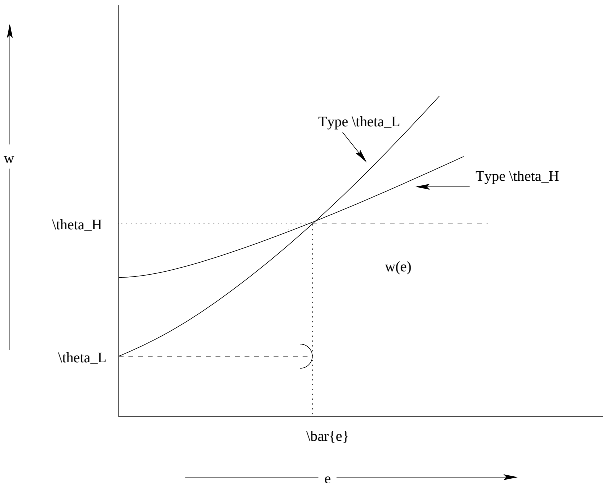

Figure 14: Wage schedule in a separating PBE

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. This is because the indifference curve of low type will be shifted "up" when he deviates to education level e. Similarly, any education level $e > e_1$ for high type cannot be sustained in equilibrium. This is because at any education level greater than e_1, 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. Finally, any education level in $[\bar{e}, e_1]$ can be sustained in a separating equilibrium of high type by appropriate wage schedule as we have illustrated.

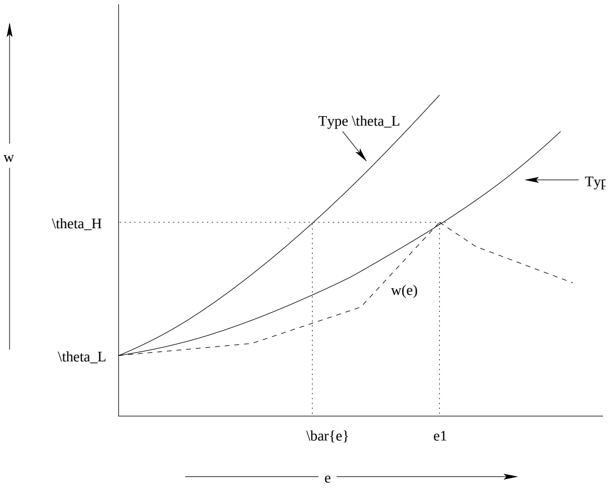

Figure 15: Wage schedule in a separating PBE

A striking feature of these separating equilibria is that they can all be Pareto ranked. First, the firms earn zero expected profit and the low type earns a payoff equal to θ_L. The high type worker gets maximum payoff at $\bar{e}$ (since at higher education levels, his wage remains same but costs go up). Hence, the separating equilibria (equilibria and not equilibrium because at any given education level, there are many possible wage schedules) where the high type chooses education level $\bar{e}$ Pareto dominates all other separating equilibria.

We now compare the welfare of these equilibria to the case where signaling is not possible. Note that in the absence of signaling, the wage offered is $E[\theta]$. This is larger than θ_L. Hence, the low type workers are worse off in the presence of signaling. The payoffs of high type workers may go either way (this is surprising since we expected that signaling will improve the payoff of high type workers). Consider two scenarios. First, suppose that $E[\theta]$ is low, then high type workers will get a high wage θ_H in equilibrium and this will offset their cost of education by sufficient amounts to increase payoff from no signaling. On the other hand, consider the case when $E[\theta]$ 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 workers. The problem here is that when signaling is present, if the high type chooses zero education level, he will be treated as low type and given θ_L wage. So, he is forced to get this costly education even though it reduces his welfare from no signaling.

Note that high value of $E[\theta]$ means high value of λ (greater number of high type workers). So, in the presence of higher fraction of high type workers, high type workers are made worse off in signaling than in no signaling. So, everyone gets costly education to avoid being mislabeled as low type, who are very few in number.

11.2 APPLYING INTUITIVE CRITERION

Consider any education level $e > \bar{e}$. At these education levels, the type θ_L worker earns a payoff less than θ_L, his equilibrium payoff in any separating equilibrium. Hence, if we impose intuitive criterion, the firm must place a belief of $\mu(e) = 1$ for any education level greater than $\bar{e}$. So, $w(e) = \theta_H$ for all $e > \bar{e}$. Consider a separating equilibrium where high type chooses an education level $e > \bar{e}$. Then, the high type can deviate to $\bar{e}$, and get higher payoff. So, no such equilibria can exist. Thus, the only equilibria surviving intuitive criterion is when the high type worker chooses education level $\bar{e}$.

11.3 POOLING EQUILIBRIA

We now examine the pooling equilibria where both type workers choose an education level $e^*(\theta_L) = e^*(\theta_H) = e^*$. Since the firms' beliefs are computed using Bayes' rule, they must assign probability λ for the high type worker when they see signal e^*. This means that the equilibrium wage offered must be $\lambda\theta_H + (1 - \lambda)\theta_L = E[\theta]$.

In general the pooling equilibrium may be complicated (in terms of wage schedule and the type of beliefs off equilibrium path). Figure 16 gives one pooling equilibrium. However,

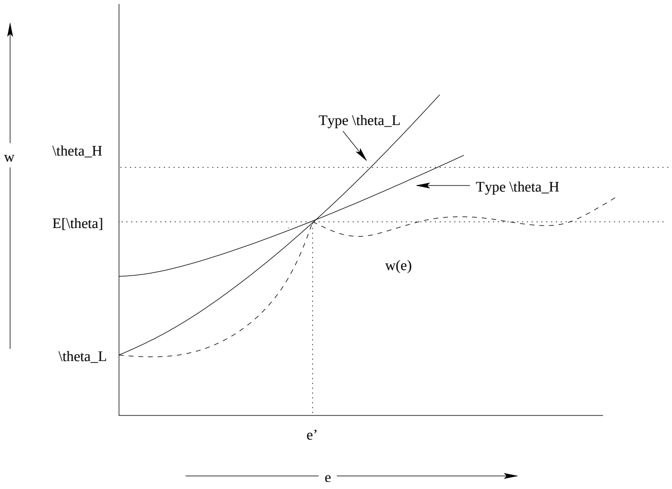

Figure 16: Wage schedule in a pooling PBE

using intuitive criterion eliminates all the pooling equilibria.