Chapter 11

Dynamic Games with Incomplete Information: Signaling and Screening

When informed players move first, their *actions speak*: choices reveal, conceal, or distort private information. This chapter develops perfect Bayesian equilibrium and its two flagship applications—signaling and cheap talk—plus the screening problem faced by uninformed movers.

11.1 Perfect Bayesian equilibrium

Subgame perfection has no grip when information sets prevent the tree from decomposing into subgames. The fix: require optimality at *every information set*, given explicitly specified *beliefs* that are consistent with Bayes' rule.

Definition 11.1: Perfect Bayesian equilibrium

A *perfect Bayesian equilibrium* (PBE) consists of a strategy profile and a *belief system* (a probability distribution over nodes at each information set) such that:

1. **Sequential rationality**: at every information set, the mover's strategy is optimal from that point on, given her beliefs and others' strategies;
2. **Belief consistency**: on the equilibrium path, beliefs are derived from the strategies by Bayes' rule; off the path (probability-zero events), beliefs are unrestricted (but must still support sequentially rational behavior).

The freedom in off-path beliefs generates multiplicity; *refinements* such as the intuitive criterion (below) and sequential equilibrium (Chapter 13) discipline it.

11.2 Signaling: Spence's education model

Example 11.1: Job-market signaling (Spence 1973)

A worker is of high ability $\theta_H = 2$ (probability μ) or low ability $\theta_L = 1$. She chooses education level $e \geq 0$ *before* entering the job market; education is pure signal (does not raise productivity) and costs e/θ —*cheaper for the able*. Competitive firms then pay a wage equal to expected productivity given e. Worker's payoff: $w - e/\theta$.
Separating equilibrium: types choose different education levels, revealing themselves;

$e_L^* = 0$, $w(e_L^*) = 1$ and $e_H^* = e^s$, $w(e^s) = 2$. Incentive compatibility requires the low type not to imitate: $2 - e^s/1 \leq 1 \Rightarrow e^s \geq 1$; and the high type to prefer signaling: $2 - e^s/2 \geq 1 \Rightarrow e^s \leq 2$. Any $e^s \in [1, 2]$ works, supported by beliefs that any $e < e^s$ means low ability.

Pooling equilibria also exist: both types choose e^p and are paid $1 + \mu$, supported by pessimistic off-path beliefs.

Definition 11.2: Single-crossing property

Signaling works when the signal is *differentially costly*: the marginal cost of the signal is lower for higher types, so indifference curves of different types cross exactly once in signal–transfer space. This *single-crossing* (Spence–Mirrlees) condition is the engine of all signaling and screening models.

Theorem 11.1: The intuitive criterion (Cho–Kreps 1987)

Fix a PBE. Suppose an off-path signal is *equilibrium-dominated* for some type (that type could not gain by sending it under *any* belief-response), but some other type *could* gain if it convinced the receiver. Then beliefs after that signal must put zero probability on the dominated type. In the Spence model, this refinement eliminates all pooling equilibria and all separating equilibria except the least-cost one ($e^s = 1$: the *Riley outcome*).

Key Idea: Costly signals must be differentially costly

Burning resources credibly communicates only what your *willingness to burn* reveals. Education requirements, extravagant advertising, warranties, engagement rings, lavish courtship displays (the peacock's tail—Zahavi's handicap principle), IPO underpricing: all are equilibria in which the cost of the signal is the message. Note the social ambivalence: in pure-signaling models, the signaling expenditure is *wasteful*—the private return to education exceeds its social return.

11.3 Cheap talk

What if talk is free—no signal cost at all? Credible communication is still possible, but only partially, and only when interests align enough.

Example 11.2: Crawford–Sobel (1982) uniform-quadratic model

A sender observes a state $\theta \sim U[0, 1]$ and sends a costless message to a receiver, who then chooses action $a \in \mathbb{R}$. Payoffs: receiver loses $(a - \theta)^2$; sender loses $(a - \theta - b)^2$, where $b > 0$ measures the conflict of interest (the sender wants action biased upward).

Full revelation is never an equilibrium for $b > 0$ (the sender would inflate). All equilibria are *partitional*: the sender reveals only which interval $[\theta_{k-1}, \theta_k]$ the state lies in. Indifference at the cut points forces interval lengths to grow by $4b$ each step, so the maximal number of intervals is the largest K with $2K(K-1)b < 1$. As $b \to 0$, $K \to \infty$ (communication becomes perfect); for $b \geq 1/4$, only *babbling* (uninformative talk) survives.

Remark: Babbling is always an equilibrium

In any cheap-talk game there is an equilibrium in which messages are ignored and hence carry no information. Cheap talk can create meaning only as one equilibrium among several—language itself is an equilibrium convention.

11.4 Screening: the uninformed side designs the menu

Reverse the timing: the *uninformed* party moves first, offering a menu of contracts to induce types to self-select. (This is a bridge to mechanism design, Chapter 14.)

Example 11.3: Price discrimination via menus

A seller with cost zero faces a buyer whose taste is $\theta_H = 2$ (probability μ) or $\theta_L = 1$; a buyer of type θ enjoying quality q and paying t gets $\theta q - t$; quality costs the seller $q^2/2$. The seller offers a menu $\{(q_L, t_L), (q_H, t_H)\}$ maximizing profit subject to *individual rationality* (each type accepts) and *incentive compatibility* (each type picks its own contract). The solution:

$$q_H = \theta_H \text{ (efficient—“no distortion at the top”)}, \qquad q_L = \theta_L - \frac{\mu}{1-\mu}(\theta_H - \theta_L) \;<\; \theta_L,$$

with the low type’s quality distorted *downward* (possibly excluded entirely) and the high type earning an *information rent*. Every airline (economy vs. business), software vendor (feature tiers), and insurer (deductible menus) solves this problem; the deliberately degraded low-end product is not a design failure but an incentive device.

11.5 Reputation and dynamic signaling (a synthesis)

The chain-store analysis of Example 9.2 can now be made precise: in each period the incumbent’s fighting decision is a *signal* about its type; entrants update by Bayes’ rule; and the rational incumbent randomizes between mimicking the tough type and revealing itself. Key general results:

- **Reputation bounds** (Fudenberg–Levine 1989): a long-run player facing a sequence of short-run opponents, with any positive prior on a “commitment type,” secures nearly the payoff of committing to that type’s strategy. Small doubts have large, permanent effects.
- **Reputation dynamics**: incentives to invest in reputation are strongest when uncertainty about type is greatest; established reputations may be milked (temporary quality cuts) and rebuilt—matching observed brand dynamics.

Exercises

Exercise 11.1: Beer–Quiche and the intuitive criterion

A defendant is strong (prob. 0.9) or weak; he chooses breakfast: Beer or Quiche. Strong types prefer Beer, weak prefer Quiche (worth 1 to eat your preferred breakfast); avoiding a duel is worth 2 to either type. An observer sees breakfast, then duels iff he believes the defendant is more likely weak. Find both pooling PBEs (all-Beer and all-Quiche), then show the all-Quiche equilibrium fails the intuitive criterion.

Exercise 11.2: Separating wages

In the Spence model with $\theta_L = 1$, $\theta_H = 3$ and education cost e^2/θ, find the set of separating education levels e^s and the Riley outcome. Then let productivity itself rise with education (wage $\theta + \frac{1}{2}e$ for a worker of known type θ with education e) and recompute the least-cost separating level. Is education still (partly) wasteful?

Exercise 11.3: Two-interval cheap talk

In the Crawford–Sobel model with $b = 1/10$, compute the maximal number of intervals K, the equilibrium cutpoints for the 2-interval equilibrium, and the receiver's expected loss under babbling vs. 2-interval communication.

Exercise 11.4: Insurance screening

High-risk and low-risk drivers have accident probabilities $p_H > p_L$, common wealth w, loss L, and identical concave utility. Show that in a competitive (Rothschild–Stiglitz) screening equilibrium—if it exists—the high-risk type gets full insurance at its fair premium and the low-risk type gets *partial* coverage, and explain why a pooling contract cannot survive competition.