Professor Lars Stole
University of Chicago
April 15, 2019, (v 1.14)

Lecture notes on Adverse selection, Signaling and Competitive Screening

Please do not distribute to anyone outside of class.

We begin our study of the economics of incomplete information by considering *competitive* markets with adverse selection. We will consider two environments throughout our analysis – the labor market setting and the insurance setting. The former is explored in MWG (ch13), while the latter is the focus of JR (ch8).

- two sides to the market - informed and uninformed;
- agents on the informed side (e.g., workers who know their productivity or consumers who know their probability of accident) have information that determines their own returns from trade, and also the returns from trade to the uninformed side of the market (e.g., firms hiring workers or insurance companies who sell insurance);
- in equilibrium, the uninformed side of the market competes away any profits from trade;
- **adverse selection** arises because the informed agents who gets the most from trade (e.g., low-productivity workers and high-risk consumers) are the types of agents who impose the greatest losses on the uninformed side of the market (e.g., hiring firms and insurance companies).

1 Adverse selection in competitive markets

In our first set of models, we assume that the object being traded is exogenously fixed. In the labor market example, there is a single type of job and the worker is either employed or not; in the insurance example, there is only a full-insurance contract available and it is either purchased or not.

1.1 Labor-market model (MWG)

- Firms and workers are risk neutral.
- Each worker has a productivity type, θ, drawn from $\Theta \equiv [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}_+$ according to the distribution function $F(\theta)$. If F is a continuous distribution, we will denote the corresponding density as f. More generally, most integration will be with respect to the measure $dF(\theta)$ (rather than $f(\theta)d\theta$ which may not be well defined).

- The firm earns gross profits θ from a worker of type θ.

- The worker's payoff from working at a firm for wage w is simply w; the worker's payoff from home production is $r(\theta)$. Thus, a worker will not choose to work for at wage w if $w < r(\theta)$. We will assume throughout that a worker who is indifferent between employment and home production will choose employment. Thus a worker will choose employment if she is offered a wage $w \geq r(\theta)$.

As a benchmark, suppose that firms can observe θ and condition their wage offers accordingly. The **full-information equilibrium benchmark** will then be a function of type, $w^{fb}(\theta)$, such that each firm earns zero profit from any hired worker (which for the employed workers requires):

$$w^{fb}(\theta) = \theta.$$

The corresponding set of workers who accept employment is

$$\Theta^{fb} = \{\theta \mid r(\theta) \leq \theta\}.$$

1.1.1 Competitive equilibrium with incomplete information

- Now assume that the worker is informed about θ, but the firm is not. The firm only knows the underlying information structure. Thus, each firm offers a single wage, w.

- By "Competitive equilibrium," we mean the participants take prices as given and therefore in equilibrium all workers will be employed at the same wage, w. By assuming more than one firm, each capable of unlimited demand, we have effectively imposed that and the employing firms will earn zero expected profit.

- At this wage, the set of workers who accept employment will be

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

- If a firm believes the expected productivity of a worker is μ, then the firm's labor demand is

$$z(w) = \begin{cases} 0 & \text{if } \mu < w \\ [0, \infty] & \text{if } \mu = w \\ \infty & \text{if } \mu > w. \end{cases}$$

- **Competitive Equilibrium.** It follows that in any equilibrium with employment, if w^* is the equilibrium wage and Θ^* is the set of worker types who choose employment at that wage, then (w^*, Θ^*) jointly satisfy

$$\Theta^* = \{\theta \mid r(\theta \leq w^*\} \tag{1}$$

$$w^* = E[\theta \mid \theta \in \Theta^*] = \frac{1}{F(\theta^*)} \int_{\Theta^*} \theta dF(\theta). \tag{2}$$

Define the set of competitive equilibria as

$$CE \equiv \{(w^*, \Theta^*) \mid \Theta^* = \{\theta \mid r(\theta \leq w^*\} \text{ and } w^* = E[\theta \,|\theta \in \Theta^*]\} \,.$$

Note in general the set of CE may include more than one point.

Special case: Type-independent outside option. Before continuing further, let's consider the possible equilibria for the simple case in which $r(\theta) = r$ for all $\theta \in \Theta$. There are two possibilities in this case. Either $w^* \geq r$, in which case $\Theta^* = \Theta$ (all workers accept employment, or $w^* < r$, in which case $\Theta^* = \emptyset$ and all workers choose their home-production option. Which of the two cases arises depends entirely upon the unconditional expected worker productivity, $E[\theta]$.

- If $r \leq E[\theta]$, then the market wage is $w^* = E[\theta]$ and every worker accepts an offer from some firm.
- If $r > E[\theta]$, then it is a market equilibrium for $w^* = E[\theta]$ and all workers reject the market wage offer. For simplicity, we will assume (as does MWG) that $w^* = E[\theta]$ if $\Theta^* = \emptyset$. Note that any $w^* > E[\theta]$ is also a market equilibrium leading to no trade.

Note that in the case of no employment, $r > E[\theta]$, it must be that $r > \underline{\theta}$ and therefore some unemployment would have been Pareto efficient in the full-information setting.

In what follows, we make two assumptions:

1. We assume that $\theta \geq r(\theta)$ for all $\theta \in \Theta$. Thus the full-information allocation is full employment; this is also the unique Pareto-optimal allocation.
2. We assume that $r(\theta)$ is nondecreasing. This implies that selection, when it occurs, is *adverse*. Specifically, we say that **selection is adverse** in a market when an informed individual's trading decision depends upon private information in a way that adversely impacts the uninformed agents in the economy. In the case in which $r(\theta)$ is strictly increasing, workers who are keenest to work (i.e., low r's) are those with low marginal products of labor.

We now explore the set of competitive equilibria. To this end, it is useful to make the technical assumption that $E[\theta \,|r(\theta) \leq w]$ is a continuous function of w. We can then graph $E[\theta \,|r(\theta) \leq w]$ with w on the horizontal axis and labor productivity, θ, on the vertical axis and locate the competitive equilibria as fixed points where $E[\theta \,|r(\theta) \leq w]$ crosses the 45-degree line.

- Note that by construction, $\underline{\theta} = E[\theta \,|r(\theta) \leq r(\underline{\theta})]$, so $(r(\underline{\theta}), \underline{\theta})$ is a point in the graph
- For $w > r(\overline{\theta})$, we have $E[\theta] = E[\theta \,|r(\theta) \leq r(\overline{\theta})]$, so $E[\theta \,|r(\theta) \leq w] = E[\theta]$ for all $w > r(\underline{\theta})$
- By continuity, we can conclude that if $\underline{\theta} > r(\underline{\theta})$, then there exists a competitive equilibrium in which $\Theta^* \neq \emptyset$. If we also have $E[\theta] < r(\overline{\theta})$, then there is less than full employment, $\Theta^* \subsetneq \Theta$.

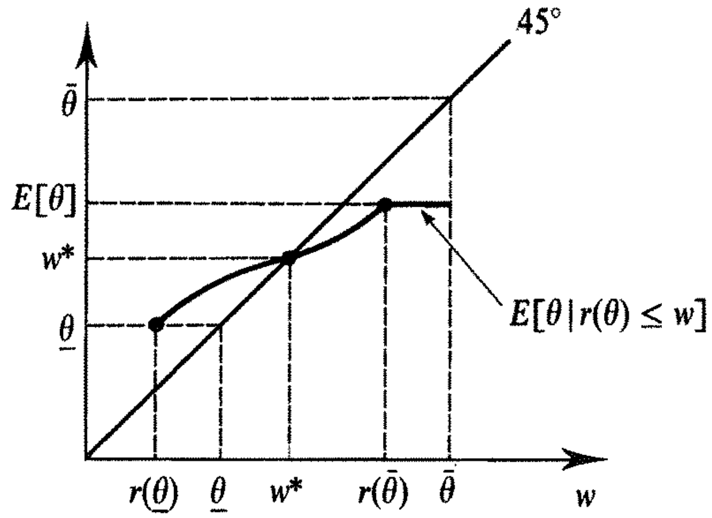

Figure 2: (MWG, Figure 13.B.1)

Example of unique interior equilibrium, $w^* \in (\underline{\theta}, \overline{\theta})$:

There are a few other special cases worth our attention.

- **No employment even though** $\theta \geq r(\theta)$ **for all** $\theta \in (\underline{\theta}, \overline{\theta})$**:** This is similar to the no-employment outcome case in which we assumed $r(\theta) = r$, except now we do not require $r(\underline{\theta}) > \underline{\theta}$. In particular, suppose that $r(\underline{\theta}) = \underline{\theta}$, but $r(\theta) < \theta$ for all $\theta \geq \underline{\theta}$. If $\frac{d}{dw}E[\theta \,|r(\theta) < w] < 1$, then the equilibrium has no measure of employment.

- **Multiple equilibria.** There is no reason why $E[\theta \,|r(\theta) \leq w]$ crosses the 45 degree line only once. In the figure below from MWG, there are three comeptitive equilibria wages, w_1^*, w_2^*, w_3^*.

  Note that these equilibria are Pareto ranked. Firms earn zero profits in every equilibrium. All workers are weakly better off with higher market wages; some types benefit strictly. Because of this, it is useful to define the set of competitive equilibrium wages and to single out the highest wage:

  $$W^* \equiv \{w^* \mid \exists \Theta^* \; s.t. \; (w^*, \Theta^*) \in CE\},$$

  and

  $$\overline{w}^* = \max \; W^*.$$

  Hence, $\overline{w}^*$ denotes the highest (and Pareto best) competitive equilibrium wage. In figure 4, $\overline{w}^* = w_3^*$.

1.1.2 Game-theoretic approach

In the previous analysis, we looked at market equilibria (i.e., equilibria in which both sides of the market were price takers with respect to the equilibrium wage). We now

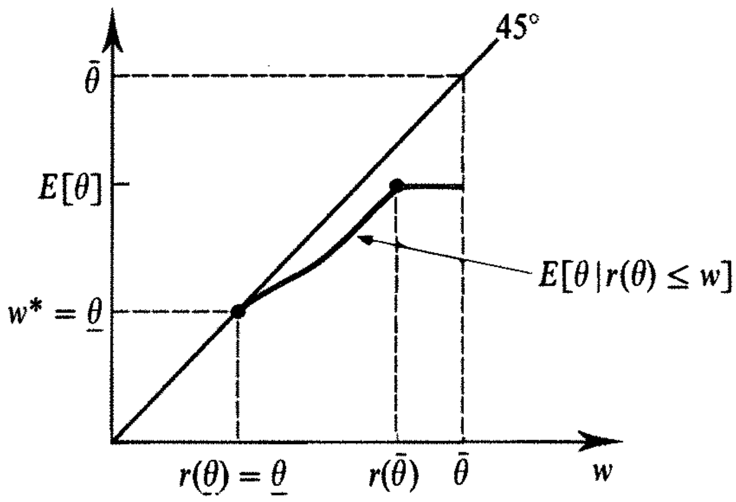

Figure 3: (MWG, Figure 13.B.2)

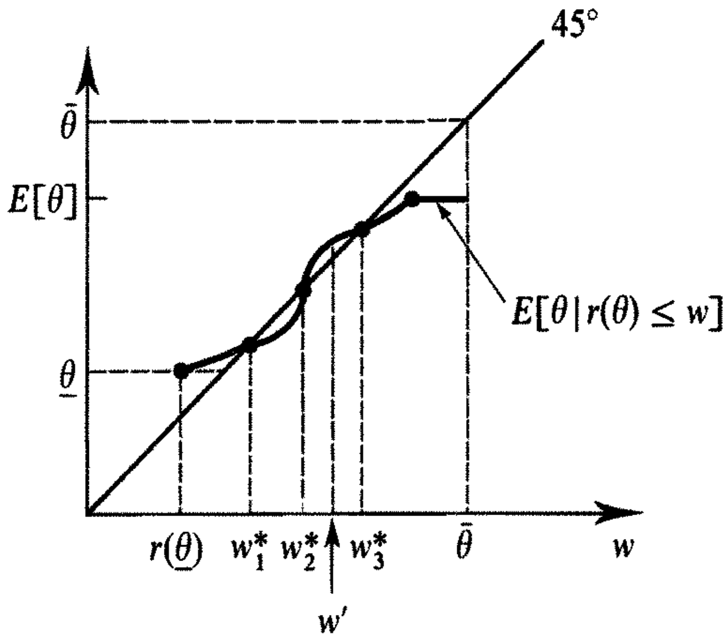

Figure 4: (MWG, Figure 13.B.3)

want to consider a game in which firms are allowed to deviate and offer an alternative wage to w^*. With the discipline imposed by allowing this strategic deviation, we want to reconsider the set of competitive equilibria and determine if some subset survives this strategic option.

Here's some preliminary intuition from Figure 4. Suppose that w_2^* was the equilibrium but a firm (who under the posited equilibrium would earn zero profit) could deviate and offer another wage. By offering w', he would attract a worker from the set $\{\theta \in \Theta \ |r(\theta) \leq w'\}$ who has an expected productivity above w'. Because this is a profitable deviation available to all firms, it seems that w_2^* cannot be an equilibrium in the more general game allowing for wage deviations. A similar analysis applies to w_1^*. The wage w_3^*, however, does not seem to suffer from such deviations. Specifically, there is no upward deviation above w_3^* that is profitable. We formalize this below.

Game:

- Stage 1: firms simultaneously announce wage offers
- Stage 2: workers decide whether to work for a firm and, if so, which one. They randomize among firms with identical wage offers.

Proposition 1. *(MWG,13.B.1)*

1. *If $\overline{w}^* > r(\underline{\theta})$ and if there is an ε such that $E[\theta \ |r(\theta) \leq w'] > w'$ for all $w' \in (\overline{w}^* - \varepsilon, \overline{w}^*)$, then in every SPNE, two or more firms offer $\overline{w}^*$ and all workers with $r(\theta) \leq \overline{w}^*$ accept these offers.*
2. *if $\overline{w}^* = r(\underline{\theta})$, then there are multiple SPNE, but in all the worker's equilibrium payoff equals $r(\theta)$.*

Note that the second condition under (i) is equivalent to the statement that the function $E[\theta \ |r(\theta) \leq w]$ crosses the 45-degree line at $\overline{w}^*$ from above. This corresponds to $w_3^* = \overline{w}^*$ in Figure 3. This condition excludes the non-generic pathology in the following figure: Regarding this case, the proposition is silent, but it is straightforward to see that both competitive equilibria in the figure will survive in the two-stage noncooperative game.

Proof of Proposition 1: (Sketch) The details are in MWG; I will only sketch it here. I will also assume that there are only 2 firms (enough for Bertrand competition) to make it simpler.

Claim (i): Step 1. Both firms must make zero profits in equilibrium. If one of the firms was making positive profits, say firm i, then firm i offers a wage $\overline{w}$ such that

$$E[\theta \ |r(\theta) \leq \overline{w}] > \overline{w}$$

and which some worker accept. There are two possibilities. Either firm j offers a higher wage which no worker takes, or firm j offers the same wage as i. In either case, firm j could then deviate and offer $\overline{w} + \varepsilon$ and steal firm i's workers, making a positive profit.

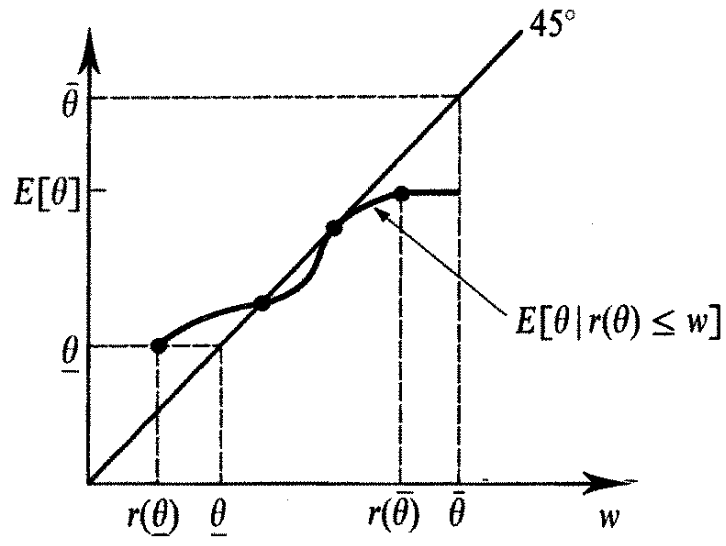

Figure 5: (MWG, Figure 13.B.4)

Hence, firms cannot make positive profits. Of course, in a SPNE firms cannot make negative profits either as they always have the option to set a wage in excess of $r(\overline{\theta})$ and hire no one.

Step 2. From step 1, we know that either the highest wage offered is a competitive equilibrium wage, $\overline{w} \in W^*$, or $\overline{w} < r(\underline{\theta})$ and no workers are employed. In either case, if $\overline{w} < \overline{w}^*$, a firm can deviate to a wage $w' \in (\overline{w}^* - \varepsilon, \overline{w}^*)$ and earn positive profits. Thus, the highest offered wage must be $\overline{w}^*$. [This is the core intuition above.]

Step 3. Both firms offering $\overline{w}^*$ is a SPNE. Any downward deviation by a firm would result in zero employment (and zero profits). Because $E[\theta \mid r(\theta) \leq w'] < w'$ for all $w' > \overline{w}^*$, there is no profitable upward deviation either.

<u>Claim (ii)</u>: By hypothesis, $E[\theta \mid r(\theta) \leq w] < w$ for all $w > \overline{w}^*$, so any upward deviation is unprofitable. Any downward deviation leads to no worker acceptance. Thus, in any equilibrium, firms offer $w^* \leq \overline{w}^*$, and no measurable set of workers is employed. $\square$

1.1.3 Is highest-wage competitive equilibrium also Pareto optimal?

Suppose that the government understood all of the details of the adverse-selection market, except the government (like the firms) does not observe the workers' types. Can the government improve on the competitive equilibrium allocations? Specifically, can it find an allocation that is Pareto superior to a market allocation (weakly better for all and strictly better for some)? If not, then we say that the market allocation is **constrained-Pareto optimal**.

To answer this question, we will impose additional structure. Suppose that F is continuous with density $f(\theta)$ and $r(\theta)$ is strictly increasing (and as before, $r(\theta) \leq \theta$ so full-employment is Pareto optimal). We assume that our central authority cannot dictate full employment; rather, workers are free to make their own decisions regarding employment. Because our central authority cannot observe θ, worker payments can only be made conditional on whether or not a worker is employed at a firm. Thus, the instruments available to the central authority are w_e and w_u. We assume that our central authority can tax all of the firms' profits, but must have a balanced budget. Note that for any set of payments, (w_e, w_u), we can define a type $\hat{\theta}$ who is indifferent between working and not:

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

(We'll set $\hat{\theta} = \underline{\theta}$ if $w_u + r(\underline{\theta}) < w_e$ and $\hat{\theta} = \overline{\theta}$ if $w_u + r(\overline{\theta}) > w_e$.) Because all types $\theta \leq \hat{\theta}$ will work, budget balance requires that

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

Using the two equations above, we see that for any threshold $\hat{\theta}$, the associated (w_u, w_e) wages are uniquely determined:

$$w_u(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta) d\theta - r(\hat{\theta}) F(\hat{\theta}) = F(\hat{\theta})(E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta})),$$

$$w_e(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta) d\theta + r(\hat{\theta})(1 - F(\hat{\theta})) = F(\hat{\theta})(E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta})) + r(\hat{\theta}).$$

We can therefore think about the central authority's problem as choosing a $\hat{\theta}$ to improve welfare.

We want to prove the following:

Proposition 2. *(MWG, Prop 13.B.2) The high-wage competitive equilibrium is constrained-Pareto optimal.*

Define $\hat{\theta}^*$ to be the threshold employment type in the highest-wage CE equilibrium. That is, $\overline{w}^* = r(\hat{\theta}^*)$. Now consider a choice by the central authority that is not $\hat{\theta} \neq \hat{\theta}^*$.

Case 1: Suppose $\hat{\theta} < \hat{\theta}^*$. Then $r(\hat{\theta}^*) > r(\hat{\theta})$ and from above

$$w_e(\hat{\theta}) \leq F(\hat{\theta})(E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta}^*)) + r(\hat{\theta}^*),$$

thus

$$w_e(\hat{\theta}) - r(\hat{\theta}^*) \leq F(\hat{\theta}) \left(E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta}^*) \right),$$

or finally

$$w_e(\hat{\theta}) - r(\hat{\theta}^*) \leq F(\hat{\theta}) \left(E[\theta \mid \theta \leq \hat{\theta}] - E[\theta \mid \theta \leq \hat{\theta}^*] \right) < 0.$$

This implies $\overline{w}^* > w_e(\hat{\theta})$ and hence a $\underline{\theta}$-type worker will be worse off.

Case 2: Suppose $\hat{\theta} > \hat{\theta}^*$. We know $E[\theta \,|r(\theta) \leq w] < w$ for all $w > \overline{w}^*$. By construction, $r(\hat{\theta}^*) = \overline{w}^*$ and because r is strictly increasing, $r(\hat{\theta}) > r(\hat{\theta}^*)$, thus for all $\hat{\theta} > \hat{\theta}^*$

$$E[\theta \,|r(\theta) \leq r(\hat{\theta})] < r(\hat{\theta}).$$

But because r is strictly increasing, this is equivalent to

$$E[\theta \,|\theta \leq \hat{\theta}] < r(\hat{\theta}), \; \forall \, \hat{\theta} > \hat{\theta}^*.$$

But then $w_u < 0$ and $\overline{\theta}$-type workers are worse off. $\square$

1.2 Insurance-market model (JR)

- Each consumer has a probability of suffering an accident, $\pi \in \Pi \equiv [\underline{\pi}, \overline{\pi}] \subseteq [0, 1]$, where the monetary cost of the accident is L dollars;
- Each consumer privately knows their accident type, π, but the insurance firms only know the distribution of types, F, defined over the support $[\underline{\pi}, \overline{\pi}]$.
- Consumers are risk averse with vNM utility represented by the strictly concave function, $u(\cdot)$; consumers have initial wealth, y.
- Firms are risk neutral.
- For now, we assume only full-insurance policies can be sold.

1.2.1 Full-information equilibrium benchmark

As a full-information benchmark, suppose that insurers can condition the price of their full-insurance policies on the consumer's accident probability – i.e., there is full information. In this case, perfect competition among insurers requires that the price of each policy is set to earn zero profits:

$$p^{fb}(\pi) = \pi \cdot L.$$

At these prices, all consumers purchase full insurance and obtain expected utility $u(y - \pi L)$.

1.2.2 Competitive equilibrium with incomplete information

Now return to our setting in which only consumers observe π; firms are uninformed. Consider the π-type consumer's decision. It is optimal to buy a full-insurance policy at price p if

$$u(y - p) \geq \pi u(y - L) + (1 - \pi) u(y).$$

Rearranging, it is optimal to buy if

$$\pi \geq h(p) \equiv \frac{u(y) - u(y - p)}{u(y) - u(y - L)}.$$

Note that $h(p)$ is increasing in p. We thus have a setting of **adverse selection** because the consumer's who are most interested in buying full insurance are those with higher accident probabilities (and this trade adversely impacts the uninformed agents in the market).

As in the case of the labor market setting, a competitive equilibrium will be a price, p^*, and a set of types, Π^*, such that

$$p^* = E[\pi \mid \pi \in \Pi^*]L,$$

$$\Pi^* = \{\pi \mid \pi \geq h(p^*)\}.$$

In terms of the premium alone, we have the condition

$$p^* = E[\pi \mid \pi \geq h(p^*)]L.$$

Analogous to the labor-market setting, we have several results:

1. Existence. It is useful to define the expected cost of a full-insurance contract as a function of p on the domain $[0, \overline{\pi}L]$:

   $$g(p) = E[\pi \mid \pi \geq h(p)]L.$$

   Note that for all $p \in [0, \overline{\pi}L]$, $h(p) \in [0, \overline{\pi}]$. (We know that a $\overline{\pi}$-type consumer will buy actuarial fair full-insurance at $p = \overline{\pi}L$, so $h(p) < \overline{\pi}$.) From this it follows that $E[\pi \mid \pi \geq h(p)]L \leq \overline{\pi}L$. Hence, g maps from $[0, \overline{\pi}]$ to itself. Moreover, it is a nondecreasing function of price. We can thus use Tarski's fixed-point theorem to conclude there must be a fixed point $p^* = g(p^*)$ which is a competitive equilibrium. (If g is continuous, we could also apply Brouwer's FP theorem too, but if the distribution of π is discrete, g is not necessarily continuous.)

2. Inefficient equilibria always exist if $\overline{\pi} = 1$. Specifically, let $p^* = \overline{\pi}L = L$ and thus only the highest type, $\overline{\pi}$, purchases full insurance, which is entirely inconsequential. This is because $h(L) = 1$ and thus $g(L) = L$. Contrast this with the labor-market setting in which it is possible that the unique competitive equilibrium has full employment (i.e., assume $r(\theta) = r < E[\theta]$).

3. Equilibria in which all consumers purchase insurance exist if

   $$\underline{\pi} \geq h(E[\pi]L).$$

   In this case, the competitive equilibrium is $p^* = E[\pi]L$ and all types purchase.

4. Some consumers do not buy insurance in any competitive equilibrium if

   $$\underline{\pi} < h(E[\pi]L).$$

   Thus, if $\Pi = [0, \overline{\pi}]$, there will be an interval of types who do not insure.

5. Multiple competitive equilibria may exist. This is hardly surprising. Here is a quick example. Suppose that $\pi \in [0, 1]$ with density $f(\pi) = 2(1 - \pi)$, $u(x) = x - \frac{1}{4}x^2$ and $y = 2$ (with $y > L$). Then there are two equilibria, $p^* = L$ and $p^* = \frac{L}{2}$.

6. If there are multiple competitive equilibria, then they are Pareto ranked. Firms earn zero profits in all equilibria. Consumers weakly prefer lower prices, and those who buy insurance strictly benefit. Thus the best competitive equilibrium is the one with the lowest price.

7. Two-stage game. We can embed the competitive insurance market into a two-stage game as in MWG and show that (excepting pathological examples), the unique equilibrium outcome coincides with the low-price competitive equilibrium.

We will return to the labor-market and insurance market settings when we consider endogenous contracts for the purpose of signaling or screening private information. For now, however, we present a few additional ideas related to adverse-selection and advantageous selection in insurance markets.

1.3 *Advantageous* versus *Adverse* selection in markets

[Here, I follow Einav, Finkelstein, "Selection in Insurance Markets," *Journal of Economic Perspectives*, 2011; see also Einav, Finkelstein, Cullen (*QJE*, 2010).]

Consider our results for adverse selection in competitive markets when only full-insurance can be sold. [When we consider signaling and screening models, we'll endogenize the choice of insurance coverage.] If selection is *adverse*, we mean that those who are willing to pay the most for insurance also have the highest risks. For example,

$$E[\pi L | \pi \geq \hat{\pi}]$$

gives the the average cost of insuring the population of consumers with $\pi \in [\hat{\pi}, \overline{\pi}]$. As $\hat{\pi}$ decreases (i.e., more people buy insurance), the average cost of a policy falls. Thus we can think about putting the fraction of the population who buys insurance on the horizontal axis and plot costs on the vertical axis. This is the average cost curve, but it is also the market supply curve given there is perfect competition. On the same graph, we can illustrate the demand relationship as the fraction of the population who have $\pi \geq h(p)$, which is a downward sloping curve. Where these demand and supply curves intersect is the fraction of the market that purchases insurance. We can also graph in the same figure (Figure 6) the marginal cost of an additional measure of consumers. MC is falling, but because the AC falls, it is everywhere below the AC curve.

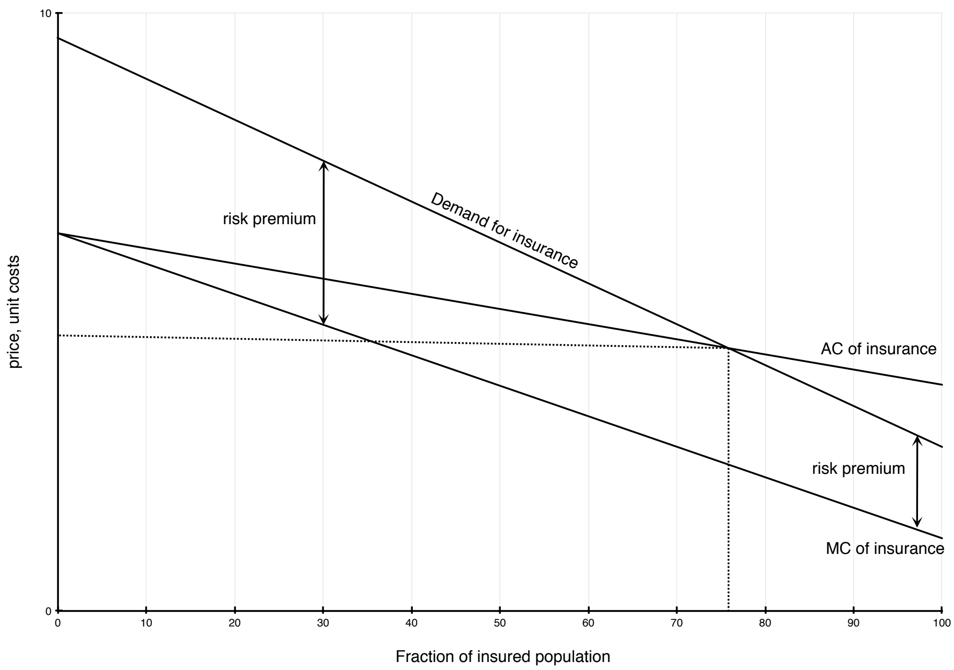

Figure 6: Adverse selection and unit cost curves

Notice that market equilibrium is driven by the intersection of demand and AC, but efficiency requires coverage to occur at the intersection of demand and MC. In this figure, as is typical in the textbook models, the market provides too little insurance. But recall that we can also construct examples where the market unravels, but no insurance is socially efficient.

Now consider *advantageous* or beneficial selection. This can arise in a richer model of preference heterogeneity. Suppose that high-risk individuals are also less risk averse. Or, if we enter the world of behavioral economics for the moment, perhaps insurance and safety are demand complements and normal goods, so that rich people buy insurance (for a sense of well-being) and also choose lower-risk environments, while poor people are more likely to pass on insurance and to also make risky choices in their lives. Regardless, when advantageous selection is present, the marginal cost of the (marginal) person buying insurance is higher than the average cost of the population buying insurance, because the low-risk consumers are inframarginal. As a result, average cost is upward sloping, and marginal cost lies above average cost. Our figure (Figure 7) is now flipped.

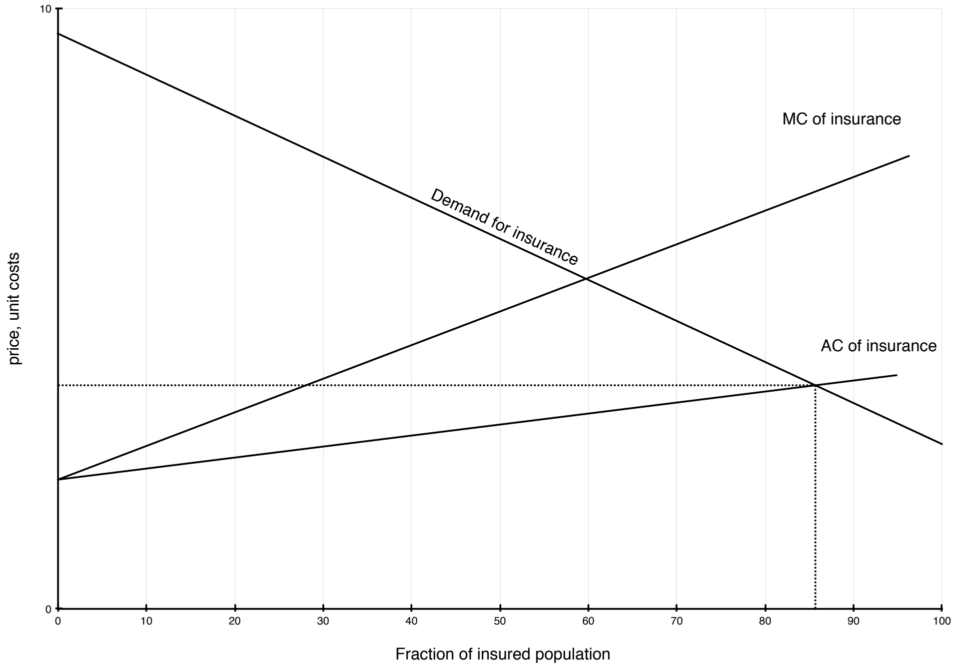

Figure 7: Advantageous selection and unit cost curves

One of the big empirical research programs of recent years has been to test whether a market has adverse or advantageous selection. Look at the Einav, Finkelstein *JEP* (2011) paper for a quick survey of previous work.

In a more recent paper, Mahoney and Weyl, "Imperfect Competition in Selection Markets," *REStat.*, 2017, have done a similar analysis contrasting advantageous and adverse selection as Einav and Finkelstein, but with imperfect competition. To accomplish this, they build a model of conjectural variations (which nests monopoly and perfect competition as special cases). The interesting point that they make is that in the world of advantageous selection, imperfect competition has the marginal benefit or reducing supply (which a competitive market may be excessively produced). The basic logic can easily be seen by introduing a monopolist's marginal revenue curves into our previous graphs. The figure below is from Mahoney and Weyl and tells the whole story:

1.4 Adverse selection: No-trade theorem

There is a close connection between unraveling in markets of adverse selection and "No-Trade" theorems. The latter basically claims that if rational traders have a common prior and start with an initial allocation of goods that is Pareto optimal, then after receiving additional private information, meaningful trade will not take place. Specifically, *if it is common knowledge that every player weakly prefers to trade relative to no trade, then every agent must be indifferent between trade and no trade.* The seminal paper with this result is Milgrom and Stokey (1982), "Information, Trade and Common Knowledge," *JET*, 26(1). The

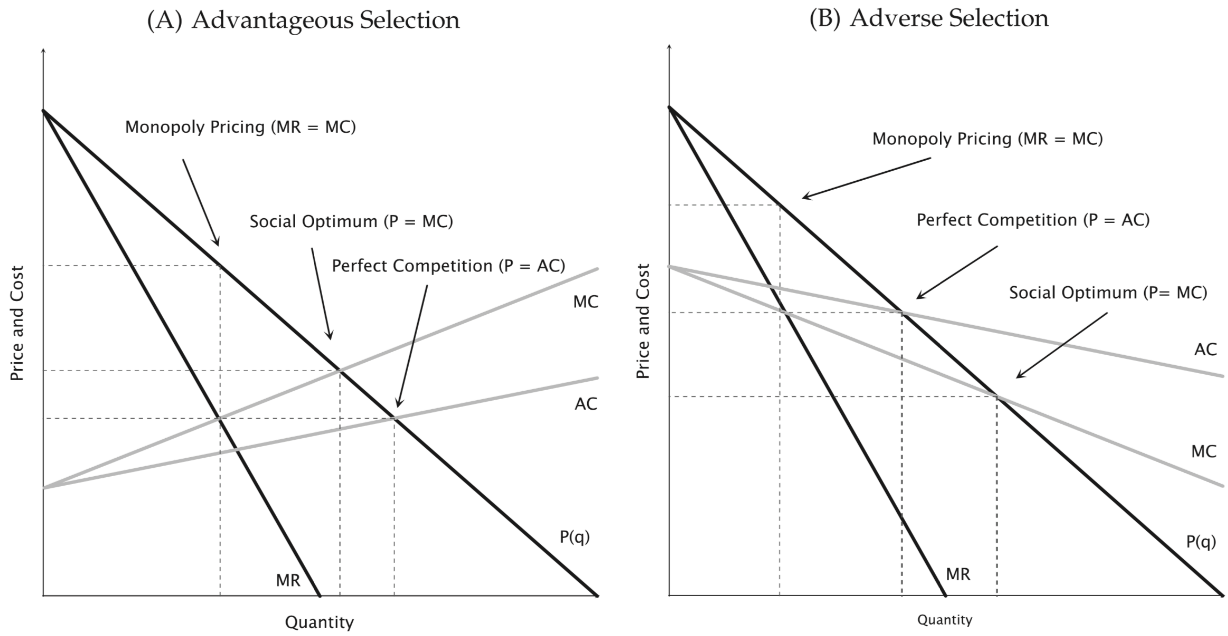

Figure 8: Advantageous versus adverse selection with monopoly and competition (from Mahoney and Weyl, 2017)

intuition is the same as with market equilibria with selection: all parties to the trade will condition expectations on their private signal and on the willingness of the other agent(s) to trade. This conditioning is what generates the no trade result. I'll give a toy example here to illustrate the reasoning, though it is far more general allowing for risk aversion, multiple goods and arbitrary endowments.

Simple example of no trade. Consider the case in which an asset is owned by player 1 today, and it's value tomorrow (date 2) is θ. Player 2 does not own any asset today. Both players share the same common prior over θ and both are risk neutral. Hence, the original allocation is Pareto optimal. At date 1, each player i receives a private signal, $s_i \in \Sigma_i$. It is common knowledge that the joint distribution of the asset value and individual signals is given by $\phi(\theta, s_1, s_2)$, which is common knowledge. For simplicity, assume that Σ_i is finite.

To make the intuition clear, assume the players engage in a simple trading game in which there is an exogenous posted price, p, and after observing their private signals but prior to observing the asset value θ, each player simultaneously chooses either "trade" or "no trade". If both choose "trade", they trade at price p; if either chooses "no trade", trade does not take place. Can the signals ever be such that player 1 and player 2 agree to meaningful trade for any price p?

To see what happens, define the set of Σ_1 signals for which player 1 is willing to sell her asset at the posted price as $T_1 \subseteq \Sigma_1$, and define the set of signals for which player 2 is willing to buy the asset at price p as $T_2 \subseteq \Sigma_2$. Then for player 1 to be willing to trade, it must be that for all $s_i \in T_1$, her expected returns (conditional on $s_2 \in T_2$) are nonnegative:

$$E[\theta \,|s_1, s_2 \in T_2] \leq p.$$

Because this holds for all $s_1 \in T_1$, it must also hold if we condition on the set T_1:

$$E[\theta \,|s_1 \in T_1, s_2 \in T_2] \leq p.$$

Now consider player 2. If $s_2 \in T_2$, it must be that he expects to earn nonnegative returns conditional on $s_1 \in T_1$:

$$E[\theta \,|s_1 \in T_1, s_2] \geq p.$$

Summing over all $s_2 \in T_2$, we have the implication that

$$E[\theta \,|s_1 \in T_1, s_2 \in T_2] \geq p.$$

Together, we must have

$$p = E[\theta \,|s_1 \in T_1, s_2 \in T_2].$$

Hence, in any states for which trade arises, it is common knowledge that trade is not beneficial. Hence, for the set of signals for which there is trade, $(s_1, s_2) \in T_1 \times T_2$, it is common knowledge by both parties that the expected value of the asset is equal to p and thus both players are indifferent between trade and no trade. Meaningful trade conditional on new information cannot arise.

2 Signaling by Disclosing Certifiable Information

Given the Pareto inefficiencies caused by private information in the previous models of competitive markets, it seems natural that the existence of a *costless* signal or certification technology – either for worker productivities or consumer risks – might improve welfare. In this section, we assume that such a costless technology exists, but leave the control of the technology in the hands of the party with private information. Specifically, we suppose that a worker (for example) can send a credible and verifiable message that her type is in some subset of the type space, $m \subseteq \Theta$, and that her message can be relied upon by the labor market. (That's what we mean by certified information.) But perhaps some worker does not want to reveal their private information; in that case, what should we expect will be revealed in equilibrium? The answer is that under general conditions, one should expect the market to unravel with every worker revealing their type (except possibly the lowest type). The intuition is that if any workers are pooled in equilibrium, then the best worker in the pool has an incentive to deviate and reveal her type to avoid being pooled with lower individuals. Here's a toy model in the context of our MWG labor-market mmodel where the result is easy to see.

Suppose that there are n types of workers, $\theta_1 < \cdots < \theta_n$, each with a positive probability of arising, ϕ_i. Suppose that $\theta > r(\theta)$ for every worker, so each worker would prefer to be employed if the worker could get $w = \theta$. Finally, suppose that each worker of type θ_i can certify to the labor market that her ability is θ_i at no cost. The timing of the game is

(1) the workers choose whether or not to publicly certify their types to the labor market, and (2) the firms in the labor market simultaneously offer wages to each worker that may condition on any certified information. Here's why full-information must emerge in any equilibrium. Suppose otherwise that that two or more types decide not to certify with some probability in an equilibrium. Let θ_k be the highest type that chooses not to certify with some probability. When θ_k does not certify, the market assigns an expected value to her that is lower than θ_k because the market groups type θ_k with lower type(s) that also do not certify. Because it is costless to reveal her type and she benefits from a higher wage, the θ_k worker will deviate and provide evidence of her type. This logic applies to any type other than the lowest, θ_1. But if the other $n-1$ higher types show evidence of their type, the lowest type is discovered whether or not she reveals evidence.

The seminal articles which first formalized this intuition are Milgrom (1981), Grossman and Hart (1980) and Grossman (1981). A paper by Okuno-Fujiwara, Postlewaite and Suzumura (OFPS) (*REStud.* 1990) generalizes this result by focusing on two key sufficient conditions, and by illustrating with a series of examples that violating either condition can prevent unraveling.[2] What should be clear from the intuition above is that there is a requirement that types are ordered, running from best to worst, and that each privately-informed individual prefers to be thought of as a "better" type. Less obvious, the certification technology must allow the type-θ player to send messages of the form "I am at least type θ." OFPS formalize these conditions.

First, we need to be clear about the game. Each player i draws a type $\theta_i \in \Theta$ according to some known distribution function. For simplicity, let's assume that Θ_i is finite with types running from $j = 1, \ldots, n$, with $\theta_1 < \cdots < \theta_n$, and the probabilities of each type are given by $\phi = (\phi_1, \ldots, \phi_n)$. Assume that types are independently distributed across players. After nature draws the players' types, each player i simultaneously sends a certifiable public message, m_i, that is any subset of Θ_i that contains θ_i. Upon viewing all of the messages, players form beliefs about each other player, $\mu_i(\cdot|m_i)$. Players then play some stage game with equilibrium payoffs given in reduced form by

$$u_i^*(\mu_i, \theta_i).$$

In our labor-market example, this object is simply

$$u^*(\mu, \theta) = E_\mu[\theta].$$

We need once last concept. We say that beliefs are *skeptical* following message m_i from player i if the beliefs place probability 1 on the message coming from the lowest feasible type, $\min\{\theta \in m_i\}$, with $\mu_i(\min\{\theta \in m_i\}|m_i) = 1$.

We can now state a variation of OFPS:

Proposition 3. *Suppose that the following two conditions are satisfied:*

1. *For any player i and type θ_i, player i can send a certified message such that θ_i is the minimum type of the message;*

[2] A more recent paper by Seidmann and Winter (*Ema*, 1997) further generalizes disclosure-game results.

2. *for each player i, if belief μ_i stochastically (first-order) dominates $\tilde{\mu}_i$, i.e.*

$$\sum_{j=1}^{k} \tilde{\mu}_i(\theta_j) \geq \sum_{j=1}^{k} \mu_i(\theta_j), \text{ for all } k \in \{1, \ldots, n\} \text{ with strict inequality for some } k$$

then the final stage payoff satisfies for any θ_i

$$u_i^*(\mu_i, \theta_i) > u_i^*(\tilde{\mu}_i, \theta_i).$$

Then the equilibrium involves full revelation of information and skeptical market beliefs.

Proof: First, note that we can rule out incomplete disclosure in equilibrium. If two or more players send the same message, the higher type can send a certified statement separating her from lower types (condition 1) and strictly raise her payoff (by condition 2).

Second, suppose beliefs are not skeptical for some message, m, that is sent in equilibrium. In this case, it must be that m contains two or more types, and the lowest type in m can send m and be thought better than her type and therefore raise her payoff above the full-information equilibrium payoff. A contradiction.

Third, it is clear that it is an equilibrium to send messages $m_i(\theta) = \{\theta\}$ if beliefs are skeptical. $\square$

Remarks:

1. Note that unraveling requires an implicit assumption that certification is costless.

2. Note also that if some player is unable to certify their type with some probability, then condition 1 is violated.

3 Signaling in competitive markets

We now assume that costless, verifiable signaling is not available. In what follows, we address situations in which an informed player (e.g., the workers or consumers in our previous labor market and insurance market examples) moves first and (possibly) signals private information to the uninformed side of the market.

Following Spence (1973), we begin our analysis with the role of education as a signal of worker productivity. We then return to our insurance market example to consider how partial coverage can arise in equilibrium as a signal of low risk.

3.1 Labor-market model (MWG)

Return to our labor market setting but with one significant change. We assume that workers choose education, $e \geq 0$, to possibly signal productivity to the labor market prior to obtaining a wage offer; the labor market observes the level of education, makes an inference about the worker's productivity, and then offers a wage.

Simple 2-type model of labor-market signaling:

- worker's productivity, θ, takes on two values, $\theta_h > \theta_l > 0$, with the high productivity type arising with probability ϕ.
- for now, we assume that $r(\theta) = 0$, so it is efficient for both worker types to be employed
- workers can choose a level of education, $e \geq 0$, with cost $c(e, \theta)$; the worker's payoff from working at wage w, having chosen education level, e, is

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

  where $c(0, \theta) = 0$, $c(\cdot, \theta)$ is increasing and convex in e, $c(e, \cdot)$ is decreasing in θ and $c_{e\theta}(e, \theta) < 0$
- firms only care about wages and productivity; education has no direct value; we can relax this assumption later

The key assumption about c is that higher productivity workers have lower marginal costs of education.[3] This is often referred to as the "single-crossing property" because of the implication that the indifference curves of any two types of workers will cross just once. To be clear, think of a worker's indifference curve in effort-wage space. The marginal rate of substitution of for a type-θ worker's indifference curve is

$$MRS_{e,w}(\theta) = -\frac{MU_e(\theta)}{MU_w} = c_e(e, \theta).$$

Higher types have flatter indifference curves given $c_{e\theta} < 0$ as illustrated in Figure 9 for indifference curves intersecting at $(\hat{e}, \hat{w})$.

[3]Note that $c_{e\theta} < 0$ and $c(0, \theta) = 0$ implies $c_\theta < 0$, so single-crossing is the key assumption: specifically,

$$c(e, \theta) - c(0, \theta) = \int_0^e c_e(x, \theta) de;$$

differentiating with respect to θ (and using $c_\theta(0, \theta) = 0$) yields

$$c_\theta(e, \theta) = \int_0^e c_{e\theta}(x, \theta) de < 0.$$

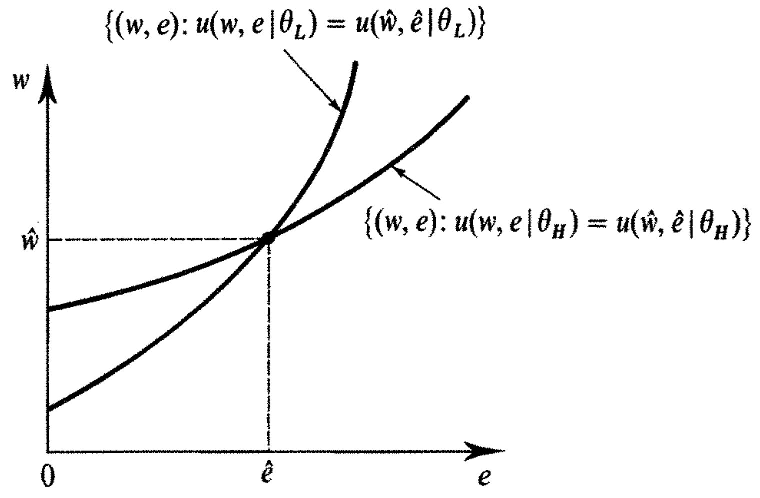

Figure 9: (MWG, Figure 13.C.2)

The signaling game: We consider perfect Bayesian equilibria in what follows. This requires actions be sequentially rational with respect to beliefs, that beliefs satisfy Bayes' rule along the equilibrium path, and that off-path beliefs are consistent across players with the same information. In the context of signaling games, this is the same as focusing on sequential equilibria, but we must make extra efforts to extend the definition of consistency to a setting in which actions (i.e., education levels) are infinite. JR(ch8) explains how to think about this. We will also restrict attention to pure-strategy equilibria.

Formally, let $e(\theta)$ be a worker's pure-strategy choice of education, let $\mu(\theta_h \mid e) = \mu(e)$ be the firms' belief that a worker who selected e is θ_h, let $w_i(e)$ be firm i's wage offer conditional on chosen e. We leave unmodeled the worker's choice among wage offers and assume that she randomly chooses among the set of highest offered wages. Figure 10 shows the (partial) game tree:

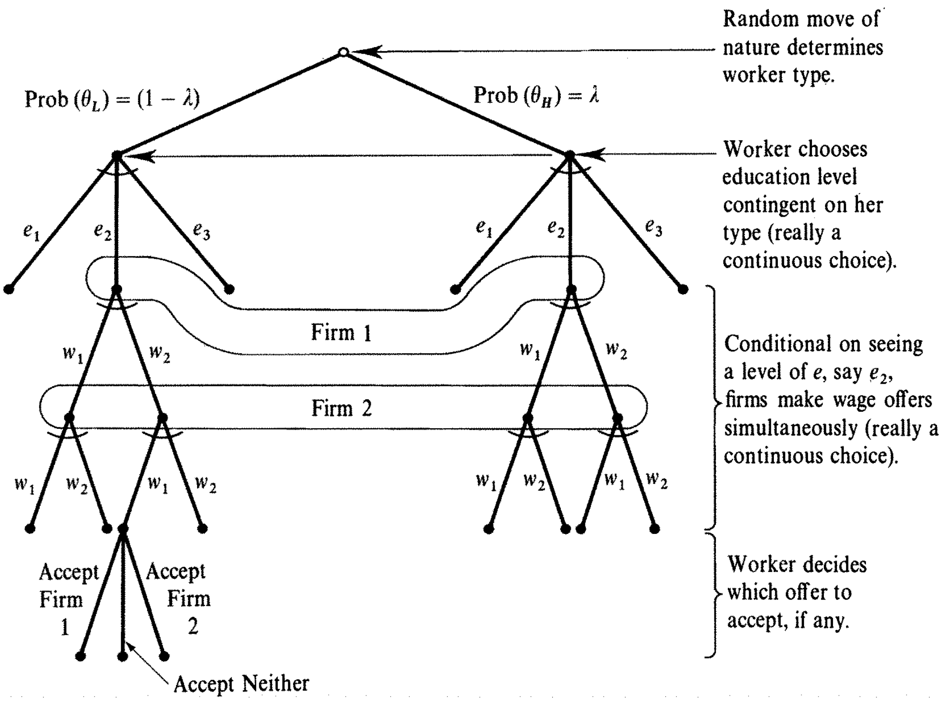

Figure 10: (MWG, Figure 13.C.1)

A **perfect Bayesian equilibrium (PBE)** in the labor-market signaling game is a strategy profile $\{e(\cdot), \mu(\cdot), (w_i(\cdot))_{i=1}^N\}$ such that

- worker's strategy, $e(\theta)$, is optimal given $(w_i(\cdot))_i$
- $\mu(e)$ is consistent with Bayes' rule wherever possible
- the wage offers $w_1(\cdot), \ldots, w_n(\cdot)$ are a Nash equilibrium in the simultaneous-move wage offer game in which each firm's belief is $\mu(e)$.

Consider the final stage. For any shared belief, $\mu(e)$, the expected productivity of the worker is

$$\mu(e)\theta_h + (1 - \mu(e)\theta_l.$$

The simultaneous-move wage offer game is a simple Bertrand pricing game, and the unique equilibrium wage (with two firms) is

$$w(e) = \mu(e)\theta_h + (1 - \mu(e))\theta_l,$$

and expected profits are zero. Given $\mu(e)$, the worker's effective wage offer is uniquely determined. Given the wage offer, $w(e)$, the worker's optimal choice of education is immediate, $e(\theta)$. Everything is driven by beliefs, $\mu(e)$.

Remark: Note that in the labor-market signaling game, the worker chooses e; but the wage offers are made by the firms. We could have alternatively modeled the game as one in which the worker chooses both e and w, leaving the firms with a simple decision to accept or reject. In this modified game, however, it is possible that firms earn profits in some equilibria (e.g., if there are negative beliefs for some zero-profit wage offers). Such equilibria, however, will not survive basic refinements, such as the Cho-Kreps "Intuitive Criterion" which we apply later; the sensible equilibrium outcomes from both games coincide. In the case of insurance markets, we will face the same modeling choice. The contract space is B (the benefit is the signal) and p (the premium is the price). We could model the game as one where the consumer chooses B and the insurance companies respond with p (that's the analogue of the labor market signaling game here), or we could model the game where the consumer chooses both B and p. When we study the insurance signaling game, we will instead follow the latter approach and allow consumers to offer contract pairs, (B, p), in the first stage.

There are two classes of pure-strategy equilibria in the signaling game. Separating equilibria and pooling equilibria. Hybird semi-separating equilibria may also exist, but for simplicity we ignore them. They do not typically survive standard refinements.

3.1.1 Separating equilibria

A **separating equilibrium** is a PBE where each worker chooses a different education level, $e(\theta_h) = e_h \neq e(\theta_l) = e_l$. Beliefs must satisfy $\mu(e_h) = 1$ and $\mu(e_l) = 0$; and hence $w(e_h) = \theta_h$ and $w(e_l) = \theta_l$. For education other than e_h and e_l, $\mu(e)$ (and therefore $w(\cdot)$) is not pinned down by Bayes' Rule.

The following lemma is very useful:

Lemma 1. *In any separating PBE,* $e_l^* = 0$ *and* $u_l^* = \theta_l$.

Proof: If the worker chooses some $e_l > 0$ in equilibrium, then the worker obtains $w(e_l) = \theta_l$. If the worker deviates and chooses $e = 0$, the worst that can happen is the worker earns $w(0) = \theta_l$. Thus

$$w(0) - c(0, \theta_l) = \theta_l \geq \theta_l - c(e_l, \theta_l),$$

and the deviation is profitable. Given $e_l^* = 0$ (and $w_l^* = \theta_l$ in any separating equilibrium), it follows that $u_l^* = \theta_l$. □

Thus, in any separating equilibrium, the indifference curve of the θ_l worker defines the set of (e, w) pairs that are weakly worse than $(e_l^* = 0, w_l^*) = (0, \theta_l)$. These weakly less-preferred points are candidates for the separating choice of the high-ability worker.

In particular, the choice $(e_h^* = \tilde{e}, w_h^* = \theta_h)$ is a point that is not preferred by the θ_l-type agent. Indeed, $e_h^* = \tilde{e}$ is the lowest level of education along the $w = \theta_h$ locus that is not preferred by the low-type agent. It represents the **least-cost separating equilibrium**.[4] To

[4] To be precise, there are actually an infinite number of equilibria with least-cost education levels, $(e_l, e_h) = (0, \tilde{e})$, all with different belief and wage functions differing over un-important regions. The phrase "least-cost separating equilibrium" really refers to any equilibrium with these least-cost education levels.

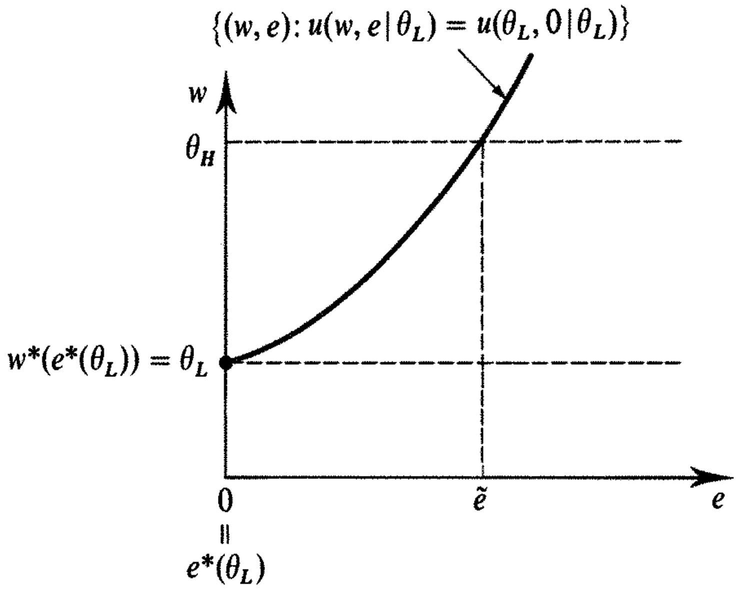

Figure 11: (MWG, Figure 13.C.4)

support this equilibrium, it suffices to choose

$$\mu(e) = \begin{cases} 1 & \text{if } e \geq \tilde{e} \\ 0 & \text{if } e < \tilde{e}. \end{cases}$$

With such beliefs, firms offer

$$w^*(e) = \begin{cases} \theta_h & \text{if } e \geq \tilde{e} \\ \theta_l & \text{if } e < \tilde{e}. \end{cases}$$

Finally, given this wage function, the high type strictly prefers to choose $e_h^* = \tilde{e}$; the low type worker is indifferent between $(0, \theta_l)$ and $(\tilde{e}, \theta_h)$, so it is optimal for her to choose $e_l^* = 0$. Note that there are an infinite number of wage functions that would also implement the same least-cost education levels as an equilibrium outcome.

Other separating equilibria also exist, held together by beliefs which require an education strictly above $\tilde{e}$ in order for $\mu(e) = 1$. What is the range of feasible education levels? It must be that the high-type prefers to separate from the low type rather than obtaining the low-type's allocation, $(0, \theta_l)$. Consider the indifference curves for the θ_l and θ_h workers, both going through $(0, \theta_l)$. The education levels between the θ_l and θ_h indifference curves, along the $w = \theta_h$ wage, represent the set of feasible education signals. In Figure 10, this is the set $[\tilde{e}, e_1]$. The depicted wage function implements $e_h^* = e_1$ as part of a separating equilibrium.

Example: Suppose that $c(e, \theta) = e(k - \theta)$ where $k > \theta_h$. In this case, the least cost-separating education level of θ_h is determined by the θ_l-type's indifference:

$$\theta_l = \theta_h - \tilde{e}(k - \theta_l).$$

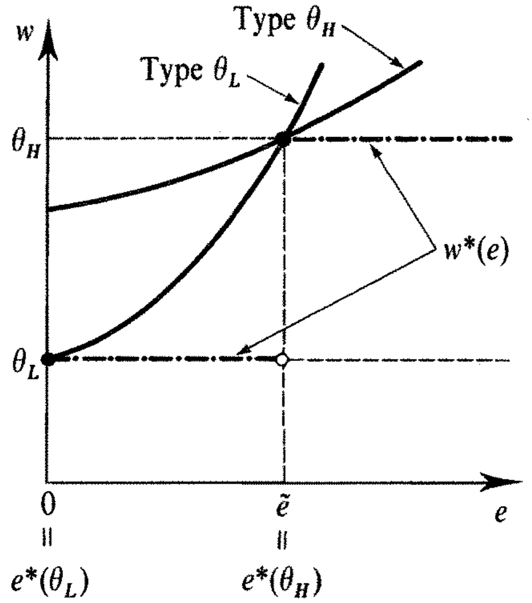

Figure 12: (MWG, Figure 13.C.6)

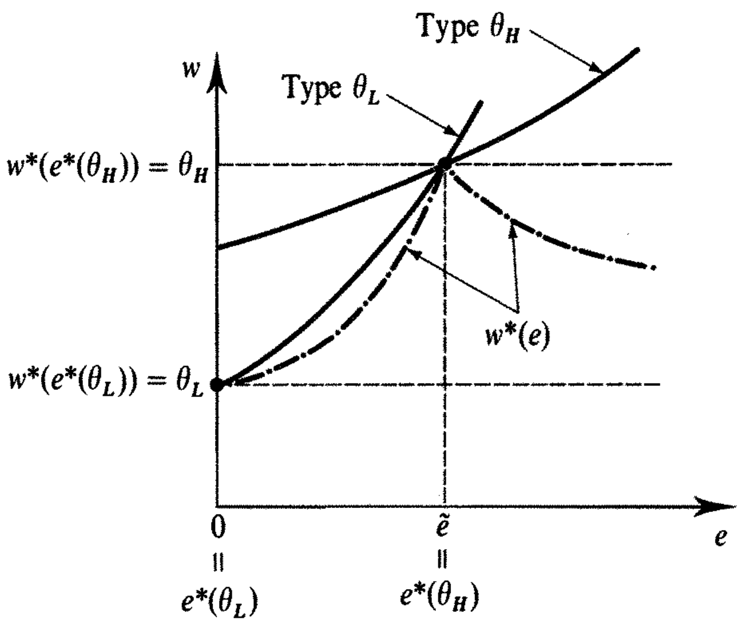

Figure 13: (MWG, Figure 13.C.5)

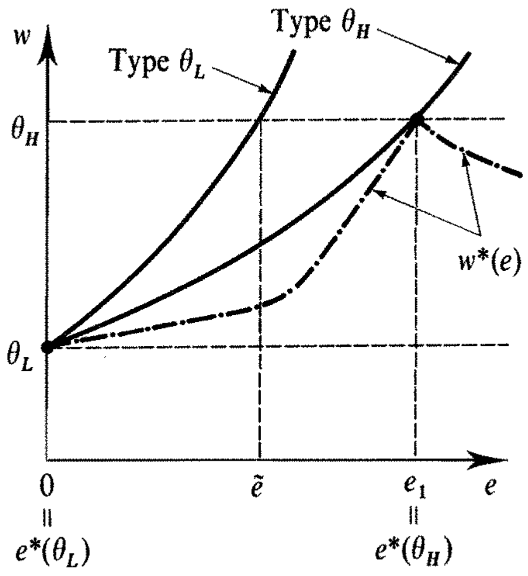

Figure 14: (MWG, Figure 13.C.7)

The highest level of education that can arise in a separating equilibrium is determined by the θ_h-type's indifference:

$$\theta_l = \theta_h - e_1(k - \theta_h).$$

Hence, the set of equilibrium education levels is given by

$$E^*_{sep} = \left\{ (e_l, e_h) \mid e_l = 0,\ e_h \in \left[\frac{\theta_h - \theta_l}{k - \theta_l}, \frac{\theta_h - \theta_l}{k - \theta_h} \right] \right\}.$$

Pareto-optimal separating equilibrium. Note that in all of the separating equilibria, the low type earns $u = \theta_l$ and firms earn zero expected profits. High-type workers are better off in separating equilibria which have lower education requirements. Hence, the Pareto-optimal equilibria among all separating equilibria are those with least-cost separating education levels, $\{(0, \theta_l), (\tilde{e}, \theta_h)\}$. Note that even the least-cost separating equilibrium outcome may not be Pareto dominated by an equilibrium without any signaling. That is, the high-type may prefer to pool with the low type and be paid $w = E[\theta]$ rather than undertake costly signaling (see panel(b) in Figure 15).

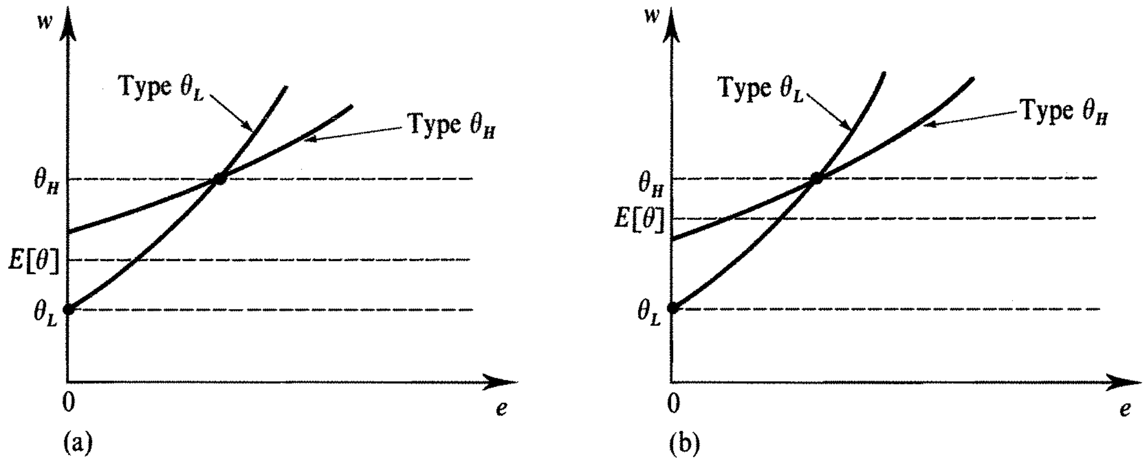

Figure 15: (MWG, Figure 13.C.8)

This is more likely to arise the higher is ϕ – the ex ante probability of θ_h. As ϕ increases toward 1, $E[\theta]$ increases toward θ_h, as in panel (b).

3.1.2 Pooling equilibria

We now turn to pooling equilibria. In all such equilibria, we have $e^*(\theta_h) = e^*(\theta_l) = e_p^*$, $\mu(e_p^*) = \phi$, $w(e_p^*) = E[\theta] = \phi\theta_h + (1-\phi)\theta_l$.

There is one key condition which must be satisfied in any pooling equilibrium: the low-type must be willing to choose e_p^* rather than deviate with $e = 0$ and $w = \theta_l$. Thus, e_p^* must satisfy

$$\theta_l \leq E[\theta] - c(e_p^*, \theta_l).$$

This defines a maximum pooling education level. In the figure below from MWG, it is denoted e'. Any education level $e_p \in [0, e']$ can arise in a pooling equilibrium. To support such an education choice, we can choose

$$\mu(e) = \begin{cases} 0 & \text{if } e < e_p^* \\ \phi & \text{if } e \geq e_p^*. \end{cases}$$

With these beliefs, the labor market will offer $w(e) = \theta_l$ for $e < e_p^*$ and $w(e) = E[\theta]$ for $e \geq e_p^*$. As a consequence, both types of workers will prefer to pool at e_p^* rather than choose $e = 0$.

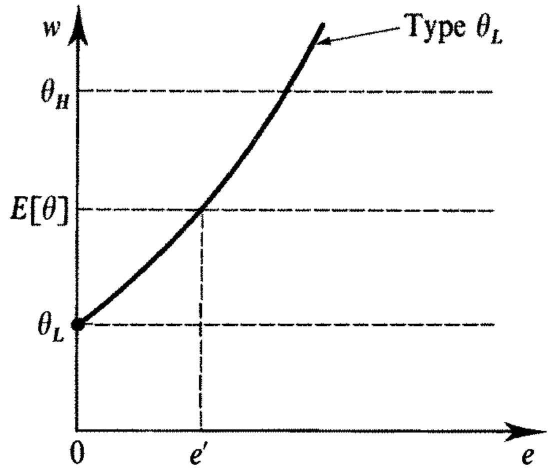

Figure 16: (MWG, Figure 13.C.9)

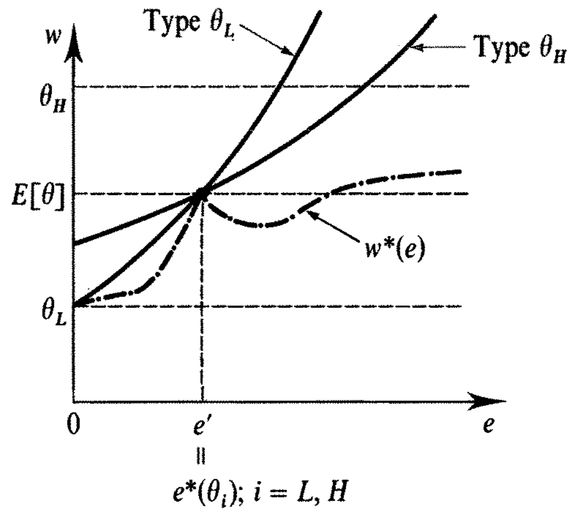

Figure 17: (MWG, Figure 13.C.10)

Example: Return to our previous example in which $c(e, \theta) = e(k - \theta)$ where $k > \theta_h$. In this case, the set of equilibrium education pooling levels is given by

$$E^*_{pool} = \left\{ e \in \left[0, \frac{E[\theta] - \theta_l}{k - \theta_l} \right] \right\}.$$

Pareto-optimal pooling equilibria. The lowest education choice results in the greatest welfare among all pooling equilibria. Thus, $e^*_p = 0$, is the Pareto-optimal pooling equilibrium. This also corresponds to the no-signal outcome.

3.1.3 Multiple equilibria and refinements

There are many, many equilibria in the labor-market signaling game. Are all reasonable?

As a motivating example, consider a pooling equilibrium in which $e_p^* > 0$ and this choice is supported by market beliefs $\mu(e) = \phi$ for all $e \geq e_p^*$. Are these beliefs reasonable? Suppose, for example, a worker chooses a higher education level, $\hat{e}$, which satisfies two properties: (1) the θ_l worker would be worse off making such a choice, even if rewarded with a belief of $\mu = 1$:

$$E[\theta] - c(e_p^*, \theta_l) = u_l^* > \theta_h - c(\hat{e}, \theta_l),$$

and (2) the θ_h type worker prefers such an action if it would lead to the same belief $\mu(\hat{e}) = 1$:

$$E[\theta] - c(e_P^*, \theta_h) = u_h^* \leq \theta_h - c(\hat{e}, \theta_h).$$

In this case, many have argued that the market should have beliefs $\mu(\hat{e}) = 1$. We want to formalize this restriction as the *intuitive criterion* of Cho and (1987).

There are many ways to present this (see for example MWG, 13.Appendix A). Here, I am going to follow Cho and Kreps (1987) (and also Cho and Sobel (1990)), mostly using their notation for signaling games, except I will continue to use θ for type, rather than t. (MWG will use t to represent "tasks" and later we will use t to represent transfer functions, so I want to reserve that notation.)

In the canonical signaling game, there are two players, a sender of information (the informed party), S, and a receiver of information, R (the uninformed party). The timing of the game is (1) nature chooses a type $\theta \in \Theta$ for the sender using the probability function, $\phi(\theta)$, (2) S privately observes θ and sends a message/signal, $m \in \mathcal{M}$ to the receiver, and (3) R observes m and takes a response/action $a \in \mathcal{A}$ and the game ends. Payoff functions for the sender and receiver are represented as $u^S(\theta, m, a)$ and $u^R(\theta, m, a)$. [We can later match up the notation to our labor-market setting, but for now note that in our labor market setting in which education has no value to the firm, the signal m is absent from the payoff function of u^R.] Define the best-response function of the receiver upon observing m with beliefs, $\mu(\cdot|m)$:

$$BR(\mu, m) \equiv \arg\max_{a \in \mathcal{A}} \sum_{\theta \in \Theta} u^R(\theta, m, a)\mu(\theta|m),$$

and define the collection of the receiver's best-responses for every μ that is a distribution on the set $\hat{\Theta} \subseteq \Theta$:

$$BR(\hat{\Theta}, m) \equiv \bigcup_{\mu \in \Delta(\hat{\Theta})} BR(\mu, m).$$

We can now define the Cho-Krep's *Intuitive Criterion*.

> **Cho-Kreps Intuitive Criterion** For each out of equilibrium message m, define the set $\hat{\Theta}(m)$ consisting of those types who would never deviate from the equilibrium:
>
> $$u^*(\theta) > \max_{a \in BR(\Theta, m)} u^S(\theta, m, a) \iff \theta \in \hat{\Theta}(m).$$

If there exists some type $\tilde{\theta}$ such that

$$u^*(\tilde{\theta}) < \min_{a \in BR(\Theta \setminus \hat{\Theta}(m), m)} u^S(\tilde{\theta}, m, a),$$

then the equilibrium does not survive the intuitive criterion.

Remarks:

1. In any extensive form game, there is a sequential equilibrium which satisfies the intuitive criterion. This is proven in Cho Kreps (1987) by showing that Kohlberg and Merten's *stable solution set* is a stronger refinement, and KM have shown that a sequential equilibrium always exists which is in the stable set.

2. Note that if $\hat{\Theta}(m)$ is empty, the intuitive criterion has nothing to say about $\mu(\theta|m)$.

3. Another way to read the intuitive criterion is that "intuitive" beliefs should only put probability weight on $\Theta \setminus \hat{\Theta}$, but be careful. The second equation in the definition, however, states that there will be some type $\tilde{\theta}$ who will prefer to deviate for any beliefs satisfying $\mu(\theta|m) = 0$ for all $\theta \in \hat{\Theta}(m)$. When there are more than two types, this is a slightly weaker test (allows more equilibria to survive) than requiring beliefs in the sequential equilibrium satisfy $\mu(\hat{\Theta}(m)|m) = 0$. The latter test reverses the order of quantifiers: an equilibrium fails if for each belief such that $\mu(\hat{\Theta}(m)|m) = 0$, there exists some type (maybe not the same type) who prefers to deviate.

Returning to our labor-market example, $m = e$ and $a = w$ and $\mathcal{A} = [\theta_l, \theta_h]$. We can reduce all of the firms into a single receiver who is hard-wired to always choose action w to equal the worker's expected marginal product, given beliefs. Now consider any separating equilibrium in Figure 14 (MWG 13.C.7). Recall that any education level between $\tilde{e}$ and e_1 arises in some equilibrium. Suppose that $e_h^* > \tilde{e}$, the least-cost separating education level. Consider the deviation to $e_h = \tilde{e} + \varepsilon$, where $\varepsilon > 0$ and $\tilde{e} + \varepsilon < e_h^*$. Such a deviation lies to the right of the θ_l type's equilibrium indifference curve, and so $\theta_l \in \hat{\Theta}(\tilde{e} + \varepsilon)$. But if e_h leads to beliefs $\mu(\theta_h) = 1$, the θ_h worker does better than in equilibrium. Hence, $\theta_h \notin \Theta(\tilde{e} + \varepsilon)$. Applying the second part of the intuitive criterion requires, therefore, that $\mu(\theta_h|\tilde{e} + \varepsilon) = 1$, and hence $w_h(\tilde{e} + \varepsilon) = \theta_h$. Payoffs to the θ_h worker are

$$\theta_h - c(\tilde{e} + \varepsilon, \theta_h) > \theta_h - c(e_h^*, \theta_h).$$

Thus, the posited equilibrium does not satisfy the Intuitive Criterion. Indeed, the above argument demonstrates that the only separating equilibrium education level that survives is $e_h^* = \tilde{e}$.

The argument that rules out pooling equilibria is similar. Choose any pooling equilibrium and consider the equilibrium indifference curves for the two agents. Because the θ_h indifference curve is flatter than that of θ_l, there will be an interval of education levels along the $w = \theta_h$ line that is to the right of the low-type's curve and to the left of the high-type's curve. The high type can deviate by offering on of these education levels. A similar argument as above shows that the pooling equilibrium cannot survive the Intuitive Criterion. Hence, pooling equilibria do not survive the refinement.

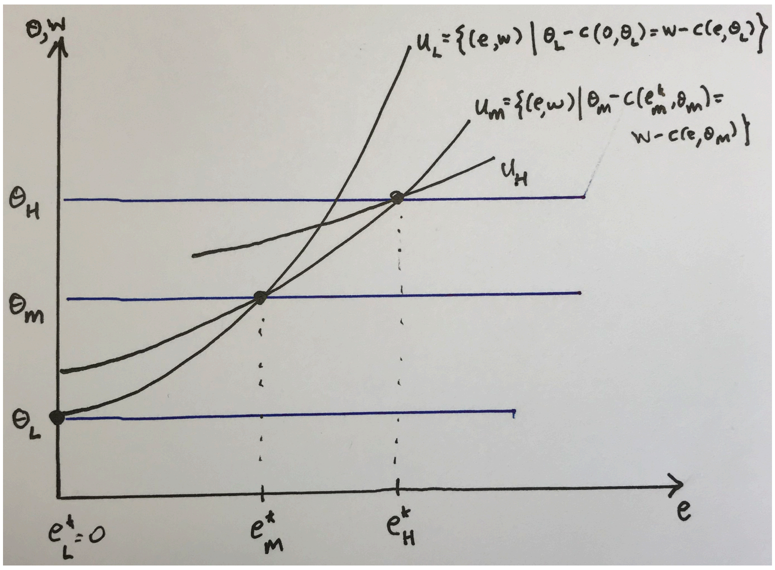

Figure 18: Example of least-cost separating equilibrium with three types, θ_l, θ_m and θ_h. This is sometimes called the *Riley* equilibrium or the *Rothschild-Stiglitz* equilibrium (as we will see when we discuss competitive screening).

There are some reasons not being entirely comfortable with this refinement. First, as the probability ϕ goes to 1, the least-cost signaling equilibrium (which is independent of ϕ) requires that the θ_h types incur costly signaling, even though the pooling outcome is Pareto superior. Second, when $\phi = 1$ and there is no private information over workers, the unique equilibrium is for workers to choose $e = 0$. This is the limit of the pooling equilibrium and not any separating equilibrium.

Multiple (more than 2) types:

In signaling games with two types, the Intuitive Criterion eliminates all hybrid and pooling equilibria, and eliminates all signaling equilibria except the least-cost equilibrium. With more than two types, the Intuitive Criterion has less force. The reason is that when $\Theta \setminus \hat{\Theta}$ contains two or more types, it is much harder to satisfy the requirement that for all beliefs on $\Theta \setminus \hat{\Theta}$, some type $\tilde{\theta}$ will want to deviate. This is because some beliefs may be unfavorable for $\theta' \in \Theta \setminus \hat{\Theta}$ while being favorable for $\theta'' \in \Theta \setminus \hat{\Theta}$, and vice versa.

The natural extension of the Intuitive Criterion is to simply force beliefs to satisfy $\mu(\theta|m) = 0$ whenever beliefs are such that θ is always "less likely" to send m than $\tilde{\theta}$. One formal notion of "less likely" is a variation of *Divinity* (Banks and Sobel (1987) that is called $D1$ in Cho and Kreps (1987). To define it, we need a little more notation. Define $MBR(\mu, m)$ to be the mixed-best-response set for the receiver. This is analogous to $BR(\mu, m) \subseteq MBR(\mu, m)$. Then define the analogous collection of mixed best responses for some belief μ supported by a set $\hat{\Theta}$ as $MBR(\hat{\Theta}, m)$. Lastly, define

$$D(\theta, m) = \{\sigma_R(a) \in MBR(\Theta, m) \mid u^*(\theta) < \sum_{a \in \mathcal{A}} u^S(\theta, m, a)\sigma_R(a)\},$$

$$D^o(\theta, m) = \{\sigma_R(a) \in MBR(\Theta, m) \mid u^*(\theta) = \sum_{a \in \mathcal{A}} u^S(\theta, m, a)\sigma_R(a)\}.$$

We can now state the requirement for $D1$-beliefs:

> **D1 refinement:** Beliefs μ satisfy $D1$ if for any out of equilibrium message m, if there exists a θ and $\tilde{\theta}$ such that
>
> $$D(\theta, m) \cup D^o(\theta, m) \subseteq D(\tilde{\theta}, m),$$
>
> then $\mu(\theta|m) = 0$.

Remarks:

- Because $D1$ is not as strong of a refinement as stability (and a stable sequential equilibrium always exists), a sequential equilibrium satisfying $D1$ always exists.
- A test that is only slightly stronger than the Intuitive criterion (exactly the same if $n = 2$) is the requirement that if $D^0(\theta, m) \cup D(\theta, m) = \emptyset$ for some type θ and there is at least one other type such that $D(\tilde{\theta}, m) \neq \emptyset$, then beliefs should satisfy $\mu(\theta|m) = 0$.
- A weaker test then $D1$ is the requirement of *divinity* or *divine beliefs*. If

  $$D(\theta, m) \cup D^o(\theta, m) \subseteq D(\tilde{\theta}, m),$$

  then

  $$\frac{\mu(\tilde{\theta}|m)}{\mu(\theta|m)} \geq \frac{\mu(\tilde{\theta})}{\mu(\theta)}.$$

  In words, the posterior belief of $\tilde{\theta}$ relative to θ should be at least as great as the prior belief. $D1$ requires the most extreme belief satisfying this condition: $\mu(\theta|m) = 0$.
- There is a sequence of nested, non-empty refinement sets:

  Stable set (non-empty) $\subseteq$ NWBR $\subseteq$ D1 $\subseteq$ Divine
  $\subseteq$ Intuitive Criterion $\subseteq$ Sequential equilibrium=PBE
  $\subseteq$ weak Seq'l Eq/weak PBE $\subseteq$ Nash Eq

- In a large class of signaling games, there is a unique equilibrium satisfying D1 (the least-cost signaling equilibrium). In the least-cost separating equilibrium, the lowest type chooses $e = 0$; the second lowest type chooses an education level that leaves the lowest type indifferent between choosing $e = 0$ and the second-lowest-type's wage and education level; the third lowest type chooses an education level that makes the second lowest type indifferent between her signal and the third-type's signal; etc. The details are in Cho and Sobel (1990) who show that the set of $D1$ equilibria coincides with the set of stable equilibria in the monotone signaling games that we have been studying.

- Another interesting refinement that is not part of the stability-nest is due to Farrell (1985) and Grossman and Perry (1986), sometimes called the Farrell-Grossman-Perry (FGP) equilibrium. It is sufficiently strong that non-pathological games sometimes fails to have FGP equilibria. It is, however, usefully applied to Cheap-talk games which we consider below.

3.2 Insurance-market model (JR)

We now return to our insurance-market setting. As with the labor market setting, we assume now there are just two types, $1 > \pi_h > \pi_l > 0$, and the probability of the low-risk (good) type is $\phi \in (0, 1)$. We assume that there is a single insurance firm.

Insurance signaling game: The timing is similar to the labor-market setting.

- First, nature chooses a consumer's risk.
- Second, the consumer chooses a policy, $\psi = (B, p)$, which is both a level of benefit or coverage, B, and a price or premium, p. [As noted before, this is not the same as the labor-market setting because prices are chosen by the informed side of the market rather than the uninformed side.]
- Third, insurance company (there is just one) forms beliefs that the consumer is low risk, $\mu(\pi_l|\psi) = \mu(\psi)$, and either accepts or rejects the consumer's offer, $\sigma(\psi) \in \{A, R\}$. Note that although there is an insurance monopoly, the consumer has the bargaining power. We could have easily included multiple firms, providing that we insist they only accept or reject (and not offer new terms) and the consumer randomizes over all firms who accept.

Equilibrium

As mentioned above, in equilibrium we require that the players exhibit equilibrium, that beliefs satisfy Bayes' rule wherever possible, and that beliefs are consistent off the equilibrium path as required by PBE and a variation of sequential equilibrium with infinite strategies (see JR, page 387).

To be precise, a pure-strategy PBE (or sequential equilibrium) is a profile $\{\psi_l, \psi_h, \sigma(\cdot), \mu(\cdot)\}$ such that

- given $\sigma(\cdot)$, the choice ψ_l maximizes the expected utility of the low-risk consumer, and ψ_h maximizes the expected utility of the high-risk consumer;
- $\mu(\psi)$ satisfies Bayes' Rule wherever possible;
- for every ψ, $\sigma(\psi)$ maximizes the firm's expected profit given beliefs $\mu(\psi)$.

For ease of notation, define the indirect utilities that arise for the consumers as functions of $\psi = (B, p)$:

$$u_l(B, p) \equiv \pi_l u(y - L + B - p) + (1 - \pi_l)u(y - p),$$

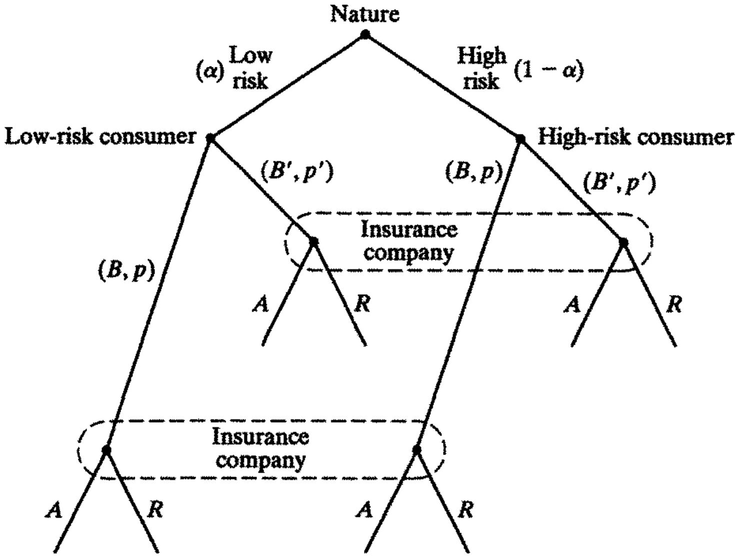

Figure 19: (JR, Figure 8.1)

$$u_h(B,p) \equiv \pi_h u(y-L+B-p) + (1-\pi_h)u(y-p).$$

These functions are continuous, differentiable, strictly concave in (B,p), strictly increasing in B and strictly decreasing in p. Graphing price on the vertical axis, the slope of a consumer's indifference curve (in (B,p) space) is

$$MRS_{B,p} = -\frac{MU_B}{MU_p} = \frac{\pi_i u'(y-L+B-p)}{\pi_i u'(y-L+B-p) + (1-\pi_i)u'(y-p)} \in (0,1).$$

Furthermore, note that $MRS_{B,p}$ is increasing in π_i. Thus, the indifference curves satisfy a single-crossing property and, moreover, high-risk consumers have steeper indifference curves than low-risk consumers.

The firm's preferences are simple. Given $\mu(\psi)$ is the probability of the low risk type, the company will accept any offer such that

$$p > E[\pi|\ \psi] \cdot B = (\mu(\psi)\pi_l + (1-\mu(\psi))\pi_h)B.$$

In a separating equilibrium, of course, $\mu(\psi) = \{0,1\}$ in equilibrium, and so there is value to considering the zero-profit loci of each type. In the figure, ψ_1 is profitable under even the most pessimistic beliefs ($\mu(\psi_1) = 0$); ψ_2 is profitable for $\mu(\psi_2) =$ nut not for $\mu(\psi_2) = 0$; and ψ_3 is unprofitable under even the most optimistic beliefs, $\mu(\psi_3) = 1$.

The following Lemma is a useful fact.

Lemma 2. *Let (u_l^*, u_h^*) be the equilibrium payoffs to the two consumer types. Necessarily,*

$$u_l^* \geq \underline{u}_l = \max_{(p,B),p=\pi_H B} u_l(B,p),$$

$$u_h^* \geq \underline{u}_h = \max_{(p,B),p=\pi_h B} u_h(B,p) = u_h(L, \pi_h L).$$

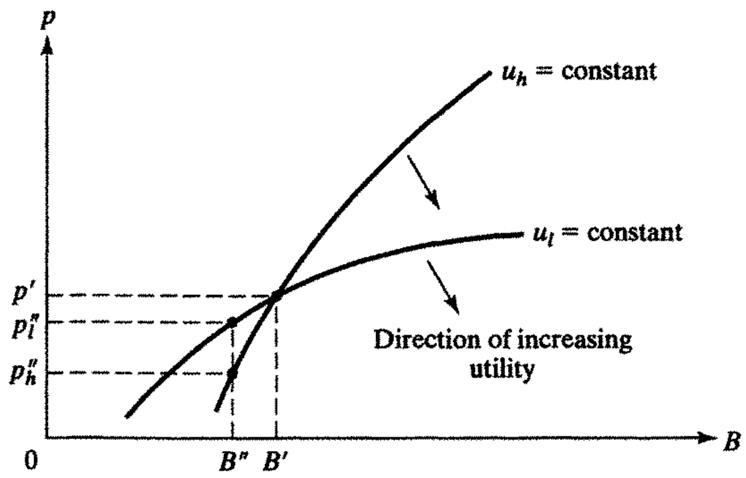

Figure 20: (JR, Figure 8.2)

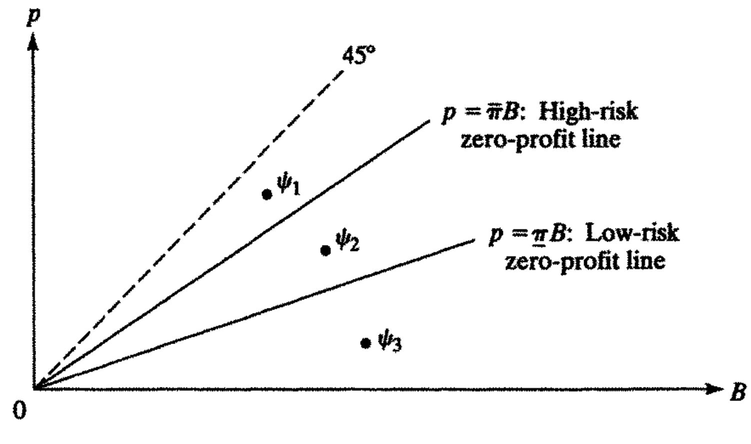

Figure 21: (JR, Figure 8.3)

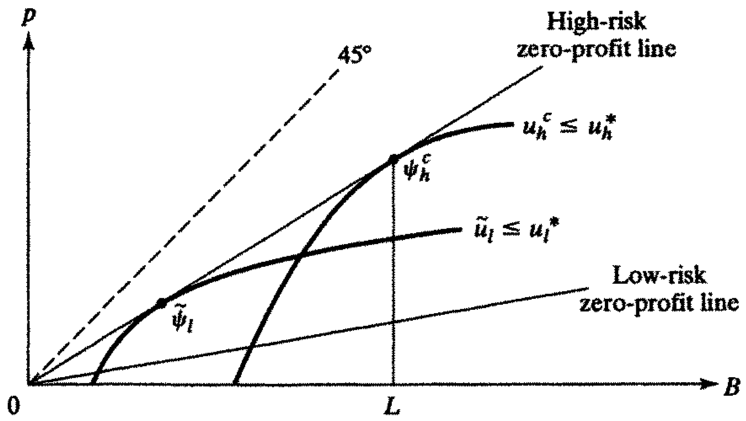

Figure 22: (JR, Figure 8.5)

The proof of the lemma follows either consumer type can always make acceptable offers of (B, p) which satisfy the zero-profit constraint for the firm.

The lemma states that in any pure-strategy equilibrium, the players always have the option to offer an insurance contract that earns zero profit for π_h beliefs. In this case, the high-risk consumer will purchase full insurance (because it is actuarially fair), and thus $u_h^* = u(L, \pi_h L)$, but the low-risk consumer will typically buy less than full insurance given its terms are not fair for type π_l. Indeed, the low-risk consumer may choose to buy no insurance at all at these terms, in which case $u_l^* = u_l(0, 0)$. [Note that we are using slightly different notation than JR(ch8). We are using $\underline{u}_l$ in place of $\tilde{u}_l$ and $\underline{u}_h$ in place of u_h^c.] **Aside:** Looking back at our labor-market setting, the analogue to the above lemma is that each worker earns at least $\underline{u}_l = \underline{u}_h = \theta_l - c(0, \theta_i) = \theta_l$ in any pure-strategy equilibrium.

3.2.1 Separating equilibria

In any separating equilibrium, it is the case that $\psi_l \neq \psi_h$, $\mu(\psi_l) = 1$ and $\mu(\psi_h) = 0$. Furthermore, it must be the case that each consumer type weakly prefers their contract offer to that of the other type, and the firm must expect nonnegative profits. These conditions yield the following lemma.

Proposition 4. *The policies $\psi_l = (B_l, p_l)$ and $\psi_h = (B_h, p_h)$, are accepted in some separating equilibrium iff*

1. *$B_h = L$ and $p_h = \pi_h L$ (high-risk consumer purchases full insurance);*

2. *$p_l \geq \pi_l B_l$ (low-risk policy earns nonnegative profits)*

3. *$u_l(\psi_l) \geq \underline{u}_l$ (from Lemma 2)*

4. *$u_h(L, \pi_h L) \geq u_h(\psi_l)$ (high-risk type does not prefer low-risk contract)*

Proof of Proposition: Necessity: In any separating equilibrium, ψ_h reveals the high-risk consumer. The requirement of nonnegative profit implies that $p_l \geq \pi_h B_h$. But $(L, \pi_h L)$ is the best possible contract for the high-type which satisfies this constraint, and the previous Lemma requires that $u_h^* \geq \underline{u}_h = u_h(L, \pi_h L)$. Hence, (1) must hold. (2) must hold for ψ_l since, by Bayes' rule, $\mu(\psi_l) = 1$ and profits must be nonnegative. (3) follows directly from Lemma 2. (4) is required for the two consumer types to be willing to choose separate offers.

We prove sufficiency by construction. For any $\{\psi_l, \psi_h\}$ which satisfy the conditions of the proposition, define $\mu(\psi) = 1$ if $\psi = \psi_l$ and $\mu(\psi) = 0$ for any other offer. The corresponding optimal strategy for the firm is

$$\sigma(\psi) = \begin{cases} A & \text{if } \psi = \psi_l \text{ or } p \geq \pi_h B \\ R & \text{otherwise.} \end{cases}$$

It is straight forward to see that μ satisfies Bayes' rule and the firm is sequentially rational given μ. We need only check the consumer's choice. Given condition (4), the high-risk consumer does better by choosing ψ_h rather than ψ_l. Any other choice by the high-risk consumer will require $p \geq \pi_h B$ for acceptance, and we already know that $\psi_l = (L, \pi_h L)$ is the best among these. The low-risk consumer will choose ψ_l given (3), because any deviation will result in high-risk beliefs by the firm and thus the requirement that $p \geq \pi_h B$ for acceptance. □

The consequence of the proposition is that all separating equilibria have the construction given in Figure 23 from JR(ch8):

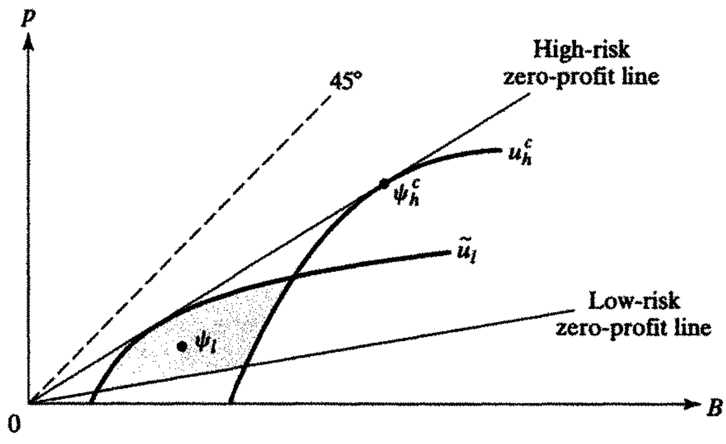

Figure 23: (JR, Figure 8.6)

Remarks of insurance signaling equilibria:

1. Notice that ψ_h is uniquely determined for *every* separating equilibrium, but there is a large set of equilibrium contracts for the low-risk consumer (this is the shaded

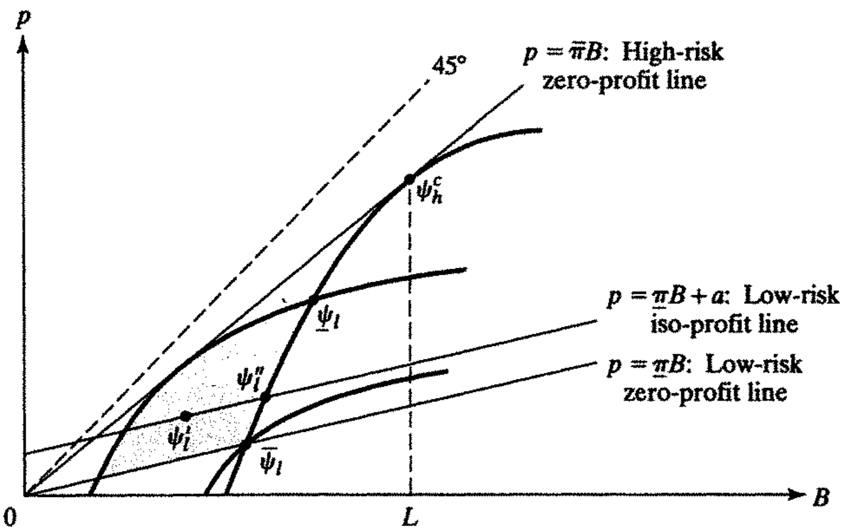

Figure 24: (JR, Figure 8.7)

region in the figure.) This is a more complex setting than MWG's labor market because in that setting, for any separating equilibrium, $w_i = \theta_i$; as a consequence, the set of equilibrium separating contracts was a one-dimensional interval of education levels, E^*_{sep}.

2. Notice that there are equilibria in which the insurance firm makes positive profits from the low-risk consumers. In contrast, expected profits are always zero in the MWG labor-market setting. Why? What is the critical difference between these games?

3. Similar to MWG, if the probability of the good type (low-risk type in this model) goes to zero, the good type would prefer a law that requires all insurance contracts to be efficient (full coverage, in the case of insurance) to prevent signaling. Recall in the labor market signaling game, the high-productivity worker would prefer that $e = 0$ be required in order to eliminate costly signaling.

4. Notice that in all separating equilibria, the high-risk consumer earns the same payoff. The payoffs to the low-risk consumer and the firm, however, vary across the shaded region. That said, only a subset of the shaded region is Pareto efficient. At any point in the shaded region, the firm is indifferent to any contract along the iso-profit line through the point (which has slope π_l). As such, a Pareto improvement can be made by moving northeast along the iso-profit until reaching the high-risk consumer's indifference curve. (Note that the $MRS_{B,p}$ exceeds π_l for the low risk consumer if $B < L$.)

5. The **best separating equilibrium** for the cosumers is the **least-cost signaling equilibrium** ($\overline{\psi}_l$ in Figure 17). Formally, $\overline{\psi}_l = (\overline{B}_l, \overline{p}_l)$ is the solution to

$$u_h(L, \pi_h L) = u_h(\overline{B}, \overline{p}_l),$$

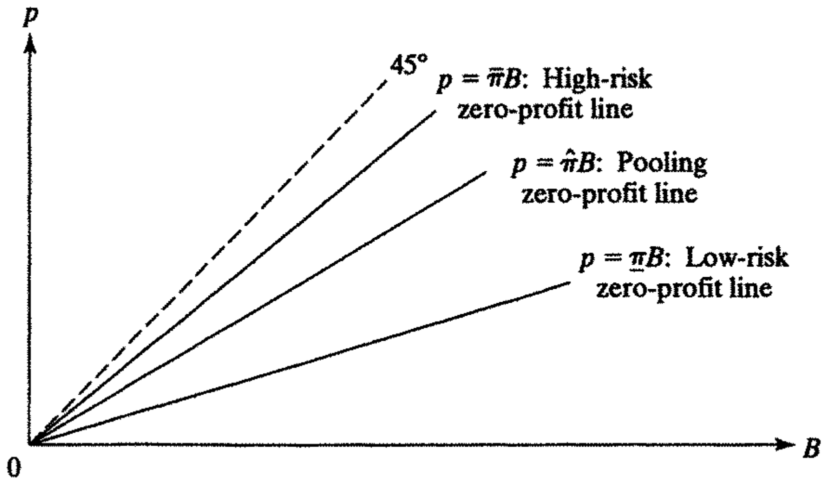

Figure 25: (JR, Figure 8.8)

$$\overline{p}_l = \pi_l \overline{B}_l.$$

3.2.2 Pooling equilibria

In a pooling equilibrium, both consumers offer $\psi_p = \psi_l = \psi_h$ and the firm believes $\mu(\psi_p) = \phi$, the prior probability that the consumer is of low risk. Consequently, the firm will only accept such a policy if it is not unprofitable:

$$p \geq E[\pi]B.$$

Because of the importance of this condition for pooling equilibria, it is useful to introduce a zero-profit line for pooling contracts: In addition to the nonnegative profit condition, recall from Lemma 2 that in any equilibrium we must have $u_h^* \geq \underline{u}_h = u_h(L, \pi_h L)$ and $u_l^* \geq \underline{u}_l = \max_{B \geq 0} \; u_l(B, \pi_h B)$. It turns out that these necessary conditions are also sufficient.

Proposition 5. *If $\psi_p = (\bar{B}, \bar{p})$ is an equilibrium pooling contract iff*

$$\bar{p} \geq E[\pi]\bar{B},$$

$$u_h^* \geq \underline{u}_h = u_h(L, \pi_h L)$$

$$u_l^* \geq \underline{u}_l = \max_{B \geq 0} \; u_l(B, \pi_h B).$$

Proof: Only sufficiency remains to be proven. Construct the following beliefs: $\mu(\psi) = \phi$ if $\psi = \psi_p$ and $\mu(\psi) = 0$ for all $\psi \neq \psi_p$. The given belief satisfies Bayes' rule for the equilibrium choice of ψ_p as required.

With these beliefs, the firm's sequentially optimal strategy is

$$\sigma(\psi) = \begin{cases} A & \text{if } \psi = \psi_p \text{ or } p \geq \pi_h B \\ R & \text{otherwise.} \end{cases}$$

Note that the firm's strategy is to accept whenever the consumer follows the equilibrium strategies (which yield nonnegative profits) or the consumer offers something different that generates nonnegative profits with beliefs $\mu(\psi) = 0$; the firm rejects whenever profits would be negative given these beliefs.

Given the firm's behavior, the consumer can choose ψ_p or any (B, p) such that $p \geq \pi_h B$. The second and third conditions in the proposition imply that the utility from choosing ψ_p, given by u_i^*, weakly exceeds $\max_{B \geq 0} u_i(B, \pi_h B)$ as required. □

The proposition indicates that a large set of ψ_p may arise in any pooling equilibrium. The shaded region in Figure 26 from Jehle and Reny (ch. 8) indicate the set.

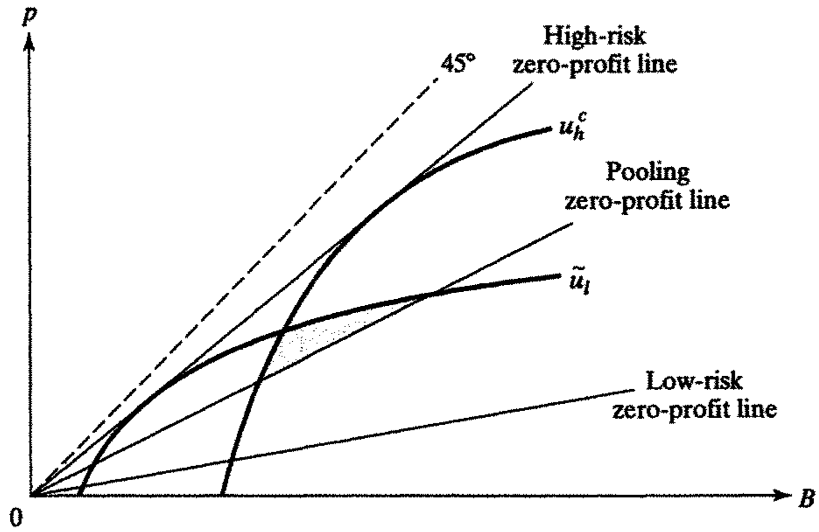

Figure 26: (JR, Figure 8.9)

Remarks on pooling equilibria in insurance markets

1. Note that as ϕ decreases (and a larger proportion of consumers are high risk) the pooling zero-iso-profit line rotates upwards toward the high-risk zero-iso-profit line. As a consequence, there is a minimum ϕ such that pooling cannot arise with a lower proportion of good (low-risk) consumers. Unlike separating equilibria in signaling games, *pooling equilibria may not exist*.

2. As the proportion of low-risk consumers increases and the pooling zero-iso-profit approaches the low-risk zero-iso-profit line, low-risk consumers will prefer the zero-profit pooling equilibrium to the zero-profit separating equilibrium. An example of this can be seen in JR Figure 8.10 (omitted).

3. Notice again that unlike the labor-market signaling game of MWG (ch13), it is possible that firms make positive profit in the JR insurance-market game.

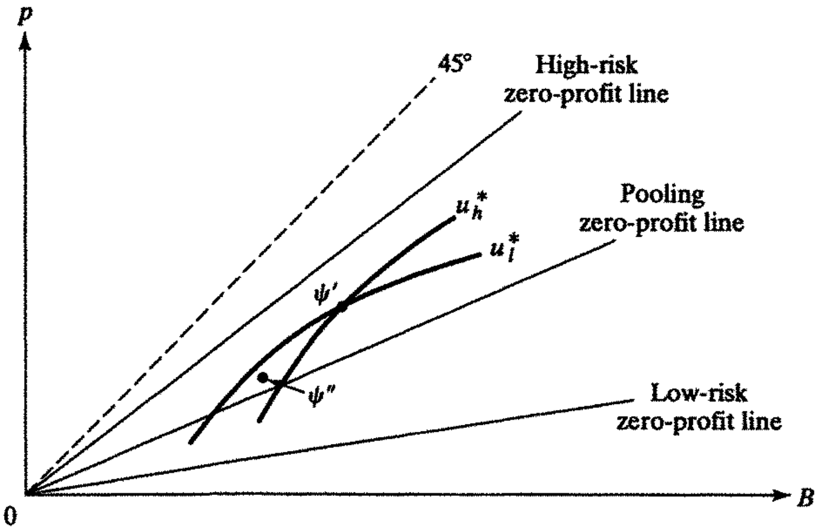

Figure 27: (JR, Figure 8.11)

3.2.3 Multiple equilibria and refinements in the insurance signaling game

As with labor-market signaling equilibria, some insurance signaling equilibria appear to be held together by unreasonable beliefs. Consider the following pooling equilibrium $\psi_p = \psi'$ and consider an out-of-equilibrium offer ψ'' in the figure below. Because $u_l(\psi'') > u_l^*$ and $u_h(\psi'') < u_h^*$, it seems more reasonable to suppose that ψ'' was offered by a low-risk consumer, and thus $\mu(\psi'') = 1$. With these "reasonable" beliefs, the original contract ψ' cannot be a pooling equilibrium.

The reasonable beliefs we are talking about here are, of course, those that survive the intuitive criterion. Let's suppose in what follows that for some out-of-equilibrium $\psi = (B, p)$, we have $u_h(\psi) > 0$ (i.e., the high-risk agent prefers ψ to no insurance) and $p > \pi_l B$ (i.e., ψ is strictly profitable if the consumer's type is known to be π_l). What does the intuitive criterion require when ψ is chosen? There are two conditions to consider. First, what is the best outcome for the high-risk consumer? Because $u_h(\psi) > 0$, the best outcome is that the firm accepts. Suppose that

$$u_h^* > u_h(\psi).$$

In this case, we need to check what is the worst outcome for the low-risk consumer, assuming that the firm believes $\pi = \pi_l$? Because ψ is profitable under those beliefs, the firm will always accept. Thus, the worst possible outcome is $u_l(\psi)$. If

$$u_l(\psi) > u_l^*,$$

then the equilibrium fails the intuitive criterion. We conclude that an equilibrium fails the intuitive criterion if there exists an out-of-equilibrium offer $\psi = (p, B)$ such that (i) $p > \pi_l B$, (ii) $u_h^* > u_h(\psi)$, and (iii) $u_l(\psi) > u_l^*$.

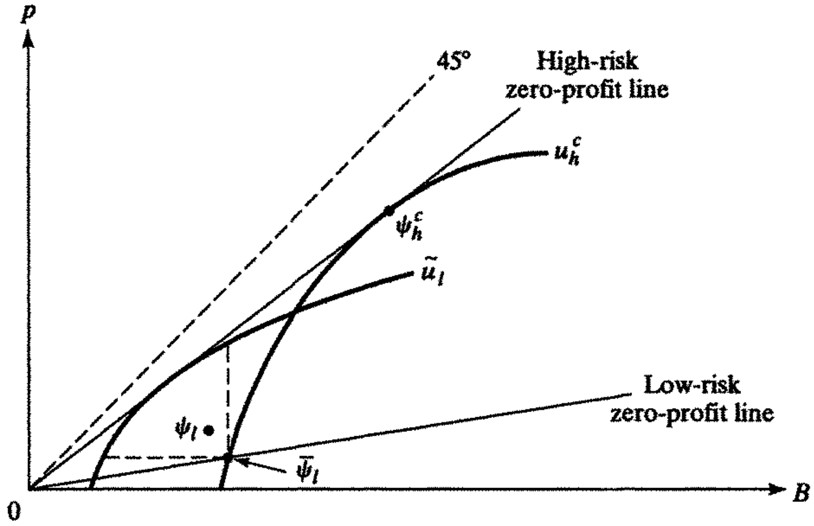

Figure 28: (JR, Figure 8.12)

In any pooling equilibrium, one can find a contract ψ that satisfies these conditions. In the example above, suppose that the low-risk consumer offers ψ'' instead of ψ'. From the point of view of high-risk consumer, the best possible outcome is acceptance and the high-risk consumer earns $u_h(\psi'')$ which is less than u_h^*. Now restricting attention to beliefs that place zero probability on π_h (and thus $\mu(\psi'') = 1$), and providing the firm earns positive profits from ψ'' when $\pi = \pi_l$, the firm will always accept the offer and the low-risk type earns $u_l(\psi'') > u_l^*$. Hence, the original pooling equilibrium offer ψ' cannot arise under the Intuitive Criterion. This argument extends to any pooling equilibrium, so we conclude that pooling equilibria do not survive the intuitive criterion in insurance signaling games with two types.

Consider instead a separating equilibrium as in Figure 21. For any equilibrium outcome ψ_l that is not the least-cost signaling contract, $\overline{\psi}_l$ (i.e., ψ_l is not at the point of intersection between the high-risk indifference curve and the low-risk zero-iso-profit line), the low-risk type could deviate and offer a contract slightly above the low-risk zero-iso-profit line and slightly to the left of the high-risk consumer's indifference curve that achieves a higher utility for the low-risk type. Such a deviation would be profitable for a firm that believed $\pi = \pi_l$; because it is to the left of the π_h-indifference curve, it would be equilibrium dominated for the high-risk consumer ($u(\psi) < u_h^*$); and it would be strictly preferred by the low-risk agent relative to u_l^*. Hence, the original ψ_l choice by the low-risk agent cannot survive the intuitive criterion. The only separating equilibrium which survives is the least-cost equilibrium in which (p_l, B_l) satisfies $u_h(L, \pi_h L) = u_h(B_l, \pi_l B_l)$ and $p_l = \pi_l B_l$.

Another way to look at the intuitive criterion in the context of the insurance signaling game with two types is that beliefs must satisfy

$$\mu(\psi) = \begin{cases} 1 & \text{if } u_h^* > u_h(\psi),\, u_l(\psi) > u_l^*, \text{ and } p > \pi_l B \\ 0 & \text{if } u_l^* > u_l(\psi),\, u_h(\psi) > u_h^*, \text{ and } p > \pi_h B. \end{cases}$$

Remarks:

- Note that the unique equilibrium contracts which satisfy the intuitive criterion is the best separating equilibrium for the low-risk consumer and firms earn zero profits.

- For ϕ close to 1, the low-risk consumer strictly prefers the pooling equilibrium to the separating equilibrium.

- Maskin and Tirole (*Econometrica*, 1992) consider a related signaling game in which the privately-informed type offers a menu of contracts (rather than an education level), the uninformed party forms beliefs about the type of agent who would send such a menu and accepts or rejects the proposal accordingly. When the outside option is the null contract and the informed party has a choice variable which satisfies the single-crossing property, then the high-type worker can always offer the least-cost separating equilibrium allocation (called the Rothschild-Stiglitz-Wilson (RSW) allocation in Maskin and Tirole) and such a menu would be accepted by any firm. Thus, in any equilibrium, the set of outcomes must weakly Pareto dominate the RSW allocation. When ϕ is low (i.e., when the high type prefers the RSW separating equilibrium to the pooling equilibrium), the unique equilibrium outcome is the least-cost separating allocation; no refinements are needed. (MT92, Prop 6) When, however, the pooling equilibrium Pareto dominates the separating equilibrium, then there are multiple equilibria in the Maskin-Tirole game, including the RSW allocation. If one applies the intuitive criterion, however, (D1 is not needed), the unique equilibrium allocation is the least-cost separating (RSW) allocation. (MT92, Prop 7).

4 Cheap-talk games

Our previous analysis of signaling games costly signals (e.g., education in labor markets and incomplete coverage insurance markets). Here, we consider the case where signals are costless (i.e., they have no direct payoff effects for either Sender or Receiver). One can therefore think of the signals as "Cheap Talk". That said, is it possible that the signals/messages may nonetheless signal information?

Let's first be clear about the general structure of Cheap-talk games. First, the timing is unchanged from our standard Sender-Receiver signaling game.

1. Nature chooses Sender type θ_i from Θ according to some prior distribution. The distribution is common knowledge but θ_i is private information to the Sender.

2. The sender chooses some message $m \in \mathcal{M}$, the set of feasible messages. A strategy will be a conditional probability distribution over messages, $\sigma^S(m|\theta)$.

3. Receiver observes m (forms a belief $\mu(\theta_i|m)$) and chooses an action $a \in \mathcal{A}$; a strategy is a conditional probability distribution over actions, $\sigma^R(a|m)$.

The only difference with the standard Sender-Receiver signaling game is that talk is cheap, so payoffs are independent of messages/signals. Thus we can write payoffs as $u^S(\theta, a)$ and $u^R(\theta, a)$. That said, a message may nonetheless be informative in some equilibria.

What are the necessary conditions for meaningful communication to emerge in an equilibrium? (1) different sender types must have different preferences over Receiver actions (otherwise all Senders would send "most favorable" message); (2) the Receiver must prefer different actions depending upon θ_i (otherwise communication is irrelevant to the Receiver); (3) Receiver's preferences over actions must not be entirely opposed to the Sender's preferences (i.e., there must be some commonality in preferences).

To illustrate (3) in a simple setting, consider two variations of the same game. In the first, Alice wishes to hire Bob. Bob's is either"sharp" or "dull" with equal probability. If he is sharp, then Alice prefers to assign him to a "hard" task and the payoffs are (2,2) for Alice and Bob; if Bob is dull, Alice prefers to assign him to the "easy" task and payoffs are (1,1). For the other task assignments, either Bob is unable to do it (dull Bob and hard task) or Bob is bored (smart Bob with easy task). In either case, payoffs are (0,0).

Now imagine that Bob's type is unknown to Alice, but he can make a speech to Alice before she assigns him to a task. There are two equilibrium outcomes.

- First, Alice can think that Bob is babbling (that his speech has no information) and ignore it. In this case she assigns the hard task and gets an expected payoff of 1. Because she ignores Bob, it is an equilibrium for Bob to babble in his speech. *Notice that the "babbling equilibrium" will be an equilibrium in any cheap talk game*, not just this game.

- Bob can instead say truthfully "I am smart" or "I am dull"; Alice responds appropriately with the optimal task assignment. This is also an equilibrium. Note, however, that the messages could be anything with meaning. In equilibrium Bob could say "turnip" if he is smart and "carrot" if he is dull, providing that Alice understood the meaning of "turnip" and "carrot." So in these games, we will focus on equilibria where messages have their commonly understood meaning.

Now change the game so that Bob always prefers the hard task because it comes with a higher wage (embedded in the payoffs). Specifically, change the previous outcome for "dull Bob with hard task" from (0,0) to (1.5,-1.5), where 1.5 is Bob's payoff. Notice that now Bob will always prefer the "hard" task assignment, so if there is a message that makes it more likely Bob gets assigned the hard task, both types of Bob will send it. For this reason, the messages cannot credibly convey information about type in equilibrium. Given messages have no content, Alice chooses the "dull" task assignment regardless of Bob's type.

Thus, for meaningful information transmission to arise in an equilibrium, the preferences of the sender and receiver must be partially aligned. Crawford and Sobel (1982) consider the first formal model of cheap talk in which meaningful communication arises in some equilibria. There primary assumptions about preferences are

- $\theta \in [0, 1]$ (without loss of generality)
- $u^S(\theta, a)$ and $u^R(\theta, a)$ are twice continuously differentiable and strictly concave in a, where $a \in \mathbb{R}$.
- Define the ideal actions for each player given θ as

$$a^S(\theta) = \max a \in \mathcal{A} u^S(\theta, a),$$

$$a^R(\theta) = \max a \in \mathcal{A} u^S(\theta, a).$$

  Assume that $a^S(\theta)$ and $a^R(\theta)$ are strictly increasing in θ (i.e., $U^k_{\theta a} > 0$) and that $a^S(\theta) > a^R(\theta)$ for all θ (i.e., the Sender is biased toward higher actions than the Receiver prefers).

Strategies

Crawford and Sobel (1982) show in Lemma 1 that in any equilibrium, the set of equilibrium actions is finite which implies that the equilibrium must involve noisy signals (i.e., messages cannot be perfectly invertible). Using this fact, Crawford and Sobel then establish (Theorem 1) that any equilibrium outcome (i.e., a mapping from θ to a) can arise as part of a partition equilibrium with the following structure:

- the set of messages is $\mathcal{M} = [0, 1]$;
- there exists an n-type partition: $\{[0, x_1), [x_1, x_2), \ldots, [x_{n-1}, 1]\}$;
- the Sender of type θ randomly chooses a message m from the interval in the partition containing his type;
- the Receiver's beliefs $\mu(\theta|m)$ are the prior beliefs conditioned on the partition of the message; the Receiver's action (pure-strategy) is sequentially rational given this belief:

$$\overline{a}^R([x_i, x_{i+1}]) \equiv \max_{a \in \mathbb{R}} \int_{[} a_i, a_{i+1}] u^R(\theta, a) dF(\theta), \quad \text{is chosen for } m \in [x_i, x_{i+1}];$$

- the Sender type $\theta = x_i$ is indifferent between adjacent actions at x_i:

$$u^S(x_i, \overline{a}^R([x_{i-1}, x_i])) = u^S(x_i, \overline{a}^R([x_i, x_{i+1}])).$$

4.1 A simple cheap-talk game with quadratic preferences

Here is a simple uniform-quadratic game of cheap talk which illustrates what can happen.

- θ is uniformly distributed on $[0, 1]$,
- Receiver preferences are quadratic
$$u^R(\theta, a) = -(a - \theta)^2$$
and therefore $a^R(\theta) = \theta$;
- Sender's preferences are also quadratic but with bias term $b > 0$:
$$u^S(\theta, a) = -(a - (\theta + b))^2,$$
and therefore $a^S(\theta) = \theta + b$.

Think of b as the bias between the send and receiver. If $b = 0$, then their preferences are perfectly aligned. For $b > 0$, however, the sender always prefers a higher action (but not too much higher action) than the receiver.

We want to characterize the equilibrium partition and verify that a given partition will generate the predicted play of the game. Let's begin by trying to find an equilibrium with 2 steps, $[0, x_1)$ and $[x_1, 1]$. Let m_1 denote a generic message from $[0, x1)$ and m_2 a generic message from $[x_1, 1]$. In an equilibrium, it must be that the receiver forms beliefs

$$\mu(\theta|m_1) = \begin{cases} \frac{1}{x_1} & \text{if } \theta \in [0, x_1) \\ 0 & \text{if } \theta \notin [0, x_1) \end{cases}$$

$$\mu(\theta|m_2) = \begin{cases} \frac{1}{1-x_1} & \text{if } \theta \in [x_1, 1] \\ 0 & \text{if } \theta \notin [x_1, 1]. \end{cases}$$

Consistent with these beliefs, the Receiver optimally chooses

$$\overline{a}^R(m_1) = \frac{x_1}{2},$$

$$\overline{a}^R(m_2) = \frac{x_1 + 1}{2}.$$

Now consider the Sender's incentives for sending messages. Any message $m_1 \in [0, x_1)$ will generate $a_1 = \frac{x_1}{2}$ and any message $m_2 \in [x_1, 1]$ will induce $a_2 = \frac{x_1+1}{2}$. In general, the θ Sender will prefer a_1 to a_2 if the midpoint between these actions is greater than $\theta + b$, and will prefer a_2 to a_1 otherwise. For both messages to be sent in equilibrium, therefore, there must exist a type who is exactly indifferent between these two outcomes (assuming such a type exists for now):

$$u^S(\hat{\theta}, a_1) = u^S(\hat{\theta}, a_2),$$

or using quadratic preferences

$$\hat{\theta} + b = \frac{1}{2}\left[\frac{x_1}{2} + \frac{x_1 + 1}{2}\right]$$

Thus, $\hat{\theta} = \frac{1}{2} - 2b$. All types below $\hat{\theta}$ send m_1 and all types above send m_2, and thus in equilibrium x_1 must equal $\hat{\theta} = \frac{1}{2} - 2b$.

Note, however, that if $b \geq \frac{1}{4}$, the required x_1 does not lie in $[0, 1]$, in which case a two-partition equilibrium cannot exist. Indeed, using our notation above, $n(b) = 1$ for all $b \geq \frac{1}{4}$. If $b < \frac{1}{4}$, then we have established the Sender will follow the equilibrium strategy of randomly sending $m_i \in \Theta_i$ if $\theta \in \Theta_1$.

What about n-step equilibria where $n > 2$? Note that the upper step in the 2-step equilibrium partition is $4b$ longer than the lower step. This is entirely due the indifference requirement of the type at x_1. More generally, for any type at x_i who is indifferent between message m_i and m_{i+1}, it must be that

$$x_i + b = \frac{1}{2}\left[\frac{x_i - x_{i-1}}{2} + \frac{x_i + x_{i+1}}{2}\right].$$

Simplifying this difference equation, we obtain

$$(x_{i+1} - x_i) = (x_i - x_{i-1}) + 4b.$$

Hence, in any n-step equilibrium, each successive interval must grow in length by $4b$. Note that the first step, $[0, x_1)$ has length x_1 and all of the steps must increase by $4b$. Thus, we have the second requirement that in any n-step equilibrium,

$$x_1 + (x_1 + 4b) + \cdots + (x_1 + (n-1)4b) = 1.$$

Using the fact that $1 + 2 + \cdots + (n-1) = n(n-1)/2$, we have

$$nx_1 + n(n-1)2b = 1.$$

Given $n(n-1)2b < 1$, there exists a value of $x_1 \in (0, 1)$ that solves this equation. The inequality is satisfied if n is an integer less than

$$\overline{n} = \frac{1}{2}\left(1 + \sqrt{1 + \frac{2}{b}}\right).$$

Hence, $n(b)$ is the greatest integer less than $\overline{n}$.

Remarks:

1. Note that as b approaches zero, $n(b)$ approaches infinity. But for any given $b > 0$, $n(b)$ is finite. Less bias means more informative equilibria can arise.

2. Welfare. From an ex ante point of view (before θ is learned), both Sender and Receiver prefer to play the most informative equilibrium. This is a result in Crawford and Sobel (1982) and is not immediate. That said, once θ is known by the sender, the Sender may prefer a different equilibrium so we need to be careful about using ex ante Pareto optimality as a refinement.

3. Our previous refinements (of t)the intuitive criterion and D1) have no power in reducing the set of equilibria because talk is cheap. Farrell (1985) and Grossman-Perry (1986) developed a refinement called *Neologism-proof equlibria* (Farrell's name) *Perfect Sequential Equilibrium* (Grossman and Perry's name). Equilibria satisfying this refinement don't always exist for dynamic games of incomplete information. In the case of the quadratic cheap-talk model of Crawford and Sobel, however, the unique equilibrium that satisfies the Farrell-Grossman-Perry refinement is the most informative equilibrium with $n(b)$ partitions. In more recent work by Chen, Kartik and Sobel (*Ema*, 2008), a simpler refinement of NITS ("no incentive to separate") is developed and applied to eliminate all but the most informative equilibria.

4. Cheap-talk models have been applied to many applied problems in economics, political science, finance and organizational design. Here are a few examples:

   - Dessein (*REStud.*, 2002). Suppose the receiver is the owner of a firm (and de facto, controls the decision a), while the sender is the manager who learns information. Suppose $b = \frac{1}{12}$ so that $n(b) = 3$. Then direct computation yields an owner payoff of $-\frac{1}{36}$ in the most informative equilibrium. Suppose instead that the owner decides to delegate the decision to the manager and allow the manager to choose any $a \in [0, 1]$. In this case, $a = \theta + b$ and the owner's payoff is $-b^2 = -\frac{1}{144}$ so delegation is better. More generally, you can imagine the owner offers the manager a delegation set $\mathcal{A}_d \subsetneq [0, 1]$ with even better outcomes.
   - Farrell and Gibbons (*JET*, 1989) consider the possibility that there is a round of cheap talk before the Chatterjee-Samuelson bargaining game. Cheap talk can matter and some types of buyers and sellers prefer an informative 2-step round of Cheap talk followed by the CS game, to simply playing the CS game.
   - Quint and Hendricks (*AEJmicro*, 2018). A recent auction application in the spirit of Farrell-Gibbons (1989). Sometimes sellers of complex assets contact numerous bidders and ask each for a preliminary indication of interest. Potential bidders respond with cheap talk messages that are crude (e.g., "I think I would bid in this range"), but still informative. The seller selects a subset of the highest indicative bids (the maximum number selected is set beforehand) and the selected bidders are asked to incur due-diligence costs (e.g., reading and analyzing numerous legal and accounting documents related to the asset value), after which they obtain a precise indication of their value of the asset. Then a real auction with real bids occurs. Because the buyers and the seller have common interests to avoid unnecessary costs in preparing the real bids, cheap talk can be informative in this setting.

5 "Signal-jamming" (applied to Career concerns)

So far we have considered three kinds of information-revelation games in which the informed player moves first and communicates information to the uninformed seller through her actions: certifiable-disclosure games, Spence-style signaling games, and cheap-talk games. We now turn to a fourth class of games that is conceptually different from the other three: signal-jamming games. The name "signal-jamming" was coined by Fudenberg and Tirole (1986) in a paper about predation. The general idea is that the market generates observations that are noisy signals about the state of the world and a player may attempt to "jam these signals" so that another player would misinterpret the observations. In the case of Fudenberg and Tirole's model, an incumbent firm secretly prices low in an attempt to fool a recent entrant into thinking the market is unprofitable and so to induce exit. The entrant is not fooled in equilibrium, but the incumbent still has an incentive to secretly price low in equilibrium which makes entry less profitable. More generally, signal jamming models involve one player choosing actions to manipulate the inferences of another. Unlike the other signaling models we have considered, here it is not necessary that the player who jams the signal has more information.

This will all make more sense with a worked out example. The example we will pursue here is the career-concerns model in Holmström (1982, reprinted in *RESud.*, 1999).

- There is a risk-neutral worker and a pair of risk neutral firms, and payoffs are assessed using a common discount factor, δ.

- In each period, the worker chooses an unobserved action, a_t, which influences output, which is observed. Output is

  $$q_t = \theta + a_t + \varepsilon_t,$$

  where θ is a private productivity parameter for the worker, chosen by nature from a normal distribution, $N(\mu_\theta, \sigma_\theta^2)$ and ε_t is output noise realized in period t, and where ε_t is normally distributed $\mathcal{N}(0, \sigma_\varepsilon^2)$. Effort has disutility $\psi(a)$ which is strictly increasing, strictly convex and $\psi(0) = \psi'(0) = 0$. The worker's payoff in period t is

  $$w_t - \psi(a_t).$$

- Firms cannot offer incentive contracts nor can they offer long-term contracts. They can only offer the worker a wage w_t at the start of the period in exchange for the output q_t at the end of the period.

- An employing firm's payoff in period t is

  $$q_t - w_t.$$

  We assume Bertrand competition between the firms so that each firm offers the worker her expected output conditional on her history of output and assumptions regarding equilibrium play:

  $$w_t = E[q_t | q_1, \ldots, q_{t-1}].$$

- We assume that neither the worker nor the firms know θ; instead, the firms will make inferences about θ based on past output which will give rise to the worker's *career concerns*.

Suppose that there are T periods, $t = 1, \ldots, T$. At the start of the game, Nature chooses θ for the worker, but neither the firms nor the worker know anything other than the prior mean μ_θ and variance σ_θ^2 of the normal distribution.

At the start of each period t, firms and the worker have observed the previous outputs of the worker, $q^{t-1} = (q_1, \ldots, q_{t-1})$. The firms use this information in tandem with knowledge about the equilibrium strategy of the worker, $(a_t^*(q^{t-1}))_t$, to compute $E[\theta|q^{t-1}]$. The firms will offer the worker the expected output in period t:

$$w_t = a_t^*(q^{t-1}) + E[\theta|q^{t-1}].$$

The worker accepts this wage and then chooses effort a_t. Finally, output is determined by $q_t = \theta + a_t + \varepsilon_t$, and the game continues to the next period if $t < T$ and concludes otherwise.

Let's start with the $T = 2$ version of this game. At $t = 2$, the worker has no incentive to choose any effort, so $a_2^* = 0$, regardless of period 1 output. As a consequence,

$$w_2 = E[\theta|q_1].$$

We'll come back to this conditional expectation in a moment.

Now consider period $t = 1$; there is no previous output history, so a_1^* is chosen independent of output history. In equilibrium, the firms know this number, and so they will offer

$$w_1^* = E[\theta] + a_1^* = \mu_\theta + a_1^*.$$

Now let's return to $E[\theta|q_1]$. Because the firms observe q_1, they can compute

$$q_1 - a_1^* = \theta + \varepsilon_1.$$

This is an unbiased estimate of θ. But they also have the prior estimate of $\theta = \mu_\theta$ that is also unbiased. The best estimate is found by combining these two signals to form a posterior of minimum variance. The formula is very simple:

$$E[\theta|q_1] = \lambda\mu_\theta + (1 - \lambda)(q_1 - a_1^*),$$

where

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\theta^2 + \sigma_\varepsilon^2} \in (0, 1).$$

What remains to be determined is a_1^*. Consider the worker's choice of a_1, knowing the inference that the market will subsequently make.

$$w_2 = \lambda\mu_\theta + (1 - \lambda)(q_1 - a_1^*)$$

and so

$$w_2 = \lambda\mu_\theta + (1 - \lambda)(\theta + a_1 + \varepsilon_1 - a_1^*).$$

The worker's present-value payoff is therefore

$$w_1^* - \psi(a_1) + \delta\left(\lambda\mu_\theta + (1 - \lambda)(\theta + a_1 + \varepsilon_1 - a_1^*)\right).$$

First-order conditions for a_1 yield

$$\psi'(a_1) = \delta(1-\lambda) = \delta\left(\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\varepsilon^2}\right) \in (0,1).$$

Thus, first-period effort is positive (due to career concerns), but in the 2-period model still lower than the first-best effort given by $\psi'(a^{fb}) = 1$.

Key: Why does the agent work in period 1? Because increasing effort on the margin increases the market's inference that the worker is more productive, which in turn leads to higher wages in the future. In equilibrium, the firms are not misled, and the worker chooses an effort level so that the marginal benefit of increasing future wages just equals the marginal cost of effort.

T**-period model:** In the T-period model, the market inference of θ turns out to be straightforward. It is easier to work with normal *precisions* rather than variances. Define the precision of a random variable, h, to be the reciprocal of its variance. Thus, high precision translates into low variance. Note that prior to period 1, the firms' best estimate of θ, is $z_0 = \mu_\theta$ which has precision $h_\theta = \frac{1}{\sigma_\theta^2}$. In each subsequent period, the firm learns the additional signal

$$z_t = q_t - a_t^*(q^{t-1}) = \theta + \varepsilon_t,$$

which is distributed normally with mean θ and precision $h_\varepsilon = \frac{1}{\sigma_\varepsilon^2}$. The best estimate of θ after observing q^{t-1} for $t > 1$ is easily determined to be

$$E[\theta|z_0, z_1, \ldots, z_t] = \frac{h_\theta\mu_0 + h_\varepsilon\sum_{s=1}^{t-1} z_s}{h_\theta + th_\varepsilon},$$

and the precision of this posterior is $h_t = h_{t-1} + h_\varepsilon$.

When the agent chooses q_t, the agent will consider the present cost of the effort, and the future gain in the improvement of the wage path. Note the marginal effect of an increase in a_t on the marginal wage is constant and depends only on the discount factor, the precisions and time. The marginal benefit is independent of q^t, and thus a_t^* is independent of previous outputs! This allows us to solve for each periods a_t by looking the first-order conditions period-by-period.

It is perhaps easiest to consider the $T = \infty$ case. It is a rather messy computation (I'll leave you to consult the original paper by Holmström (1999)), but the marginal return from effort a_t can be shown to equal a marginal increase in wage w_s equal to h_ε/h_s. Thus, the total present-discounted return to effort a_t is

$$\gamma_t \equiv \sum_{s=t}^{\infty} \delta^{s-t}\frac{h_\varepsilon}{h_s}.$$

Hence, $\psi'(a_t^*) = \gamma_t$. Because precisions increase linearly in time, γ_t will decrease over time. This reflects the fact that the market has almost learned the worker's type and thus additional manipulation has little impact. Hence, we should see higher efforts earlier in the relationship (indeed it can be $a_1^* > a^{fb}$ with multiple periods) than near the end.

Further Remarks:

- With large T, the firms learn the worker's type in the long run. Holmstrom (1999) offers an extension in which θ follows a random walk, $\theta_t = \theta_{t-1} + \eta_t$, with η_t distributed normally with mean 0 and precision h_η. One can apply the same toolkit as before and consider steady-state paths. Holmström shows that f $\delta = 1$, the steady state effort level is equal to the first best, a^{fb}. Otherwise, steady-state effort is below the first-best level. The steady-state level of effort is also closer to the first best as σ_η^2 is higher or σ_ε^2 is lower. Indeed, if $\varepsilon_t \equiv 0$, then steady-state effort is again first-best efficient.

- Holmstrom (1999) provides a somewhat different model of career concerns with project selection in the last part of the paper. There, he supposes that there are some projects that, if chosen by the manager, would reveal information about their type. If managers are even slightly risk neutral, then their incentives will be marginally distorted away from positive NPV projects because any project will increase the dispersion of the market's posterior, and thereby create wage risk for the manager. As a further extension, if the worker observes some information about the project that the firm does not, a lemons problem may develop leading to no projects being chosen in equilibrium.

- Many extensions have been made to Holmström's career concern model. Take a look at the chapter in Bolton and Dewatripont, especially the two papers by Dewatripont, Jewitt and Tirole (*RESTud.*, 1999) that are referenced in BD.

6 Screening in competitive markets

We now return to our labor-market and insurance market environments but switch the timing, letting the uninformed parties move first and make offers to the informed side of the market. Such endogenous offers will (sometimes) separate or "screen" out the different types, and hence we refer to them as *screening contracts*.

6.1 Labor-market model (MWG)

For simplicity, we assume that there are just two types, $\theta_h > \theta_l$ and the ex ante probability of the high type is ϕ. We will also continue to assume that reservation values are type independent and set $r(\theta_l) = r(\theta_h) = 0$. To maintain our symmetry with the case of signaling, we assume that there is a screening variable t, a task, that can be required of the worker, and this task is unproductive (just as education was in the signaling game). The type-θ worker's payoff from accepting a wage w and a task, t, is

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

where c has the same properties as our eduction cost function used before (i.e., c is increasing and convex in t, decreasing in θ, $c(0, \theta) = 0$, and c satisfies single-crossing in (t, θ)).

Screening Game

1. Nature chooses the worker's type, θ with probability ϕ of θ_h and $(1 - \phi)$ of θ_l.

2. Firms (we'll assume there are just two) simultaneously offer a menus of contracts; a contract is a pair (w, t), and a menu is a collection of contracts (perhaps just one);

3. Given the firms' offers ($n = 2$ for simplicity), workers choose whether to accept a contract and, if so, which contract and firm. (We assume that workers resolve indifferences toward lower tasks and toward employment; workers randomize if more than one firm offers the worker's preferred contract.

Remark: Note that in the last stage of the game, the informed player (the worker) moves. Because the information set is a singleton at this decision node, it represents a proper subgame. As such, weak Perfect Bayesian equilibrium and SPNE (and sequential equilibrium, properly extended) are all equivalent.

Full-information outcome

As a benchmark, suppose that the worker's type is observable but otherwise the timing is the same as above. What happens?

Result: In any subgame perfect Nash equilibrium of the complete information game, type θ_i worker is offered and accepts contract $(w_i^*, t_i^*) = (\theta_i, 0)$.

The proof of this largely follows the standard logic of Bertrand games:
(1). Any individual contract must generate zero expected profits. If this were not the case, then $w_i^* > \theta_i$ for some firm and that firm must be earning negative profits; in this case the firm could drop the unprofitable contract. Suppose instead that some firm offers $w_i^* < \theta_i$, in which case positive profits are being made by at least one of the two firms. Let the total industry profit for type θ_i be Π. In this case, the weakly lower profitable firm could offer $(w_i^* + \varepsilon, t_i^*)$ and obtain $\Pi - \varepsilon > \Pi/2$ for ε small. Thus, $w_i^* = \theta_i$ for $i = l, h$.
(2). $t_i^* = 0$ for $i = l, h$. Suppose otherwise that $t_i^* > 0$. In this case, task allocation is inefficient and either firm could deviate by offerings $t_i^* - \varepsilon$ and adjusting the wage downward so as to capture almost all of the efficiency gains (leaving a small gain to he worker so that they are strictly willing to deviate). Only at $t_i^* = 0$ is such a deviation impossible. $\square$

Remark: Note that the complete-information equilibrium offers, $\{(\theta_h, 0), (\theta_l, 0)\}$ cannot be part of the equilibrium to an incomplete-information game because both types of workers would choose $(\theta_h, 0)$ and the firms would lose money on the θ_l-type workers.

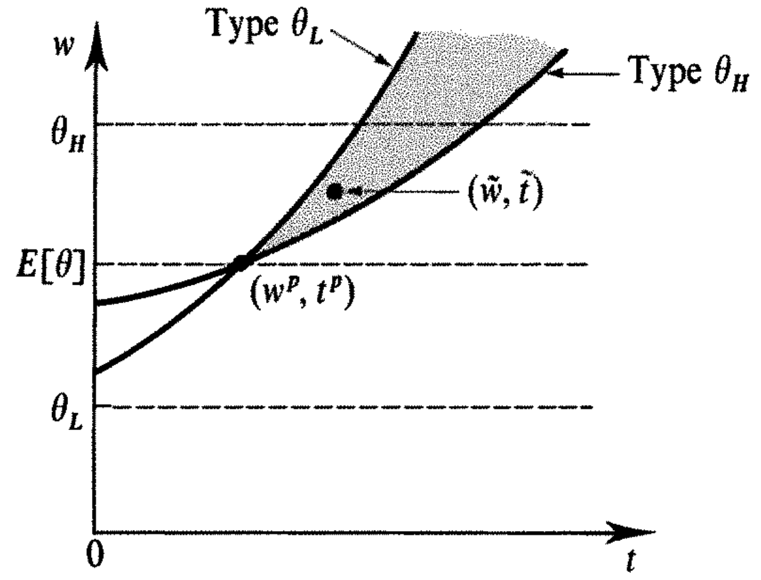

Figure 29: (MWG, Figure 13.D.3)

6.1.1 Incomplete information game

We proceed with a series of lemmata.

Lemma 3. *In any pure-strategy equilibrium (pooling or separating), both firms earn zero profits in expectation.*

Proof: Let $\{(w_h, t_h), (w_l, t_l)\}$ be the contracts that are chosen in equilibrium by the two types of two workers (perhaps the same contract). Suppose that industry expected profits are $\Pi > 0$. Consider a deviation by a firm earning $\Pi/2$ or less: $\{(w_h + \varepsilon, t_h), (w_l + \varepsilon, t_l)\}$ for $\varepsilon > 0$. This contract would be acceptable to both types and would still satisfy incentive compatibility in a separating equilibrium, leaving the deviating firm with $\Pi - \varepsilon > \Pi/2$. Thus, $\Pi = 0$. $\square$

Lemma 4. *No pooling equilibria exist.*

Proof: If a pooling equilibrium exists, (w_p, t_p), Lemma 3 implies it must generate zero profits and thus $w_p = \phi\theta_h + (1 - \phi)\theta_l$. But then single-crossing implies that there is a profitable deviation contract $(\tilde{w}, \tilde{t})$ that is preferred by type θ_h and not by type θ_l. See figure. $\square$

Lemma 5. *In any separating equilibrium, $w_h = \theta_h$ and $w_l = \theta_l$.*

Proof: Suppose that $w_l < \theta_l$. Then either firm could deviate and offer only $(w_l + \varepsilon, t_l)$, attracting the θ_l type worker and making positive profit (even if θ_h are also attracted). Hence, $\theta_l \geq w_l$.

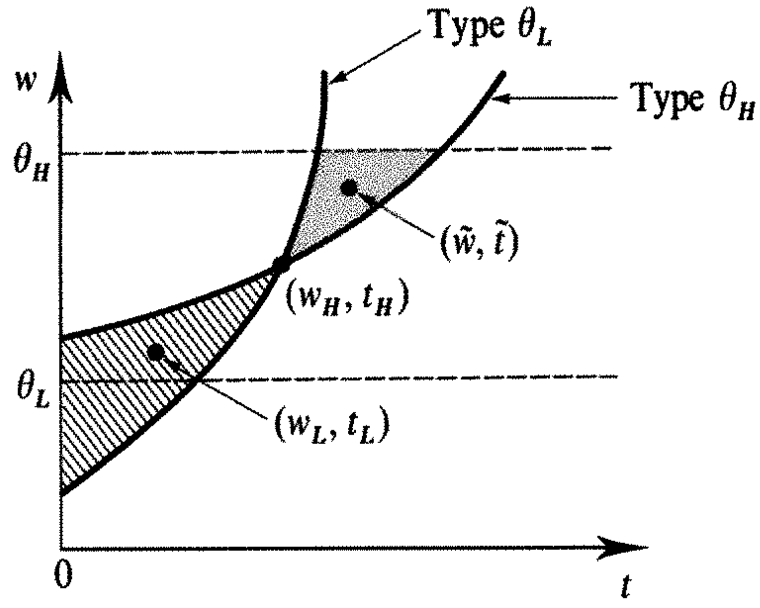

Figure 30: (MWG, Figure 13.D.4)

Suppose that $w_h < \theta_h$. Because Lemma 3 establishes that expected profits must be zero, it must be the case that $w_l > \theta_l$. (See figure below.) As such, any separating equilibrium $\{(w_h, t_h), (w_l, t_l)\}$ has the exhibited features in the graph – namley (i) $w_l > \theta_l$ and, (ii) (w_l, t_l) lies in the cross-hatched area (otherwise it is not incentive compatible). Because of single-crossing, there exists a contract $(\tilde{w}, \tilde{t})$ in the shaded region such that θ_h would choose $(\tilde{w}, \tilde{t})$ instead of (w_h, t_h). Thus, a deviating firm would earn more profit on the new contract. Hence, $w_h \geq \theta_h$. Because $w_l \geq \theta_l$ and $w_h \geq \theta_h$, the zero-profit conclusion of Lemma 3 implies $w_h = \theta_h$ and $w_l = \theta_l$. □

Lemma 6. *In any separating equilibrium,* $(w_l, t_l) = (\theta_l, 0)$.

Proof: Lemma 5 established $w_l = \theta_l$. Suppose that $t_l > 0$. Then we have the situation depicted in Figure 31:

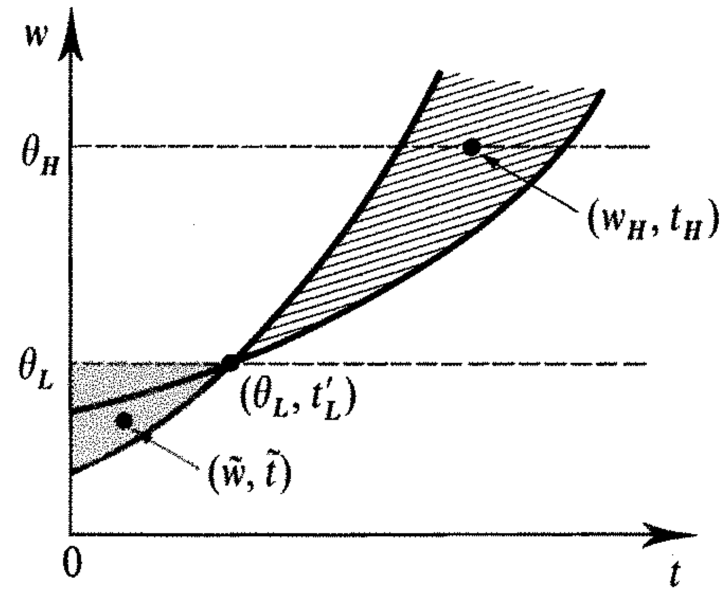

Figure 31: (MWG, Figure 13.D.5)

A deviating firm could offer $(\tilde{w}, \tilde{t})$ in the shaded region. Such an offer would be strictly preferred by the low ability type, but it would not be attractive to the high type. Given it attracts the low-type worker at a lower wage, it would be more profitable. Hence, in any separating equilibrium, $t_l = 0$. $\square$

Lemma 7. *In any separating equilibrium, the high-ability workers are offered and accept* $(w_h = \theta_h, t_h)$ *such that*

$$\theta_h - c(t_h, \theta_l) = \theta_l - c(0, \theta_l).$$

Proof: Previous lemmata have established $w_h = \theta_h$ and $(w_l, t_l) = (\theta_l, 0)$. What remains is to show that t_h is set at a level where the low-type is just indifferent between the two contract choices. Define $\hat{t}$ as

$$\theta_h - c(\hat{t}, \theta_l) = \theta_l - c(0, \theta_l).$$

Notice that in all such situations, if $t_h \neq \hat{t}$, then there exists a shaded region (as in the Figure 32) which includes all contracts that are strictly preferred by the high type to (w_h, t_h) and which are more profitable for the deviating firm.

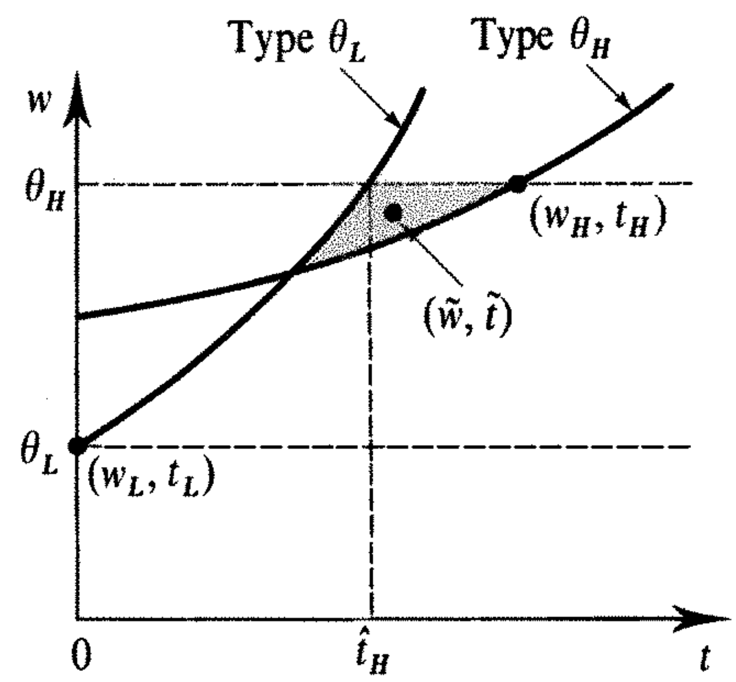

Figure 32: (MWG, Figure 13.D.6)

Thus, $t_h = \hat{t}$, as in the lemma. □

These lemmata together prove the following:

Proposition 6. *In any subgame-perfect equilibrium, firms offer the menu* $\{(\theta_h, t_h), (\theta_l, 0)\}$ *where* t_h *satisfies*

$$\theta_h - c(t_h, \theta_l) = \theta_l - c(0, \theta_l),$$

low-ability workers choose $(\theta_l, 0)$ *and high-ability workers choose* (θ_h, t_h).

Remark: This is not to say that a separating equilibrium always exists. Only that *if* a pure-strategy exists, then it is necessarily the least-cost separating equilibrium.

6.1.2 Possible nonexistence of pure-strategy equilibria

Separating equilibria can fail to exist when the probability of the high-ability type is sufficiently high that a pooling contract can be constructed which is favorable to both types. Indeed, one can easily see that as ϕ becomes sufficiently close to 1, $E[\theta]$ becomes close to θ_h, and the existence of the pooling deviation which destroys the separating equilibrium emerges as illustrated in panel (b) in Figure 33.

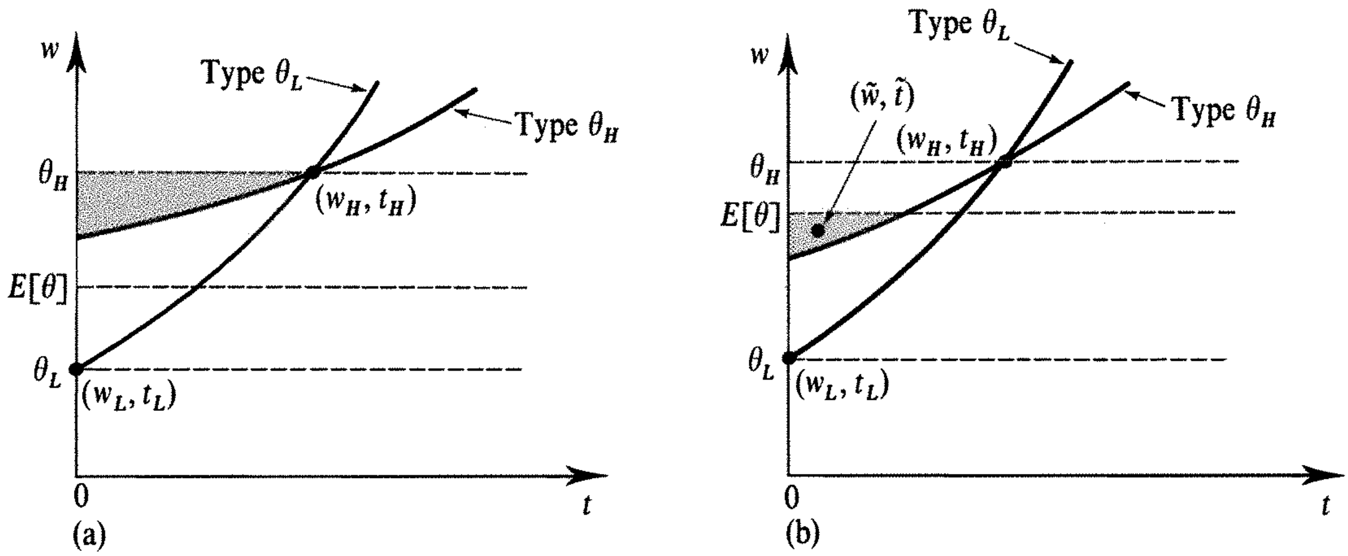

Figure 33: (MWG, Figure 13.D.7)

6.2 Insurance-market model (JR)

We now return the the insurance-market setting of JR (ch. 8) and consider the associated screening game in which the insurers (there are two) move first and offer contracts, followed by the informed consumers choice from the available offers.

Screening game We follow JR's extensive form in which nature chooses the consumer's type after contracts have been offered but before the consumer is called upon to choose. We could have equally well had nature move first, followed by the insurers, and then the consumers. In both cases, the only non-singleton information sets belong to the insurers in their simultaneous move stage and these information sets are always reached. (See figure below from JR, Figure 8.14, but note our notation is slightly different.) Consequently, we can use SPNE as our solution concept.

1. Two insurance companies simultaneously offer a menu of offers. It will be sufficient that each firm, j, offers a menu with (at most) two contracts, $\{\Psi_l^j = (B_l^j, p_l^j), \Psi_h^j = (B_h^j, p_h^j)\}$.
2. Nature chooses the consumer's type, π, with probability ϕ that the consumer is low-risk (accident probability, $\pi_l > 0$), and $(1 - \phi)$ probability that the consumer is high-risk ($1 > \pi_h > \pi_l$).
3. Consumer chooses a single policy from the presented menus, or chooses no insurance.

Full-information outcome

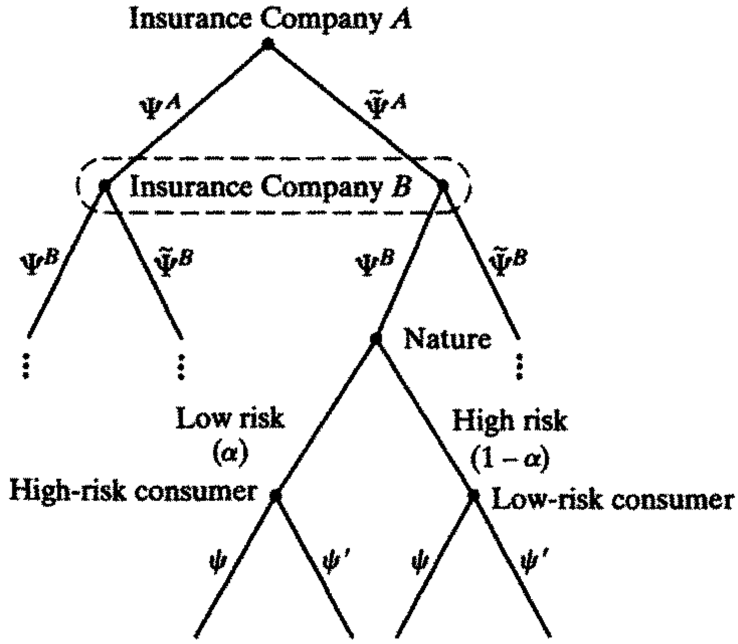

Figure 34: (JR, Figure 8.14)

Under full information (i.e., the consumer's type is known at stage 1 when offers are chosen by the insurers), the unique equilibrium contracts will be for each insurer to offer zero-profit, full insurance contracts. That is, the low-risk consumer will be offered $(L, p = \pi_l L)$ by both firms and the high-risk consumer will be offered $(L, p = \pi_h L)$ by both firms.

Incomplete information game

We will focus on pure-strategy equilibria. This first result is similar to the outcome in the labor-market setting.

Lemma 8. *In every pure-strategy SPNE, firms earn zero expected profits.*

Proof: Because firms do not have to offer any policies, expected profits cannot be negative. We shall show they cannot be positive either.

Suppose that the aggregate expected profits of the two firms is Π,

$$\Pi = \phi(p_l^* - \pi_l B_l^*) + (1-\phi)(p_h^* - \pi_h B_h^*),$$

and that firm j earns no more than $\Pi/2$. There are two cases. First, suppose that there is pooling in equilibrium, $B_l^* = B_h^* = B_p^*$ and $p_l^* = p_h^* = p_p^*$. Firm j could deviate and offer slightly more coverage, $B' = B_p^* + \varepsilon$, for the same price and attract both consumers. This would yield profits arbitrarily close to Π, yielding a contradiction. Second, suppose that there is separation in equilibrium, $(B_l^*, p_l^*) \neq (B_h^*, p_h^*)$. Because each consumer type has to weakly prefer the contract designed for them in a separating equilibrium, and because there is single-crossing, it must be that one of the two consumer types has a slack incentive constraint. That is, either

$$u_l^* = u_l(B_l^*, p_l^*) > u_l(B_h^*, p_h^*), \text{or}$$

$$u_h^* = u_h(B_h^*, p_h^*) > u_h(B_l^*, p_l^*).$$

Suppose that the first strict inequality holds. We want to design a deviation for firm j that allows him to attract (and separate) both types of consumers, and yields him close to Π profits. To this end, choose $B_l' = B_l^* + \varepsilon$ and $B_h' = B_h^* + \delta$ (using the same prices). Clearly, both consumer types will choose contracts from firm j. In order for firm j to maintain separation, we need to show that for ε and δ arbitrarily small

$$u_l(B_l^* + \varepsilon, p_l^*) > u_l(B_h^* + \delta, p_h^*),$$

$$u_h(B_h^* + \delta, p_h^*) > u_h(B_l^* + \varepsilon, p_l^*).$$

(See Figure 35.)

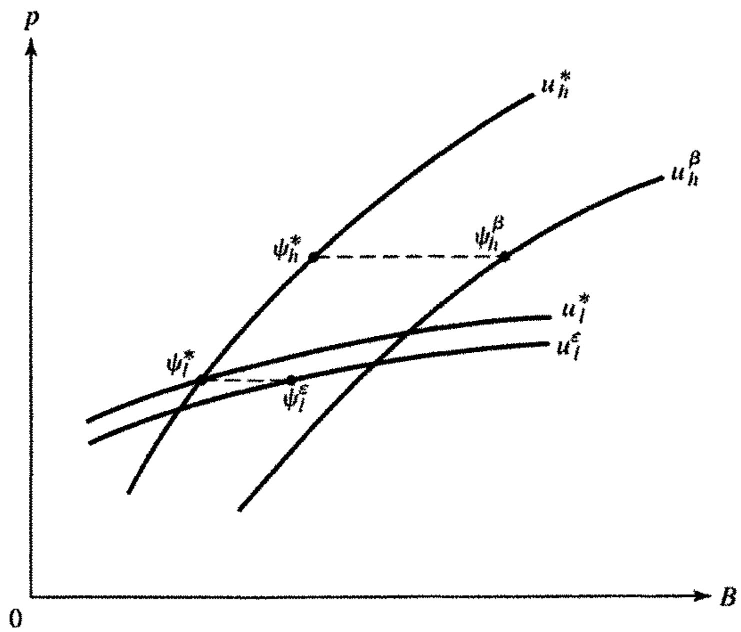

Figure 35: (JR, Figure 8.15)

For ε and δ sufficiently small, the first inequality follows from the hypothesis that $u_l^* > u_l(B_h^*, p_h^*)$. The second inequality can be satisfied by any given δ by making ε smaller:

$$u_h(B_h^* + \delta, p_h^*) > u_h(B_h^*, p_h^*) = u_h^* \geq u_h(B_l^*, p_l^*) = \lim_{\varepsilon \to 0} u_h(b_l^* + \varepsilon, p_l^*).$$

Hence, firm j can offer a separating deviation that earns almost Π. $\square$

Lemma 9. *There is no pure-strategy pooling equilibria in the insurance screening game.*

Proof: Let (B^*, p^*) be the pooling policy that is chosen by both consumer types in equilibrium. Because expected profits are zero (prev. lemma), we have

$$p^* = (\phi\pi_l + (1-\phi)\pi_h)B^*.$$

Note that if $B^* = 0$, then $p^* = 0$ and the consumer is not purchasing any insurance. In this case, a deviating firm can offer $(L, \pi_h L + \varepsilon)$ and earn positive profits on the high-risk type (and possibly profit on the low-risk type). Thus, we must have $B^* > 0$ (and hence $p^* > 0$) as in Figure 36.

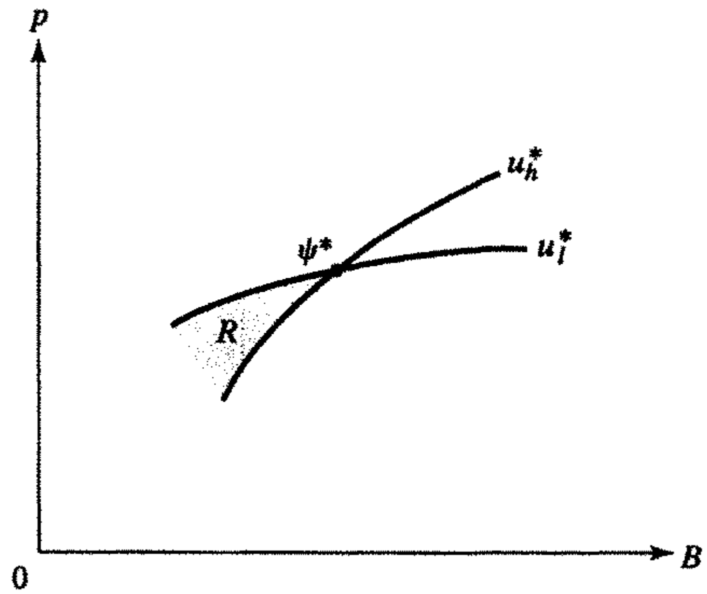

Figure 36: (JR, Figure 8.16)

But in this case, firm j could offer a single contract in the shaded region that is attractive to the low-risk consumer and not to the high-risk consumer (who would continue to buy (B^*, p^*) from the other firm. Because (B^*, p^*) generates zero expected profit when both types purchase it, it yields strictly positive profits when only the low-risk consumer purchases it. Hence, one can choose a contract arbitrarily close to (B^*, p^*) in the shaded region and earn positive profits. $\square$

Lemma 10. *In any separating equilibrium, the high-risk consumer must obtain utility $u_h(L, \pi_h L)$.*

Proof: If this were not the case, then a deviating firm could offer the high-risk consumer $(L, \pi_h L + \varepsilon)$ and earn ε as profit (and not lose anything on the low-risk type). $\square$

Lemma 11. *In any separating equilibrium, the low-risk consumer's contract must lie in the low-risk zero-profit line: given B_l, it must be that $p_l = \pi_l B_l$.*

Proof: Because the high-risk type must earn $u_h(L, \pi_h L)$ in payoff, (B_h^*, p_h^*) must lie on or below the high-risk zero-profit line, leading to nonpositive profits. Because expected profits are zero, this implies (B_l^*, p_l^*) must lie on or above the low-risk zero-profit line. Suppose it lies strictly above. Then there is a shaded region, R, as in Figure 37:

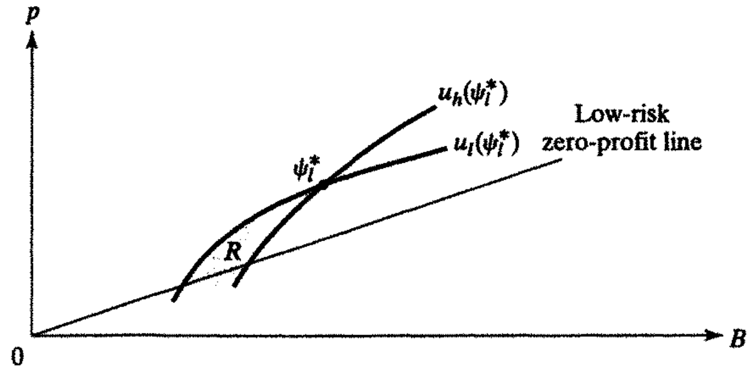

Figure 37: (JR, Figure 8.19)

In this case, one firm could offer a single policy in the shaded region and attract only the low-risk type, making positive profit. A contradiction. □

Lemma 12. *In any separating equilibrium, the high-risk consumer's contract is $(L, \pi_h L)$.*

Proof: Because (B_l^*, p_l^*) is on the low-risk zero-profit line, (B_h^*, p_h^*) must lie on the high-risk zero-profit line. $(B_h^* = l, p_h^* = \pi_h L)$ is the unique point generating utility $u_h(L, \pi_h L)$ and that earns zero profit. □

Lemma 13. *In any separating equilibrium, the low-risk consumer's contract satisfies*

$$u_h(B_l, \pi_l B_l) = u_h(L, \pi_h L).$$

Proof: We have established $(B_h^*, p_h^*) = (L, \pi_h L)$ and that $B_l^*, p_l^*) = (B_l^*, \pi_l B_l^*)$. What remains is to show that only one B_l^* can possibly arise in equilibrium. Suppose that B_l^* does not satisfy the hypothesis:

$$u_h(B_l, \pi_l B_l) < u_h(L, \pi_h L).$$

Then we have the situation in the Figure 38 in which some $\psi'' = (B'', \pi_l B'')$ is chosen by the low-risk type in equilibrium:

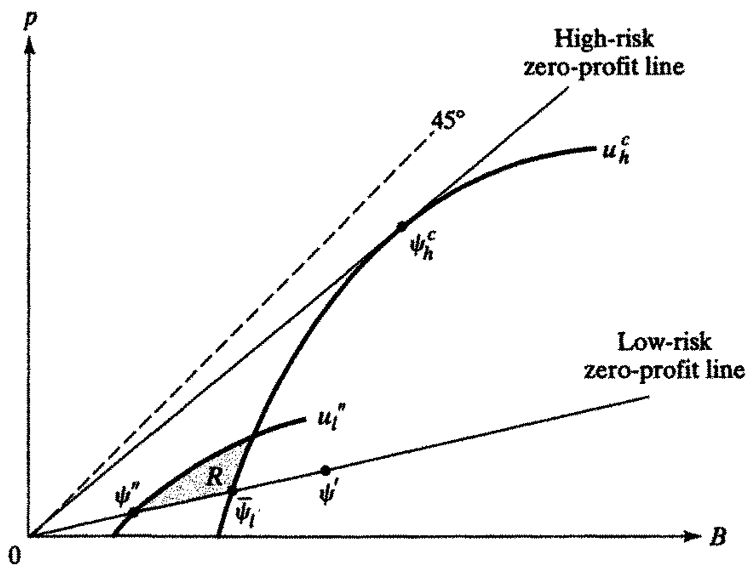

Figure 38: (JR, Figure 8.20)

This gives rise to a shaded region, R. A deviating firm can offer a contract in the shaded region, attract the low-risk type, and earn positive profit. $\square$

Threading together our results, we conclude ...

Proposition 7. *In any separating equilibrium, the chosen contracts are* $(B_h, p_h) = (L, \pi_h L)$ *and* $(B_l, \pi_l B_l)$ *where*

$$u_h(B_l, \pi_l B_l) = u_h(L, \pi_h L).$$

Of course, this does not imply that a separating equilibrium exists. Only that if one does exist, then it necessarily is the least-cost separating equilibrium.

6.2.1 Non-existence of pure-strategy equilibria

As in the case of labor-market screening games, if the probability that the informed type is the good type (i.e., low risk) is sufficiently high (i.e., ϕ is close to one), then a pooling contract deviation exists which prevents separation from arising in equilibrium. See Figure 39, for example.

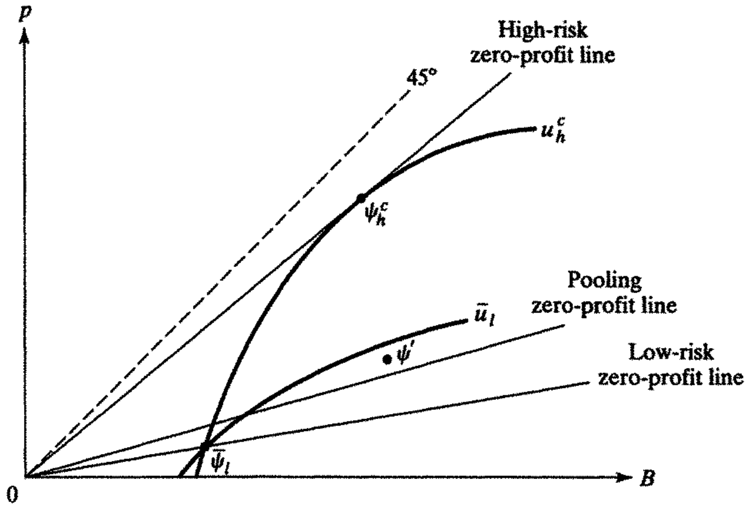

Figure 39: (JR, Figure 8.21)

Remarks:

1. Existence in mixed strategies (Dasgupta and Maskin (1986); full characterization of unique (possibly mixed-strategy) equilibrium in Luz (*Theoretical Economics*, 2017).

2. Exclusivity is very important. With non-exclusive contracts, simple Akerlof-style adverse selection contracts may emerge as the unique equilibrium. See Attar, Marriotti, Salanie (*Ema*, 2011), for example.

3. Azevedo, Gottlieb, "Perfect competition in markets with adverse selection," (*Ema*, 2017). Novel market-based solution concept that allows for endogenous characteristics, multidimensional preferences, and uniqueness (and existence).

   - Agent preferences are $u(x, p, \theta)$ where $\theta \in \Theta$ is some type (possibly multidimensional) that has measure μ, p is price and $x \in X$ is a contract (e.g., level of coverage, etc.). A firm's marginal cost of contract x taken up by consumer θ is $c(x, \theta)$.
   - AG introduce two concepts: **weak equilbrium** and **(strong) equilibrium**, where I have added "strong" for clarity.
   - A **weak equilibrium** is a price schedule, $p^*(\cdot) : X \to \mathbb{R}_+$ and an allocation α^*, which is a probability measure over (θ, x) choices. It is the natural generalization to the single-price notion of competitive equilibrium with market price p^* and set of types who purchase, Θ^*. The requirements are weaker than a competitive equilibrium, as you would guess. There are two conditions:
     (a) For all $x \in X$ such that $\int \alpha^*(\theta, x) d\mu(\theta) > 0$ (i.e., some types actually purchase x in the weak equilbrium,

$$p^*(x) = E_{\alpha^*)}[c(x, \theta)],$$

where $E_{\alpha^*}[\cdot]$ is the expectation taken with respect to measure α^*, conditional on x,

(b) for almost all (θ, x) such that $\alpha^*(\theta, x) > 0$

$$x \in \arg\max \; u(x, p^*(x), \theta).$$

Thus, a weak equilibrium requires that consumers are optimizing with respect to the given price schedule, and firms and not losing money *on contracts chosen in equilibrium*.

- The weak equilibrium is very weak. There are no profit restrictions for contracts which are not chosen in weak equilibrium. Moreover, supply does not need to equal demand. Prices $p^*(x)$ can be so high for $x \neq \mathbf{0}$ that $x = 0$ is chosen (at $p^*(0) = 0$) by all types. Obviously, there would be excess supply at these prices. In short, lots of weak equilibria exist.
- The additional requirement that AG add is meant to mimic the forces of entry in pushing down prices. Let $\bar{X}$ be a finite subset of X and suppose for each contract in $\bar{X}$, there is a behavioral-type agent, denoted $\bar{x} \in \bar{X}$ who will buy it at any price and who has zero cost to the firm, $c(\bar{x}, \bar{x}) = 0$. Suppose that there is a small probability of each of these types, and let their measure be η. (AG will ultimately take the limit as η converges to zero.) Now consider a sequence of $(\bar{x}^n)_n \in \bar{X}^n$ where $\bar{x}^n$ converges to x and $\bar{X}^n$ converges to X as $\eta^n(\bar{X}^n)$ converges to zero. For any such sequence, let (p^n, α^n) be a weak equilibrium. Then we say a weak equilibrium (p^*, α^*) is a (strong) **equilibrium** if α^n converges to α^* and for every sequence $(x^n)_n$ with $x^n \in \bar{X}^n$ and limit $\bar{x} \in X$, $p^n(x^n)$ converges to $p^*(\bar{x})$.
- Practically speaking, the perturbations require that prices decrease until they get to the point where they are attractive to some standard (non-behavioral) type. In a sense, all contracts become active with positive probability.
- **Existence.** AG demonstrate existence every game assuming reasonable technical properties (e.g., bounded marginal rates of substitution, continuity of u and c, etc.). (Theorem 1 in AG, 2017.)
- **Properties.** The most interesting property is that for every contract $\tilde{x}$ with positive price, $p^*(\tilde{x}) > 0$, there is a (θ, x) with $\alpha^*(\theta, x) > 0$ such that

$$u(x, p^*(x), \theta) = u(\tilde{x}, p^*(\tilde{x}), \theta), \;\text{ and }\; c(\tilde{x}, \theta) \geq p^*(\tilde{x}).$$

That is, for every contract, $\tilde{x}$ with a positive price there is some consumer θ and a contract, x, such that the consumer is indifferent to $\tilde{x}$, and no firm would make profits from $\tilde{x}$ if it were chosen. For many of the simple (one-dimensional) screening games we have considered, this provides a unique $p^*(x)$ which corresponds to the equilibrium in Rothschild and Stiglitz (even when the pure-strategy equilibrium in RS does not exist!).

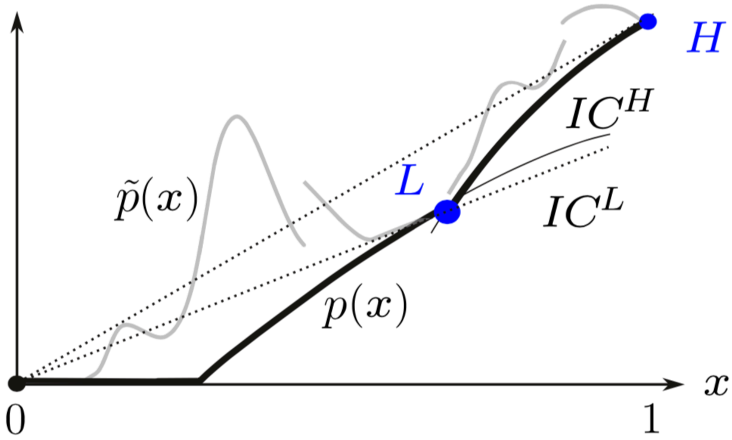

Figure 40: Figure from Azevedo and Gottlieb (2017). Dotted lines are zero-profit lines; blue dots are contracts chosen in equilibrium, and bold curve represents $p^*(\cdot)$.

- Because of the indifference property of the equilibrium $p^*(\cdot)$, it is straightforward to compute for situations with multi-dimensional uncertainty.