Chapter 1

Introduction to Contract Theory

Instructor Slides

Contract Theory: A Short Introduction

Ana Espinola-Arredondo, Felix Muñoz-Garcia, and Pak-Sing Choi

Outline of Chapter 1

- This chapter introduces the field of contract theory and provides historical and conceptual foundations for the rest of the book.
- Sections:
  - 1.1 Introduction – what is contract theory and why does it matter?
  - 1.2 A brief history of contracts – from ancient law to modern practice.
  - 1.3 A brief history of contract theory – key economists and breakthroughs.
  - 1.4 Real-world applications – labor, insurance, finance, and policy.

What you will learn in Chapter 1

- The **definition** of a contract and the key challenges of contract design.
- The two fundamental principles: **incentive compatibility** and **participation constraints**.
- How **information asymmetries** give rise to moral hazard and adverse selection.
- The **historical roots** of contract law from ancient Mesopotamia to the 19th century.
- The **intellectual history** of contract theory, from Coase to Hart and Holmström.
- How contract theory is used in **practice**: labor markets, insurance, finance, and public policy.

Section 1.1
Introduction

What is Contract Theory?

- Contract theory lies at the intersection of **microeconomics**, **game theory**, and **information economics**.
- It provides tools to analyze how agents can achieve mutually beneficial outcomes despite:
  - **Conflicting interests** – principals and agents want different things.
  - **Incomplete information** – one party knows more than the other.
  - **Unverifiable actions** – some decisions cannot be observed or contracted upon.

What is Contract Theory? – Key questions

- Key questions:
  - How should a contract be written to elicit the desired behavior?
  - Who should bear risk in a relationship?
  - How do information problems affect efficiency?

The Principal-Agent Framework

- The canonical setting in contract theory involves two parties:
  - **Principal**: the party who designs the contract (e.g., employer, insurer, regulator).
  - **Agent**: the party who takes an action or possesses private information (e.g., worker, policyholder, firm).
- The principal's goal is to design a contract that induces the agent to act in a way that is beneficial for the principal.

The Principal-Agent Framework – Examples

- Examples:
  - Employer (principal) and worker (agent).
  - Insurer (principal) and policyholder (agent).
  - Government (principal) and contractor (agent).
  - Investor (principal) and manager (agent).

Two Guiding Principles of Contract Design

- Contract theory relies on two core principles:
- **1. Incentive Compatibility (IC):**
  - The contract must be designed so that the agent finds it optimal to behave in the way the principal desires – acting in their own self-interest must align with the contract's goals.
  - Example: a sales commission contract where working harder genuinely benefits the salesperson.

Two Guiding Principles – Participation Constraint

- **2. Participation Constraint (PC):**
  - All parties must obtain at least as much utility from participating in the contract as from their best outside option (reservation utility $\overline{u}$).
  - Example: the worker must prefer the offered contract over unemployment or alternative jobs.

Incentive Compatibility – A closer look

- **Formal statement:** A contract is incentive compatible if, for every agent type or action choice a, truthful reporting or the desired action a^* satisfies:

$$U(a^*) \geq U(a) \quad \text{for all } a.$$

- The agent has no incentive to deviate from the contracted behavior.

Incentive Compatibility – Two versions

- **Two versions:**
  - *Dominant strategy IC:* the agent's best response is the desired action regardless of what others do.
  - *Bayesian IC:* the agent's best response is the desired action given expectations about others' behavior.
- IC is the cornerstone of moral hazard under asymmetric information (Chapters 3–4) and adverse selection models (Chapters 5–6).

Participation Constraint – A closer look

- **Formal statement:** An agent with reservation utility $\overline{u}$ accepts the contract if and only if:

$$U(\text{contract}) \geq \overline{u}.$$

- The reservation utility $\overline{u}$ represents:
  - Outside job opportunities.
  - Unemployment benefits.
  - The utility from leisure if the agent does not work.

Participation Constraint – Key insight

- In equilibrium, the principal will make the PC bind ($U = \overline{u}$) to extract maximum surplus from the agent.
- **Key insight:** the principal cannot extract more than the agent's outside option. Competition for agents drives $\overline{u}$ up, increasing wages.

Types of Information Problems

- Contract theory distinguishes two fundamental information asymmetries:
- **1. Moral Hazard (hidden action):**
  - The agent takes an action *after* signing the contract that the principal cannot observe.
  - Example: a worker exerts low effort after being hired.
  - *Why the name?* It comes from 19th-century **insurance**: once insured, a party may act less carefully (a covered warehouse owner is less cautious about fire), changing the "hazard" the insurer faces.
  - Studied in Chapters 2, 3, and 4.

Types of Information Problems – Adverse selection

- **2. Adverse Selection (hidden type):**
  - The agent has private information *before* the contract is signed.
  - Example: an insurance applicant knows their own health risk better than the insurer.
  - Studied in Chapters 5 and 6.

Moral Hazard vs. Adverse Selection

- **Moral Hazard:**
  - Timing: hidden action occurs *after* contracting.
  - Problem: agent may take a costly action (effort) that benefits the principal but is unobservable.
  - Solution tools: performance-based pay, monitoring, bonuses.
- **Adverse Selection:**
  - Timing: hidden information exists *before* contracting.
  - Problem: agents with different types (e.g., high vs. low ability) self-select into contracts not designed for them.
  - Solution tools: screening (principal designs a menu of contracts), signaling (agent sends a costly signal).

Moral Hazard vs. Adverse Selection – Solution tools

- Both problems lead to **inefficiency** relative to the symmetric information benchmark.

Signaling

- A key mechanism under asymmetric information: **signaling**.
- Here, the *informed party* (agent) moves first and sends a costly signal to *credibly* reveal their type to the uninformed party (principal).
- Classic example: Spence (1973) education model.
  - High-ability workers signal their type by acquiring education, even if education itself does not raise productivity.
  - Education is a credible signal because it is *less costly* for high-ability workers to acquire.

Signaling – Spence and equilibria

- Signaling leads to **separating equilibria** (types distinguished) or **pooling equilibria** (types not distinguished).
- Studied in Chapter 5.

Mathematical Tools in Contract Theory

- Contract theory uses a rich set of mathematical methods:
- **Lagrangian optimization:** solving constrained maximization problems (principal's problem with IC and PC constraints).
- **Bayesian games:** analyzing strategic interactions under incomplete information (used in signaling and mechanism design).

Mathematical Tools – Probability and mechanism design

- **Comparative statics:** how do optimal contracts change when parameters (risk aversion, outside options, effort costs) change?
- **Probability theory:** effort affects the probability distribution of outcomes; conditional probabilities $f(\pi|e)$ are central.
- **Mechanism design:** reverse-engineering the rules of the game to achieve a desired outcome (Chapter 7).

Risk and Contract Design

- **Risk attitudes** play a central role in contract theory:
  - **Risk-neutral** agents only care about expected value.
  - **Risk-averse** agents dislike variance; they prefer a certain outcome to a risky one with the same expected value.
  - **Risk-loving** agents prefer variance.
- **Optimal risk-sharing:**
  - The less risk-averse party should bear more of the risk.
  - Classic result: risk-neutral principal + risk-averse agent $\Rightarrow$ full insurance for the agent (fixed wage).
  - But full insurance destroys effort incentives $\Rightarrow$ **trade-off between risk-sharing and incentive provision**.

Risk and Contract Design – Optimal risk-sharing

- This trade-off is the central tension in Chapters 2–4.

Efficiency and Information Rents

- Under **symmetric information**: the principal achieves the first-best outcome – the contract is fully efficient.
- Under **asymmetric information**: the principal can only achieve a second-best outcome.
  - **Moral hazard:** the agent exerts less than first-best effort because hidden actions cannot be contracted upon directly.
  - **Adverse selection:** the principal must leave *information rents* to agents with better private information to induce them to reveal their type truthfully.

Efficiency and Information Rents – Key lesson

- **Key lesson:** information asymmetries generate inefficiency and redistribute surplus from the principal to informed agents.

Section 1.2
A Brief History of Contracts

Why study the history of contracts?

- Contracts are not a modern invention – they have been a cornerstone of economic and social organization for millennia.
- Understanding the history of contracts helps us appreciate:
  - Why formal agreements are necessary for social order and economic stability.
  - How contract law has evolved in response to the complexity of economic transactions.
  - The universality of the problems that contract theory addresses: trust, enforcement, and information.

Why study the history of contracts? – The core challenge

- From ancient clay tablets to digital smart contracts, the core challenge has remained the same: **how to make promises credible and enforceable**.

Ancient Mesopotamia: The First Contracts

- The earliest known formal agreements date to **ancient Mesopotamia**, around 2100 BCE.
- **Code of Ur-Nammu** (c. 2100 BCE): one of the earliest legal codes, with provisions for contracts involving sales, loans, and property.
- **Code of Hammurabi** (c. 1754 BCE): a more comprehensive code with 282 laws governing trade, employment, property, and penalties for breach of contract.

Ancient Mesopotamia – The economic challenges

- These early codes reflect the fundamental economic challenges:
  - How to incentivize farmers to work hard on leased land.
  - How to ensure merchants honor trade agreements over long distances.
  - How to allocate risk between lenders and borrowers.

Ancient Mesopotamia: Economic Relevance

- The Code of Hammurabi contains remarkably modern contract provisions:
  - **Employment contracts**: wages for different types of workers, specifying duties and penalties for shirking.
  - **Loan contracts**: interest rate limits and collateral requirements to reduce default risk.
  - **Trade contracts**: rules for merchants traveling on behalf of investors – an early form of principal-agent relationship.
- **Modern connection:** the farmer-investor relationship in Hammurabi's code is structurally identical to the principal-agent problem we study in Chapters 2–4: an investor (principal) contracts with a farmer (agent) who exerts unobservable effort.

Roman Contract Law

- Roman law (c. 500 BCE – 500 CE) introduced fundamental concepts still central to modern contract law:
- **Consensus ad idem** ("meeting of the minds"): a contract requires mutual agreement on the same terms by all parties. This is the foundation of the IC/PC framework.

Roman Contract Law – Emphasis on mutual consent

- **Classification of contracts:**
  - *Verbal contracts (stipulatio)*: binding through spoken words.
  - *Written contracts (litteris)*: recorded for evidence.
  - *Real contracts (re)*: formation by delivery of goods.
- Roman law emphasized **mutual consent** and the formal requirements for valid agreements, distinguishing enforceable from unenforceable promises.

Roman Contract Law: Legacy

- Roman contract principles were absorbed by:
  - **Medieval European legal systems**: through canon law of the Catholic Church.
  - **English common law traditions**: the concept of "consideration" (something of value exchanged) derives partly from Roman ideas.
  - **Continental civil law systems**: the Napoleonic Code (1804) drew heavily on Roman law.
- **Modern connection:** the Roman requirement of *consensus ad idem* maps directly to the participation constraint (PC): a contract is only valid if both parties willingly agree – i.e., if both earn at least their reservation utility.

The Medieval Period: The Law Merchant

- The expansion of medieval trade (10th–15th centuries) required contract rules that could operate *across* different legal systems.
- **Law Merchant (Lex Mercatoria)**: a body of commercial law developed by and for merchants, operating across European trade routes.
  - Standardized contract practices across different kingdoms and jurisdictions.
  - Resolved disputes through merchant courts, not state courts.
  - Built on reputation and repeated relationships – an early solution to the enforcement problem.

The Medieval Period – Doctrine of consideration

- **Doctrine of consideration** in English common law: a contract requires that something of value be exchanged (not just a promise). This underpins the PC requirement in modern contract theory.

The Enlightenment and 19th Century

- Enlightenment philosophers contributed theoretical foundations for contract law:
  - **John Locke**: natural rights and the social contract – individuals enter agreements to protect their rights.
  - **David Hume**: analyzed the role of conventions and reputation in sustaining cooperation without formal enforcement.
- By the **19th century**, the principles of contract law were fully established:
  - Freedom of contract: parties can agree to any terms.
  - Sanctity of contract: courts will enforce valid agreements.
  - Remedies for breach: damages, specific performance.

The Enlightenment and 19th Century – 19th-century principles

- These principles form the legal backdrop for the economic models of contract theory developed in the 20th century.

Summary: Evolution of Contract Law

- Across all historical periods, three recurring challenges drove the development of contract law:
  1. **Enforcement**: how to ensure parties honor their commitments.
  2. **Information**: how to deal with one party knowing more than the other.
  3. **Risk allocation**: how to distribute uncertainty between contracting parties.
- These three challenges are *exactly* what modern contract theory formalizes:
  - Enforcement $\rightarrow$ incentive compatibility constraints.
  - Information $\rightarrow$ moral hazard and adverse selection.
  - Risk allocation $\rightarrow$ optimal risk-sharing and wage structures.

Section 1.3
A Brief History of Contract Theory

Ronald Coase and the Nature of the Firm (1937)

- **Ronald Coase** (Nobel Prize, Economics, 1991) laid the groundwork for contract theory with his seminal 1937 paper "The Nature of the Firm."
- **Key question:** why do firms exist if markets are efficient?
- **Coase's answer:** firms exist because markets have **transaction costs** – the costs of writing, negotiating, and enforcing contracts.
  - When transaction costs are high, it is cheaper to organize production within a firm than to contract through the market.
  - This insight introduced the idea that contracts are *costly* to write – a precursor to incomplete contracts.

Ronald Coase and the Nature of the Firm – Shift of focus

- Coase's work shifted the focus from price theory to **institutional design** – how contracts shape economic organization.

Jensen and Meckling: Agency Theory (1976)

- **Michael Jensen** and **William Meckling** advanced the study of **agency theory** in their influential 1976 paper.
- **Key contribution:** formalized the conflict of interest between principals (shareholders) and agents (managers).
  - Managers may pursue their own interests (perks, empire building) rather than maximizing shareholder value.
  - This is a classic **moral hazard** problem: the manager's effort is not perfectly observable.

Jensen and Meckling – Solutions and legacy

- **Solutions:** design contracts that align manager and shareholder incentives – stock options, performance pay, monitoring.
- This work established the vocabulary and framework that became the foundation of Chapters 2–4 of this book.

Early Moral Hazard Models: 1970s

- The 1970s saw the formal mathematical development of moral hazard models:
- **Ross (1973):** developed one of the first formal principal-agent models; analyzed the conditions under which the agent's interests can be aligned with the principal's.

Early Moral Hazard Models – Shavell and legacy

- **Mirrlees (1975, 1976):** studied optimal income taxation as a principal-agent problem; introduced the **first-order approach** (replacing the IC constraint with the agent's FOC) – a key technique used in Chapter 4.
- **Shavell (1979):** analyzed risk-sharing and incentive provision in the context of liability and insurance contracts; showed the fundamental trade-off between risk-sharing and incentive provision.
- These papers established the formal structure of the moral hazard models studied in this book.

Oliver Hart and Bengt Holmström: Nobel Prize 2016

- **Oliver Hart** (Harvard) and **Bengt Holmström** (MIT) were awarded the Nobel Prize in Economics in 2016 for their contributions to contract theory.
- **Holmström's contributions:**
  - *Moral Hazard and Observability* (1979): derived conditions for optimal incentive contracts; the **informativeness principle** (any signal informative about effort should be included in the contract).
  - *Multitask Principal-Agent Analyses* (1991, with Milgrom): when agents perform multiple tasks, incentives on one task can distort effort on others – studied in Chapter 4.

Hart and Holmström – Hart's contributions

- **Hart's contributions:**
  - Incomplete contracts and the theory of the firm (with Moore and Grossman).
  - Foundations for property rights and vertical integration.

Holmström's Informativeness Principle

- One of the most important results in contract theory:
- **Informativeness Principle** (Holmström, 1979):
  - A signal s should be included in the agent's compensation contract if and only if it provides additional information about the agent's action (effort) beyond what is already contained in the observed output.
  - More precisely, if s is informative about e conditional on π, then the optimal contract $w(\pi, s)$ depends on s.

Holmström's Informativeness Principle – Practical implication

- **Practical implication:** relative performance evaluation – compensating a CEO based not just on the firm's own performance but also on competitors' performance (which removes common shocks).
- Studied in detail in **Chapter 4** (Section 4.6 on informative signals).

Adverse Selection: Akerlof and Spence

- **George Akerlof** (Nobel 2001, jointly): "The Market for Lemons" (1970) – the first formal model of adverse selection.
  - In a used-car market, sellers know quality but buyers don't.
  - Only low-quality cars ("lemons") are traded in equilibrium – the market **unravels**.
  - Studied in Chapter 5.
- **Michael Spence** (Nobel 2001, jointly): education as a signal of ability.
  - High-ability workers signal type through costly education.
  - Studied in Chapter 5 (signaling model).

Adverse Selection – Spence and Stiglitz

- **Joseph Stiglitz** (Nobel 2001, jointly): screening in insurance markets – insurers design menus of contracts to separate types.
  - Studied in Chapters 5 and 6.

Incomplete Contracts: Hart and Moore

- **Incomplete contracts** (Hart and Moore, 1988, 1990; Grossman and Hart, 1986): a major development in contract theory.
- **Key insight:** in many real-world situations, it is *impossible* to write a complete contract specifying what should happen in every contingency.
  - Some contingencies are not foreseeable at the time of contracting.
  - Some actions or states of the world are observable but not *verifiable* (cannot be proven to a court).

Incomplete Contracts – Consequences and applications

- **Consequence:** parties must renegotiate when unforeseen contingencies arise, creating **hold-up problems**.
- **Applications:** vertical integration decisions, authority within firms, ownership rights, and governance structures.
- Hart received the Nobel Prize in 2016 specifically for this line of research.

Mechanism Design: Hurwicz, Maskin, and Myerson

- **Leonid Hurwicz**, **Eric Maskin**, and **Roger Myerson** were awarded the Nobel Prize in Economics in 2007 for laying the foundations of **mechanism design theory**.
- **Mechanism design** ("reverse game theory"): instead of analyzing the outcome of a given game, the designer *chooses* the rules of the game to achieve a desired social outcome.

Mechanism Design – Myerson's contributions

- **Myerson's contributions:**
  - Revelation principle: any outcome achievable with any mechanism can be achieved with a direct truthful mechanism.
  - Optimal auction design: revenue-maximizing mechanism for a seller facing buyers with private valuations.
- Studied in **Chapter 7** of this book.

Modern Developments in Contract Theory

- Contract theory continues to evolve, integrating insights from:
- **Behavioral economics:** agents may not be fully rational.
  - Present bias, loss aversion, and social preferences affect how contracts are designed and accepted.
  - Example: automatic enrollment in pension plans exploits present bias to increase savings.

Modern Developments – Dynamic and relational contracts

- **Machine learning and algorithmic contracts:**
  - Algorithmic tools can predict agent behavior, enabling more sophisticated performance monitoring.
  - Dynamic contracting with AI-based performance evaluation.
- **Dynamic contracts:** repeated interactions and reputation effects allow for richer incentive structures over time.
- **Relational contracts:** informal self-enforcing agreements sustained by the value of the relationship.

Timeline: Key Milestones in Contract Theory

- **1937**: Coase, "The Nature of the Firm" – transaction costs and the boundaries of the firm.
- **1970**: Akerlof, "The Market for Lemons" – adverse selection.
- **1973**: Spence, "Job Market Signaling" – signaling.
- **1975–79**: Mirrlees, Shavell, Holmström – formal moral hazard models.
- **1976**: Jensen & Meckling – agency theory.

Timeline: Key Milestones (cont.)

- **1979**: Holmström – informativeness principle; moral hazard and observability.
- **1986–90**: Grossman, Hart, Moore – incomplete contracts.
- **2001**: Nobel Prize to Akerlof, Spence, Stiglitz.
- **2007**: Nobel Prize to Hurwicz, Maskin, Myerson.
- **2016**: Nobel Prize to Hart and Holmström.

Section 1.4
Real-World Applications

Overview of Applications

- Contract theory finds direct application in many real-world settings. We focus on four major domains:
  1. **Labor markets**: performance pay, efficiency wages, stock options.
  2. **Health insurance**: co-pays, Value-Based Insurance Design, moral hazard.
  3. **Financial markets**: debt covenants, venture capital, securitization.
  4. **Public policy**: procurement auctions, public-private partnerships, regulatory design.
- We also examine four detailed **case studies**:
  - The 2007–2008 financial crisis.
  - The U.S. health insurance market in the 1990s.
  - Auto insurance in Israel (1975 reform).
  - The NFL Draft.

Application 1: Labor Markets

- Labor markets are perhaps the most natural application of principal-agent theory.
- The **moral hazard problem**: employees may exert suboptimal effort when their actions are not easily monitored by employers.

Application 1: Labor Markets – Contract design solutions

- **Contract design solutions:**
  - **Performance-based pay**: bonuses, piece rates, and commissions tied directly to measurable output.
  - **Stock options**: align manager and shareholder incentives by giving managers a stake in firm performance.
  - **Efficiency wages**: paying above-market wages to reduce shirking (a worker who loses this job loses a rent).
  - **Monitoring**: CCTV, productivity tracking software, performance reviews.

Labor Markets: Empirical Evidence

- A rich empirical literature documents moral hazard in labor markets:
- **Lazear (2000):** studied the introduction of piece rates at Safelite Glass; productivity increased by 44% and wages by 10%.
- **Bandiera, Barankay, Rasul (2007):** farm workers on individual pay schemes exerted 50% more effort than those on group schemes.

Labor Markets: Empirical Evidence – Monitoring technology

- **Jensen and Murphy (1990):** estimated that CEO pay was almost insensitive to firm performance – a puzzle for incentive theory.
- **Monitoring technology:** modern employers use call monitoring, keylogging, GPS tracking, and algorithmic performance scoring – all responses to the moral hazard problem.
- **Takeaway:** the type and structure of compensation matters significantly for worker behavior.

Application 2: Health Insurance

- Health insurance markets exhibit **both** moral hazard and adverse selection:
- **Moral hazard:** individuals with comprehensive insurance may seek unnecessary medical services or engage in riskier behaviors.
    - Example: low co-pays encourage overutilization.

Application 2: Health Insurance – Contract design

- **Adverse selection:** individuals with higher health risks are more likely to purchase insurance.
  - Example: only sick people buy health insurance $\Rightarrow$ premiums rise $\Rightarrow$ healthy people drop out $\Rightarrow$ death spiral.
- **Contract design solutions:**
  - Co-payments and deductibles: make consumers bear some cost, reducing moral hazard.
  - **Value-Based Insurance Design (VBID):** lower cost-sharing for high-value treatments (preventive care, chronic disease management) to encourage cost-effective use.

Health Insurance: Empirical Evidence

- **RAND Health Insurance Experiment (1984):** the most comprehensive study of moral hazard in health insurance.
  - Individuals with comprehensive insurance spent 20–30% more on healthcare than those with high deductibles.
  - The increase was largely in discretionary services, not emergency care.
- **Chernew et al. (2008):** high co-pays reduce utilization of healthcare services, confirming the moral hazard prediction.
- **VBID programs:** Gibson et al. (2011) find VBID interventions lead to reductions in adverse health events and long-term healthcare costs.

Health Insurance: Empirical Evidence (cont.)

- **Takeaway:** the design of insurance contracts has significant real consequences for healthcare utilization and health outcomes.

Application 3: Financial Markets

- Financial markets are a rich source of contract theory applications:
- **Debt contracts:**
  - Borrowers may take excessive risks, knowing that losses are borne by lenders (moral hazard – limited liability).
  - Collateral requirements and covenants are contract design responses.

Application 3: Financial Markets – Venture capital and securitization

- **Venture capital:**
  - Convertible preferred equity, staged financing, and board control rights all manage moral hazard and adverse selection between investors and entrepreneurs.
- **Securitization:**
  - Mortgage-backed securities transfer risk but also reduce the originator's incentive to screen borrowers carefully (moral hazard).
  - "Skin in the game" requirements (retain a fraction of the risk) are a contract design response.

Case Study 1: The 2007–2008 Financial Crisis

- The 2007–2008 financial crisis was fueled in part by severe moral hazard problems among large financial institutions.
- **"Too big to fail"**: institutions expected government bailouts in case of financial distress, reducing their incentive to manage risk prudently.

Case Study 1: Financial Crisis – Evidence and root cause

- **Evidence of moral hazard:**
  - Lehman Brothers increased its leverage ratio from 24:1 in 2003 to over 30:1 by 2007.
  - AIG sold over $440 billion in credit default swaps to insure risky securities, without sufficient reserves – required a $182 billion government bailout.
- **Root cause:** implicit guarantee of a government safety net created perverse incentives; banks prioritized short-term profits over long-term stability (Acharya et al., 2011; Rajan, 2010).

Case Study 1: Contract Theory Interpretation

- The financial crisis is a textbook principal-agent problem on a macroeconomic scale:
  - **Principal**: government and taxpayers.
  - **Agent**: bank managers and executives.
  - **Hidden action**: the level of risk-taking.
  - **Misaligned incentives**: managers received bonuses based on short-term profits, not long-term stability.
- **Policy responses (contract redesign):**
  - Dodd-Frank Act (2010): stricter capital requirements, reduced leverage, improved oversight.
  - Clawback provisions: bonuses can be recovered if risk materializes later.
  - Stress tests: force banks to reveal hidden information about their risk exposure.

Case Study 2: U.S. Health Insurance in the 1990s

- During the 1990s, the rise of managed care coincided with a surge in U.S. healthcare spending:
  - Health expenditure grew from \$696 billion (1990) to \$1.2 trillion (2000) – a 70% increase in one decade.
  - Average premiums for employer-sponsored insurance grew 7% annually.
- **The moral hazard problem:** insurers offered plans with very low out-of-pocket costs to attract customers, but this encouraged overutilization:
  - Insured individuals opted for unnecessary diagnostic tests and expensive treatments.
  - RAND experiment: individuals with comprehensive insurance spent 20–30% more on healthcare than those with high deductibles.
- **Contract design response:** managed care organizations introduced gatekeeping, prior authorization, and formulary restrictions to reduce moral hazard.

Case Study 3: Auto Insurance in Israel (1975)

- Israel introduced **mandatory automobile insurance** in 1975.
- Before the reform: drivers bore the full cost of damages they caused $\Rightarrow$ strong incentive for careful driving.
- **After the reform (Cohen and Einav, 2007):**
  - Traffic accidents increased by 10–15% in the five years following mandatory insurance.
  - Insurance claims surged; average premiums increased by 20% within a decade.

Case Study 3: Israel – Interpretation and design response

- **Interpretation:** mandatory insurance reduced the perceived financial responsibility for risky driving:
  - Drivers with comprehensive coverage were more likely to engage in speeding and reckless driving.
  - The insurer cannot observe driving behavior directly $\Rightarrow$ classic moral hazard.
- **Contract design response:** usage-based insurance (UBI) with telematics – monitor driving behavior and price accordingly.

Case Study 4: The NFL Draft

- The **NFL amateur player draft** provides an interesting application of adverse selection:
- **The information problem:** teams are not perfectly informed about all players' abilities, health, and character.
  - Players and their agents are better informed about the player's true talent than the teams.
- **Teams' response:** invest heavily in experts and scouts ("combine" evaluations, medical examinations, psychological testing) to reduce the information gap.

Case Study 4: The NFL Draft – Contract theory lens

- Despite these efforts, the system is not perfect (Wooten and Kruggel, 2024): teams can make suboptimal draft choices, selecting players who are not as promising as expected.
- **Contract theory lens:** the player's type (true ability) is private information $\Rightarrow$ adverse selection problem. Rookie contracts (fixed scale based on draft position) are a screening device that accounts for the uncertainty about player quality.

Application 4: Public Policy and Procurement

- Contract theory has important applications in **public policy**:
- **Procurement contracts:**
  - Governments hire contractors to provide goods and services (defense equipment, infrastructure, IT systems).
  - Moral hazard: contractors may cut corners or overcharge once the contract is signed.
  - Adverse selection: the government doesn't know which contractor is most efficient.
  - **Design tools:** competitive bidding (auctions), fixed-price vs. cost-plus contracts, performance benchmarks.

Application 4: Public Policy – Conservation auctions

- **Procurement auctions for conservation:**
  - Governments have used reverse auctions to protect biodiversity in the U.S. and Australia – paying landowners to preserve habitat at the lowest cost.
  - These are direct applications of mechanism design (Chapter 7).

Public-Private Partnerships (PPPs)

- **Public-private partnerships** are long-term contracts between a government and a private firm to deliver public infrastructure or services.
- Examples: toll roads, hospitals, schools, water treatment plants.
- **Contract theory challenges in PPPs:**
  - *Moral hazard:* the private firm may under-invest in maintenance or quality once the contract is signed (hidden action).
  - *Adverse selection:* the government doesn't know the firm's true costs when awarding the contract.
  - *Incomplete contracts:* it is impossible to specify quality for every possible contingency over a 25-year contract.

Public-Private Partnerships – Design responses

- **Design responses:** performance bonds, output-based specifications, benchmarking, and renegotiation clauses.

Behavioral Economics and Contract Theory

- Traditional contract theory assumes fully rational agents. **Behavioral economics** challenges this assumption.
- **Present bias:** agents overweight immediate costs and underweight future benefits.
  - Application: automatic enrollment in pension plans exploits present bias to increase retirement savings (Thaler and Sunstein, 2008).
  - Implication for contracts: optimal compensation may involve commitment devices (e.g., vesting schedules) to counteract present bias.

Behavioral Economics and Contract Theory – Social preferences

- **Loss aversion:** agents feel losses more strongly than equivalent gains (Kahneman and Tversky, 1979).
  - Implication: framing bonuses as "gains" vs. penalizing under-performance as "losses" affects effort levels differently.
- **Social preferences:** agents care about fairness and reciprocity, which can enhance or undermine incentive contracts.

Machine Learning and Algorithmic Contracts

- Recent technological advances are transforming contract design:
- **Algorithmic monitoring:**
  - Machine learning can predict worker productivity, enabling more precise performance measurement.
  - Reduces the information asymmetry that gives rise to moral hazard.
  - Examples: Uber's driver rating system, Amazon's warehouse productivity tracking, ride-sharing surge pricing algorithms.

Machine Learning and Algorithmic Contracts (cont.)

- **Algorithmic contracts:**
  - Smart contracts (blockchain) automatically execute when conditions are met, reducing enforcement costs.
  - Applications: escrow services, supply chain management, financial derivatives.
- **Challenge:** algorithmic systems can create new forms of moral hazard ("gaming" the algorithm rather than doing the underlying task).

Organization of the Book

Chapter 1 – Introduction to Contract Theory

- **This chapter.**
- Introduces the field, key concepts, and historical background.
- **Key concepts:** incentive compatibility, participation constraints, moral hazard, adverse selection, signaling.
- **Historical context:** from Hammurabi to Hart.
- **Applications:** labor, insurance, finance, public policy.
- **No prerequisites** beyond basic microeconomics.

Chapter 2 – Moral Hazard under Symmetric Information

- **Setting:** both principal and agent can observe the agent's effort – the benchmark case.
- **Key model:** principal $B(\pi - w)$; agent $u(w, e) = v(w) - g(e)$; conditional probabilities $f(\pi|e)$.
- **Key results:**
  - Optimal contract: equate marginal rates of substitution across outcomes.
  - Risk sharing: less risk-averse party bears more risk.
  - Full insurance for risk-averse agent when principal is risk-neutral.
  - Franchise contract when principal is risk-averse and agent is risk-neutral.
- **Extensions:** more than two outcomes; comparative statics of wages.

Chapter 3 – Moral Hazard under Asymmetric Information

- **Setting:** the principal cannot observe the agent's effort – the realistic case.
- **New constraint:** incentive compatibility (IC) is added to the participation constraint (PC).
- **Key results:**
  - Under asymmetric information, the agent no longer receives full insurance.
  - Wages must vary with output to provide effort incentives.
  - The trade-off between risk-sharing and incentive provision is central.
  - Monotone likelihood ratio property (MLRP): optimal wages are increasing in output.
- The informativeness principle: which signals should appear in the contract?

Chapter 4 – Moral Hazard with Continuous Effort

- **Setting:** continuous effort $e \in [0, \infty)$; linear contracts $w(x) = a + bx$.
- **Key technique:** the first-order approach – replaces the IC constraint with the agent's first-order condition.
- **Key results (CARA utility, normal output):**
  - Optimal bonus: $b^* = \dfrac{1}{1 + \rho\sigma^2}$ – decreasing in risk aversion ρ and output variance σ^2.
  - Equilibrium effort: $e^* = b^*$.

Chapter 4 – Continuous Effort – Extensions

- **Extensions:**
  - Performance vs. output (β parameter).
  - Informative signals: the informativeness principle ($\alpha^* = -\sigma_{\varepsilon s}/\sigma_s^2$).
  - Multi-task moral hazard (Holmström-Milgrom).

Chapter 5 – Adverse Selection: Screening and Signaling

- **Setting:** the agent has private information (type) before signing the contract.
- **Lemons problem (Akerlof 1970):** adverse selection leads to market unraveling; threshold probability $\bar{q} = \dfrac{s_H - b_L}{b_H - b_L}$.
- **Screening:** the principal designs a menu of contracts to separate agent types.
- **Signaling (Spence model):** the agent sends a costly signal to reveal type.
  - Separating PBE: (E^H, NE^L).
  - Pooling PBE: (NE^H, NE^L).
  - PBE refinement: equilibrium dominance / intuitive criterion.
- Warranties, quality certificates, and labor market applications.

Chapter 6 – Screening

- **Setting:** integrates moral hazard and adverse selection; agent type θ affects the cost of effort.
- **Model:** profit $\pi(e) = \log(e) - w$; agent utility $u(w, e) = w - \theta e^2$; types $\theta_L < \theta_H$.
- **Symmetric information:** $e_i^{SI} = \left(\frac{1}{2\theta_i}\right)^{1/2}$; wage $w_i^{SI} = \frac{1}{2}$.
- **Asymmetric information:**
  - No distortion at top: $e_H^{AI} = e_H^{SI}$.
  - Downward distortion at bottom: $e_L^{AI} < e_L^{SI}$.
  - Information rent: $u_L^{AI} = (\theta_H - \theta_L)c(e_H) > 0$.
- **Extensions:** single-crossing property, non-linear pricing, product quality screening.

Chapter 7 – Mechanism Design

- **Setting:** the designer chooses the rules of the game to achieve a desired social outcome.
- **Key concepts:**
  - Direct vs. indirect revelation mechanisms (DRM vs. IRM).
  - BIC (Bayesian IC) and DSIC (dominant strategy IC).
  - Revelation principle: any BNE outcome is achievable with a BIC-DRM.
- **Key results:**
  - Groves mechanism: achieves allocative efficiency in dominant strategies; transfer $t_i = \sum_{j \neq i} u_j + h_i(\theta_{-i})$.
  - VCG mechanism: a special case; second-price auction.
  - Myerson's theorem: optimal Bayesian mechanism and virtual valuations.
  - Revenue equivalence theorem.

Summary

Summary – Key Concepts

- **Contract theory** studies how to design agreements between parties with different information and conflicting interests.
- **Two core principles:**
  - Incentive compatibility: the agent finds it optimal to behave as the principal desires.
  - Participation constraint: all parties willingly join the agreement.
- **Two main information problems:**
  - Moral hazard (hidden action): studied in Chapters 2–4.
  - Adverse selection (hidden type): studied in Chapters 5–6.
- **Mechanism design** (Chapter 7): designing the rules of the game to achieve a desired social outcome.

Summary – Historical Foundations

- The challenges addressed by contract theory – enforcement, information, risk-sharing – have been recognized since antiquity.
- **Historical milestones:**
  - Code of Hammurabi (c. 1754 BCE): employment, loans, trade.
  - Roman law: consensus ad idem, contract types.
  - Medieval Law Merchant: cross-border trade contracts.
  - Enlightenment: philosophical foundations of contractual obligation.
- **Intellectual history milestones:**
  - Coase (1937): transaction costs and the firm.
  - Akerlof (1970), Spence (1973): lemons problem and signaling.
  - Holmström (1979), Hart (1986): moral hazard and incomplete contracts.
  - Nobel Prizes: 2001, 2007, 2016.

Summary – Real-World Relevance

- Contract theory is not merely academic – it has shaped policy and practice across many domains:
- **Labor markets:** performance pay and efficiency wages address moral hazard.
- **Health insurance:** VBID and co-pays balance coverage and moral hazard.
- **Financial markets:** covenants, collateral, and "skin in the game" requirements address hidden risk-taking.
- **Public policy:** procurement auctions and PPP design use mechanism design principles.
- **Looking ahead:** behavioral insights and machine learning are transforming how contracts are designed and enforced.

Key Takeaways

1. Contract theory sits at the intersection of microeconomics, game theory, and information economics.
2. The two pillars of any contract: **incentive compatibility** and **participation constraints**.
3. Information asymmetries generate **inefficiency**: moral hazard leads to insufficient effort; adverse selection leads to market failure or information rents.
4. The fundamental trade-off in contract design: **risk-sharing** vs. **incentive provision**.
5. From Mesopotamia to machine learning: the challenges of contract design are timeless, even if the tools to address them continue to evolve.

End of Chapter 1

Chapter 2

Moral Hazard under Symmetric Information

Instructor Slides

Contract Theory: An Introduction with Applications

Ana Espinola-Arredondo, Felix Muñoz-Garcia, and Pak-Sing Choi

Outline of Chapter 2

- This chapter analyzes the moral hazard problem under **symmetric information** – both principal and agent observe the agent's effort.
- Sections:
  - 2.1 Introduction – examples, empirical evidence, and case studies.
  - 2.2 Model – principal, agent, and conditional probabilities.
  - 2.3 Two efforts and two outcomes – the benchmark.
  - 2.4 Risk attitudes – three cases.
  - 2.5 More than two outcomes – general framework.
  - 2.6 Summary. 2.7 Exercises.

What you will learn

- The **moral hazard problem**: definition and real-world examples.
- How to model the principal's and agent's preferences using benefit and utility functions.
- How to use **conditional probabilities** $f(\pi|e)$ to link effort to profits.
- The **symmetric information benchmark**: when effort is observable, the optimal contract solves a constrained maximization with only the participation constraint.
- The **MRS condition**: a key efficiency property.
- How **risk attitudes** of the principal and agent determine the shape of the optimal contract.

Section 2.1
Introduction

What this chapter studies

- We analyze the **moral hazard problem** in a framework of **symmetric information**.
- Both parties – the principal (employer, insurer) and the agent (employee, policyholder) – have full visibility of relevant variables.
- Even with symmetric information, contract design must address **incentive misalignments** between principal and agent.

What this chapter studies (cont.)

- The chapter:
  - Provides real-world examples and case studies.
  - Models principal-agent interactions formally.
  - Derives the optimal contract under three risk-attitude configurations.
  - Extends the analysis to settings with more than two outcomes.

What is moral hazard?

- **Moral hazard**: a situation in which one party (the agent) has an incentive to take risks or shirk responsibilities because another party (the principal) bears the consequences.
- Arises whenever:
  - One party's actions are not perfectly observable or verifiable, OR
  - One party's actions affect a payoff that is partially borne by another.
- **Key insight:** the agent's optimal behavior often differs from what would be best for the principal – creating an **incentive misalignment**.
- Even under symmetric information, moral hazard issues emerge because the principal and agent have *different preferences* (e.g., risk attitudes, disutility of effort).

2.1.1 Examples of moral hazard

Example 1 – Labor markets

- The canonical example of moral hazard: employees may shirk their responsibilities or exert suboptimal effort.
- Why? Because their actions are not easily monitored or verified by employers.

Example 1 – Labor markets (cont.)

- **Employer responses:**
  - Performance-based incentives (commissions, bonuses, stock options).
  - Monitoring technologies (CCTV, productivity tracking software, time clocks).
  - Periodic performance reviews and ratings.
- **Key trade-off:** stronger incentives boost effort but expose the worker to more risk.
- This labor-market example is the running example throughout this chapter.

Example 2 – Health insurance

- Healthcare plans with comprehensive coverage and low cost-sharing (low co-pays) induce moral hazard.
- **Consequences:**
  - Patients seek unnecessary medical services.
  - Overconsumption of healthcare resources.
  - Higher healthcare expenditures and higher insurance premiums.

Example 2 – Health insurance (cont.)

- **Insurer's response: Value-Based Insurance Design (VBID)**
  - Lower co-pays for high-value treatments (preventive care, chronic disease management).
  - Higher co-pays for low-value services.
  - Encourages cost-effective care utilization, partially mitigating moral hazard.

Example 3 – Financial markets

- Borrowers may take excessive risk because they anticipate bailouts or government support in case of default.
- Creates moral hazard problems for lenders and investors.

Example 3 – Financial markets (cont.)

- **Regulatory responses:**
  - Capital requirements (e.g., Basel III).
  - Credit scoring models to screen borrowers.
  - Central clearinghouses and margin requirements for derivatives.
- **Special case – derivative markets:**
  - Credit default swaps and collateralized debt obligations can incentivize speculative behavior.
  - Excessive risk-taking is exacerbated by complex financial instruments.

2.1.2 Empirical studies

Empirical evidence – Health insurance

- U.S. studies confirm the theoretical prediction:
  - **High co-pays reduce utilization** of healthcare services (Chernew et al., 2008).
  - Magnitude depends on service type, demographics, and health conditions (Cutler & Zeckhauser, 2000).
- **Effectiveness of VBID:**
  - VBID incentivizes preventive care, chronic disease management, and evidence-based treatments (Gibson et al., 2011).
  - Leads to reductions in adverse health events, hospitalizations, and long-term healthcare costs (Fendrick et al., 2013).
- Recent studies extend the analysis to health disparities, equity, and healthcare delivery models (Wharam et al., 2018; Maciejewski et al., 2019).

Empirical evidence – Financial markets

- **Banking sector findings:**
  - Implicit government guarantees (deposit insurance, bailout expectations) induce excessive risk-taking (Acharya & Richardson, 2009; Kashyap et al., 2009).
  - "Too-big-to-fail" policies amplify moral hazard (Mishkin, 2011).
  - Technological advancements affect risk-taking (Adrian & Ashcraft, 2016; Demirgüç-Kunt et al., 2018).
- **Derivative markets:**
  - Systemic risk is affected by financial innovation (Blundell-Wignall & Atkinson, 2010).
  - CDS and CDOs amplify systemic risk during market stress (Greenwood & Hanson, 2013; Acharya et al., 2010).
- Recent work: governance, compensation, and risk culture in banking (Bolton et al., 2020).

2.1.3 Case studies

Case Study 1 – 2007–2008 Financial Crisis

- Fueled by large financial institutions originating, securitizing, and trading subprime mortgage-backed securities.
- These institutions acted under the assumption that they were "**too big to fail**" – anticipating government bailouts.

Case Study 1 – Lehman and AIG

- **Lehman Brothers:**
  - Increased leverage ratio (assets/equity) from **24:1 in 2003** to over **30:1 by 2007**.
  - Declared bankruptcy in September 2008 with over $619 billion in debt – largest bankruptcy filing in U.S. history.
- **AIG (American International Group):**
  - Sold over $440 billion in credit default swaps (CDS) by 2008.
  - Required a $182 billion U.S. government bailout to prevent collapse.

Case Study 1 (cont.) – Perverse incentives

- The implicit guarantee of a "government safety net" created **perverse incentives**.
- Banks and insurance companies prioritized short-term profits over long-term stability.
- **Empirical evidence:**
  - Acharya et al. (2009): implicit guarantees exacerbated risk-taking among large institutions.
  - Rajan (2010): executives rewarded based on immediate financial performance, ignoring systemic risk.
- **Read as a moral-hazard problem:** the government (principal) cannot verify managers' risk choices, and the bailout expectation weakens the link between losses and the manager's own payoff.

Case Study 1 (cont.) – The Dodd-Frank response

- **Post-crisis reform – Dodd-Frank Act (2010):**
  - Stricter capital requirements.
  - Reduced leverage limits.
  - Improved oversight and stress testing.

Case Study 1 – Contract-design reading of each measure

- **Contract-design reading of each measure:**
  - Capital requirements $\rightarrow$ force managers to keep "skin in the game"; a loss now costs the bank's own equity.
  - Leverage limits $\rightarrow$ cap the size of the hidden bet.
  - Stress tests $\rightarrow$ generate a signal that reduces the principal's information asymmetry.
- Questions remain about whether these measures are strong enough – an open policy debate.

Case Study 2 – U.S. Health Insurance (1990s)

- Rise of managed care systems in the 1990s coincided with surging healthcare costs.
- Insurers offered policies with very low out-of-pocket costs to attract customers.

Case Study 2 – The moral-hazard problem

- These plans encouraged **overutilization**:
  - Patients opted for unnecessary diagnostic tests.
  - Patients chose expensive treatments when cheaper alternatives existed.
- **Key numbers:**
  - U.S. health expenditure: \$696 billion (1990) $\rightarrow$ \$1.2 trillion (2000), a **70% increase** in one decade.
  - Premiums for employer-sponsored insurance grew on average 7% annually.

Case Study 2 (cont.) – RAND and the moral-hazard reading

- **RAND Health Insurance Experiment (1984):** individuals with comprehensive coverage spent **20–30% more** on healthcare than those with high deductibles.
  - Random assignment of coverage rules $\rightarrow$ the spending gap comes from behavioral response, not sicker enrollees.
  - Most of the extra spending was on discretionary care (specialist visits, non-emergency tests), not on life-saving treatment.

Case Study 2 – Contract-theory reading

- **Contract-theory reading:**
  - The insurer (principal) cannot verify whether a given visit was medically needed.
  - Low co-pays make careful use unrewarding for the insured; expected insurer payouts rise.
  - Managed care, deductibles, and prior authorization are the principal's response – imperfect proxies for the hidden action.

Case Study 3 – Auto Insurance Israel (1975)

- Israel introduced **mandatory automobile insurance in 1975**.
- **Before the reform:** drivers bore the full cost of damages they caused $\Rightarrow$ careful driving incentivized.

Case Study 3 – Cohen and Einav (2007)

- **After the reform (Cohen and Einav, 2007):**
  - Traffic accidents increased **10–15%** in the 5 years following the reform.
  - Average premiums increased **20%** within a decade.
- The reform is a natural experiment – coverage went from a self-selected choice to a universal rule, so the increase in accidents must come from the change in incentives rather than from riskier drivers buying insurance.

Case Study 3 (cont.) – Contract-theory reading

- Mandatory insurance reduced the perceived financial responsibility for risky driving.
- Drivers with comprehensive coverage were more likely to engage in speeding and reckless driving.
- Insurers cannot observe driving behavior directly – a classic moral-hazard setting.

Case Study 3 – Design responses seen today

- **Design responses seen in the market today:**
  - Deductibles and experience-rated premiums make each accident costly to the driver.
  - Usage-based insurance (telematics) partially converts a hidden action into an observed signal – braking, speed, mileage.
  - No-claims bonuses reward the unobserved careful driving with a lower future premium.

Case Study 4 – Tanking in professional sports

- Leagues such as the NBA and NFL award the earliest draft picks to the worst-performing teams, to preserve competitive balance.
- **The incentive problem:**
  - Once a team is eliminated from playoff contention, the value of winning another game falls near zero, while losing improves its draft position.
  - Deliberately underperforming to secure a better pick is known as "tanking."
- **A clean case of hidden action:** the league (principal) wants high effort in every game, but effort is not contractible – resting healthy starters or managing injuries always has a sporting justification.

Case Study 4 (cont.) – The league's design response

- **League response:** weaken the link between losing and reward.
  - NBA **draft lottery** (introduced 1985): the worst record no longer guarantees the first pick.
  - NBA **flattened odds** (2019): the three worst teams now share identical top-pick odds ($\approx 14\%$ each), further cutting the return to losing.

Case Study 4 – The contract-theory reading

- **Contract-theory reading:**
  - The old rule paid a big prize (top pick) for a signal (last place) that was easy to game through hidden effort.
  - Randomization breaks the deterministic reward for underperforming – a mechanism-design fix, closer in spirit to the tools we develop in Chapter 7.
- Same logic applies elsewhere: performance quotas that trigger promotions or bonuses can invite the analog of tanking whenever the threshold rule is deterministic.

Section 2.2
Model

Model overview

- We consider an interaction between:
  - **Principal** – the manager of a firm.
  - **Agent** – a worker the manager hires.
- The worker exerts **effort** $e \geq 0$ on the job; effort affects the firm's profit π.
- The principal pays the worker a **wage** $w \geq 0$.
- Read e as the hours or intensity the worker supplies, and w as the price the principal pays for the profit that effort makes more likely. The tension is that effort is costly to the worker but valuable to the principal.

Model overview – Timing

- Under **symmetric information**, the principal can observe the worker's effort e – this is the benchmark. In Chapter 3 we drop this observability assumption.
- **Timing.** The principal (1) proposes a contract (e, w); (2) the agent accepts or rejects (PC); (3) if accepted, the agent chooses effort; (4) profit π realizes and the wage is paid.
- Because the principal moves first and effort is observable here, she can make the wage contingent on the effort she requires, so step (3) holds no surprises. This is exactly what breaks down in Chapter 3.

Model overview (cont.) – Three modeling ingredients

- Three modeling components:
  1. Principal's benefit function $B(\pi - w)$.
  2. Agent's utility function $u(w, e)$.
  3. Conditional probability $f(\pi|e)$ linking effort to profits.
- The first two components describe *preferences* (how each party values money and effort); the third describes the *technology* (how effort maps into profit). The optimal contract comes from combining them.

2.2.1 The Principal's benefit function

- The principal (manager) values the contract through a **benefit function**

$$B(\underbrace{\pi - w}_{\text{net return}}) \tag{2.1}$$

where $\pi \geq 0$ is the profit the worker's effort generates and $w \geq 0$ is the wage she pays him. She keeps π and hands over w, so her *net return* is $\pi - w$.

- **Most common assumption – risk-neutral principal:**

$$B(\pi - w) = \pi - w,$$

so B is linear in her net return.

- Intuitively, a risk-neutral principal ranks contracts by expected net profit alone, ignoring how much that profit fluctuates across outcomes.

The benefit function – Risk attitudes

- More generally, $B(\pi - w) = (\pi - w)^{\theta}$, where θ sets the curvature of B:
  - $\theta \in [0, 1)$: **risk-averse**, B concave in the net return.
  - $\theta = 1$: **risk-neutral**, B linear.
  - $\theta > 1$: **risk-loving**, B convex.
- The curvature is what matters: a concave B means the principal dislikes swings in her net return, while a linear B means she is indifferent to them.
- Economically, θ measures how much the principal cares about risk in profits, not just their average level – the same role the concavity of $v(w)$ will play for the worker.

2.2.2 The Agent's utility function

- The agent (worker) has utility function

$$u(w, e) = \underbrace{v(w)}_{\text{utility from wage}} - \underbrace{g(e)}_{\text{disutility of effort}}, \qquad (2.2)$$

  where $w \geq 0$ is the wage and $e \geq 0$ is the effort.
- Reading the two terms:
  - $v(w)$: the benefit the worker gets from being paid, rising in w.
  - $g(e)$: the cost he bears from exerting effort, which he subtracts.
- So the worker trades off a larger paycheck against the effort it takes to earn it, and signs the contract only if the balance clears his outside option.

The Agent's utility – Separability

- The utility in (2.2) is **separable** in wage and effort: it adds a wage term and an effort term rather than mixing them in a product.
- Two consequences we use repeatedly:
  - The marginal utility of a higher wage, $v'(w)$, does not depend on how hard the worker is working, e.
  - The marginal disutility of effort, $g'(e)$, does not depend on how much he is paid, w.
- Intuitively, separability keeps the two margins independent, so we can study the wage and the effort decisions one at a time.

Utility from wages, $v(w)$

- Properties of $v(w)$: $v'(w) > 0$ and $v''(w) \leq 0$.
  - $v'(w) > 0$: more pay is always better.
  - $v''(w) \leq 0$: each extra dollar adds a bit less utility than the last – diminishing marginal utility of income.
- Examples: $v(w) = \sqrt{w}$ and $v(w) = w^{1/3}$; more generally

$$v(w) = Aw^{1-\alpha}, \quad A > 0, \ \alpha \in [0, 1].$$

- Economically, the weak concavity of v is exactly the worker's risk aversion: a concave v makes him prefer a sure wage to a fair gamble over wages.

Utility from wages – Reading $v(w)$

- Risk-aversion parameter α:
  - $\alpha = 0$: **risk-neutral** worker, $v(w) = Aw$.
  - $\alpha = \frac{1}{2}$: **risk-averse**, $v(w) = A\sqrt{w}$.
  - $\alpha = 1$: **extremely risk-averse**, $v(w) = Aw^0 = A$ – unaffected by wage offers.
- So α is a dial for risk aversion: at $\alpha = 0$ the worker values money linearly, and as α rises toward 1 the utility function bends more sharply, so a sure wage becomes worth more to him than a risky one with the same mean.

Disutility from effort, $g(e)$

- Properties of $g(e)$: $g'(e) > 0$ and $g''(e) \geq 0$.
  - $g'(e) > 0$: more effort is more costly for the worker.
  - $g''(e) \geq 0$: the cost rises at a weakly increasing rate – each extra hour hurts a little more than the previous one.
- Captures the disutility from working longer hours, or from a more intense effort during the same shift.
- Examples: $g(e) = e$ (constant marginal cost) and $g(e) = e^2$ (increasing marginal cost).

Disutility from effort – Reading $g(e)$

- More generally:

$$g(e) = \kappa e^{\beta}, \quad \kappa > 0, \ \beta \geq 1,$$

  where we write the constant as κ, not B, to avoid clashing with the principal's benefit B.

- Convexity parameter β:
  - $\beta = 1$: linear cost, $g(e) = \kappa e$, so marginal cost is constant.
  - $\beta = 2$: quadratic cost, $g(e) = \kappa e^2$, so marginal cost rises with e.
  - Higher β means stronger convexity and a faster-rising marginal disutility of effort.
- Economically, κ shifts the whole cost up or down, while β controls how quickly extra effort becomes painful.

2.2.3 How does effort affect profits?

- Effort affects the **probability distribution** of profits.
- Otherwise, profits would be unaffected by effort and the principal would not care about how much the worker exerts!
- We model this with the **conditional probability**

$$f(\pi_i|e) = \Pr\{\pi = \pi_i|e\} \geq 0,$$

  which gives the probability that profit equals π_i when the worker exerts e.
- **Example:** when the worker exerts 3 hours of effort, the firm earns \$5,000 with probability $f(\$5{,}000|3) = \frac{1}{4}$.
- Intuitively, higher effort does not fix the profit; it tilts the odds toward better profit realizations. That is the only channel through which effort matters to the principal.

Table 2.1 – Two efforts, three profits

- Suppose two effort levels ($e_H > e_L$) and three profit levels ($\pi_H > \pi_M > \pi_L$):

		Profits		
		π_L	π_M	π_H
Effort	e_H	$f(\pi_L \mid e_H)$	$f(\pi_M \mid e_H)$	$f(\pi_H \mid e_H)$
	e_L	$f(\pi_L \mid e_L)$	$f(\pi_M \mid e_L)$	$f(\pi_H \mid e_L)$

Table 2.1. Conditional probabilities with two efforts and three profits.

Table 2.2 – Two numerical examples

Effort		Profits		
		π_L	π_M	π_H
	e_H	0.1	0.3	0.6
	e_L	0.6	0.3	0.1

Effort		Profits		
		π_L	π_M	π_H
	e_H	0.1	0.3	0.6
	e_L	0.5	0.5	0

- **Left panel:** all profits have positive probability; increasing effort shifts probability from π_L to π_H.

Table 2.2 – Two numerical examples (cont.)

- **Right panel:** some events have zero probability; increasing effort shifts probability away from π_L and π_M.

Properties of conditional probabilities

- Conditional probabilities $f(\pi_i|e)$ must satisfy two standard properties.
- **Property 1: Range.**

$$f(\pi_i|e) \in [0, 1],$$

because each $f(\pi_i|e)$ is a probability.

  - $f(\pi_i|e) = 0$: profit π_i is impossible at effort e.
  - $f(\pi_i|e) = 1$: profit π_i is certain at effort e.

- So every cell of the probability table is a number between 0 and 1, with 0 and 1 as the limiting certain cases.

Properties of conditional probabilities (cont.)

- **Property 2: Sum to one.**

$$\sum_{i=1}^{N} f(\pi_i|e) = 1,$$

because, fixing effort e, some profit realization must occur.

  - Across all N possible profit realizations, the probabilities add to 1.
  - For two outcomes ($N = 2$): $f(\pi_H|e) + f(\pi_L|e) = 1$.

- Together the two properties say each *row* of the probability table (one effort level) is a genuine distribution over profits: nonnegative entries that sum to one.

Example 2.1 – Two efforts, two outcomes

- Consider $e_H = 5$, $e_L = 0$; $\pi_H = \$1{,}200$, $\pi_L = \$300$.
- Conditional probabilities:
  - High effort: $p_H = f(\pi_H|e_H) = 0.5$, $1 - p_H = f(\pi_L|e_H) = 0.5$.
  - Low effort: $p_L = f(\pi_H|e_L) = 0.25$, $1 - p_L = f(\pi_L|e_L) = 0.75$.

		Profits	
		π_L	π_H
Effort	e_H	0.5	0.5
	e_L	0.75	0.25

Table 2.3. Conditional probabilities – Example 2.1.

- Reading the table: higher effort lifts the chance of π_H from 0.25 to 0.5, shifting probability toward the good outcome – the channel the principal pays for.

Compact notation – Two efforts, two outcomes

- For each profit level i:
  - It occurs with probability p_i when the worker exerts effort i.
  - It occurs with probability $1 - p_i$ when the worker exerts the opposite effort.

		Profits	
		π_L	π_H
Effort	e_H	$1 - p_L$	p_H
	e_L	p_L	$1 - p_H$

Table 2.3b. Two efforts and two outcomes – General setting.

- For settings with $K > 2$ effort levels or $N > 2$ profit realizations, we use the more explicit notation $f(\pi_i|e)$.

Section 2.3
Two Efforts and Two Outcomes

Symmetric information benchmark

- Both principal and agent observe the agent's effort e.
- This is the **benchmark** – in Chapter 3 we relax this assumption.
- Why study it?
  - To establish the **first-best** outcome (fully efficient contract).
  - To highlight the role of **risk-sharing** when there are no incentive constraints.
  - To provide a reference for the second-best contracts under asymmetric information.

Symmetric information benchmark – Modern relevance

- Even with modern monitoring (cameras, log-ins, productivity tracking), observation is rarely perfect – the principal still cannot see the worker's concentration or focus on the task at hand.
- Still, symmetric information is a useful approximation whenever output tracks effort closely, as in piece-rate manufacturing or commission sales.
- Treat it as the **first-best benchmark**: it tells us the best a contract could do if effort were fully observable, and Chapter 3 measures the cost of losing that observability.

The principal's problem

- Under symmetric information, the principal solves:

$$\max_{e, w_H, w_L \geq 0} \quad p_H B(\pi_H - w_H) + (1 - p_H) B(\pi_L - w_L)$$

$$\text{s.t.} \quad p_H v(w_H) + (1 - p_H) v(w_L) - g(e) \geq \bar{u}, \quad \text{(PC)}$$

  where $p_H \equiv f(\pi_H | e)$ and $1 - p_H \equiv f(\pi_L | e)$ is shorthand for the two outcome probabilities at the induced effort.

- The principal chooses:
  - Effort e she induces.
  - Wages (w_H, w_L) paid in each outcome.
- **Three choice variables**, one constraint (PC). There is only one constraint because effort is observable, so the principal need not worry about inducing it – we return to this next.

Reading the principal's problem

- The objective is the principal's **expected profit**. In each outcome i she earns the profit net of the wage she pays, $\pi_i - w_i$, and her benefit from that net return is $B(\pi_i - w_i)$:

$$\underbrace{p_H}_{\text{prob. } p_H}\ \underbrace{B(\pi_H - w_H)}_{\text{benefit if } \pi_H} + \underbrace{(1-p_H)}_{\text{prob. } 1-p_H}\underbrace{B(\pi_L - w_L)}_{\text{benefit if } \pi_L}.$$

- She weighs each benefit by how likely that outcome is, p_H or $1 - p_H$, so the objective is an **expectation** taken over the two profit realizations.
- The constraint (PC) says the worker must find the contract at least as good as his outside option $\overline{u}$; otherwise he walks away.

Why only PC? – A note on IC

- Under **symmetric information**, the principal observes effort e and can write the contract so the wage is paid *only* when the observed effort matches the required one.
- So the principal does not need to worry about *how* to induce a given effort – effort is contractible.
- That is why the problem carries only the **participation constraint** (PC).
- Example: the contract can read "exert e_H, and you are paid w_H or w_L; exert less and you are paid nothing." Since the court can verify e, this clause is enforceable, so the principal simply picks the effort she wants.

Why only PC? – Comparison with Chapter 3

- In Chapter 3 the principal cannot observe e, so a second constraint enters – the **incentive compatibility** (IC) constraint. Contrast:
  - Symmetric information: max s.t. PC.
  - Asymmetric information: max s.t. PC and IC.
- The two problems differ only in one constraint – but that single constraint drives every result in the next chapter.
- Keep this contrast in mind: whenever a Chapter 2 result "breaks" in Chapter 3, the IC constraint is the reason.

The participation constraint (PC)

- **PC:**

$$p_H v(w_H) + p_L v(w_L) - g(e) \geq \overline{u}$$

- **Left-hand side:** the worker's expected utility from the contract – expected wage utility minus disutility of effort.
- **Right-hand side: reservation utility** $\overline{u}$ – utility from the worker's outside option (alternative jobs, unemployment benefits, leisure).
- Interpretation: the worker accepts only if the contract is at least as good as his outside option; if PC is violated, he rejects it.
- At the optimum PC holds with *equality*: any slack would let the principal shave wages and raise her own benefit, as we confirm next.

The Lagrangian

- Setting up the Lagrangian:

$$\mathcal{L} = \underbrace{p_H B(\pi_H - w_H) + p_L B(\pi_L - w_L)}_{\text{Objective}} + \lambda \underbrace{[p_H v(w_H) + p_L v(w_L) - g(e) - \overline{u}]}_{\text{PC}} \tag{2.3}$$

  where $\lambda \geq 0$ is the Lagrange multiplier on the PC. We attach λ to the constraint and add it to the objective, so that setting the derivatives of $\mathcal{L}$ to zero gives the optimum.

- Two cases:
  - $\lambda > 0$: PC **binds** – the worker is just indifferent between accepting and rejecting.
  - $\lambda = 0$: PC is slack – the worker strictly prefers to accept.
- We will show $\lambda > 0$: the principal optimally drives PC to equality, leaving no slack.

First-order conditions in wages

- Differentiating (2.3) with respect to w_H and w_L (the chain rule puts a -1 on each B', since $\partial(\pi_i - w_i)/\partial w_i = -1$):

$$\begin{aligned} p_H B'(\pi_H - w_H) \times (-1) + \lambda p_H v'(w_H) &= 0, \\ p_L B'(\pi_L - w_L) \times (-1) + \lambda p_L v'(w_L) &= 0. \end{aligned}$$

- Cancel $p_H, p_L > 0$ (each multiplies its whole equation) and move the B' term across:

$$\begin{aligned} \lambda v'(w_H) &= B'(\pi_H - w_H), \\ \lambda v'(w_L) &= B'(\pi_L - w_L). \end{aligned}$$

- Divide each by $v'(w_i) > 0$ to isolate λ:

$$\lambda = \frac{B'(\pi_H - w_H)}{v'(w_H)}, \qquad \lambda = \frac{B'(\pi_L - w_L)}{v'(w_L)}.$$

Solving the FOCs – Step by step

- **Step 1.** Differentiate $\mathcal{L}$ with respect to w_H. The chain rule turns $B(\pi_H - w_H)$ into $B'(\pi_H - w_H) \times (-1)$ – the -1 is the derivative of the argument.
- **Step 2.** The PC term contributes $\lambda p_H v'(w_H)$; setting the sum to zero gives

$$-p_H B'(\pi_H - w_H) + \lambda p_H v'(w_H) = 0.$$

- **Step 3.** Cancel $p_H > 0$ (probabilities of realized outcomes never enter the wage choice symmetrically – they cancel from every wage FOC).

Solving the FOCs – Cancelling probabilities

- **Step 4.** Solve for λ to get one expression per outcome; equating them yields the MRS condition on the next slide.
- **Common student mistake.** Writing $B'(\pi_H - w_H) = -1$ (forgetting that the -1 comes from the chain rule, not from B').

The MRS condition (Eq. 2.4)

- Both FOCs equal the same multiplier λ, so they equal each other:

$$\lambda = \frac{B'(\pi_H - w_H)}{v'(w_H)} = \frac{B'(\pi_L - w_L)}{v'(w_L)}. \tag{2.4}$$

- Since $B' > 0$ and $v' > 0$, the ratio is positive, so $\lambda > 0 \Rightarrow$ **PC binds**.
- Define the **marginal rate of substitution** (MRS) in each outcome:

$$MRS_H \equiv \frac{B'(\pi_H - w_H)}{v'(w_H)}, \qquad MRS_L \equiv \frac{B'(\pi_L - w_L)}{v'(w_L)}.$$

- Then (2.4) becomes compactly:

$$MRS_H = MRS_L \tag{2.5}$$

Intuition of $MRS_H = MRS_L$

- MRS_i measures the ratio of marginal utilities from w_i between principal (numerator) and agent (denominator) in outcome i.
- **Adjustment mechanism:**
  - If $MRS_H > MRS_L$: principal is more affected by a marginal increase in w_H than the agent benefits.
  - Optimal to **decrease** w_H and **increase** w_L.
  - This change *decreases* MRS_H and *increases* MRS_L until they coincide.
- Opposite if $MRS_H < MRS_L$.

Intuition of $MRS_H = MRS_L$ (cont.)

- At the optimum the two marginal rates coincide, so the principal has no further incentive to reallocate wages across outcomes: shifting a dollar from w_H to w_L (or back) no longer raises her expected benefit.
- This mirrors the standard consumer tangency condition, where the MRS is equated across goods; here the "goods" are the wages paid in the high- and low-profit states.
- Economically, the condition is about *how* the wage bill is split across outcomes, not its total size – the PC fixes the total, while (2.5) fixes the split.

A numerical check of $MRS_H = MRS_L$

- Take a simple parametrization: $B(x) = x$ (risk-neutral principal) and $v(w) = \sqrt{w}$ (risk-averse agent).
- Then $B'(\cdot) = 1$ and $v'(w) = \dfrac{1}{2\sqrt{w}}$, so the MRS in outcome i is

$$MRS_i = \frac{1}{1/(2\sqrt{w_i})} = 2\sqrt{w_i}.$$

- Setting $MRS_H = MRS_L$ gives $2\sqrt{w_H} = 2\sqrt{w_L}$, which yields $w_H = w_L$.
- So the MRS condition already delivers **full insurance** under Case 1 – a preview of the result we prove formally on the next slides. The condition is doing real work, not just book-keeping.

Section 2.4
Risk Attitudes

Three cases

- We examine how equation (2.5) is affected by risk attitudes:
- **Case 1:** Principal is **risk-neutral**, agent is **risk-averse**.
  - The standard textbook setting.
- **Case 2:** Principal is **risk-averse**, agent is **risk-neutral**.
  - Yields a franchise contract.
- **Case 3: Both** principal and agent are **risk-averse**.
  - Generates a sharing rule: each party bears some payoff volatility.
- In every case the model and the binding PC are the same; only the curvatures of B and v change. That single change is enough to reshape the optimal contract, which is the point of the comparison.

2.4.1 Case 1: Risk-neutral principal, risk-averse agent

Case 1 – Simplifying the MRS condition

- **Step 1. Set the curvatures.** Principal risk-neutral, $B(\pi - w) = \pi - w$, so $B' = 1$; agent risk-averse, $v(w)$ strictly concave.
- **Step 2. Substitute $B' = 1$ into (2.4),** which collapses each MRS to $1/v'$:

$$\lambda = \frac{1}{v'(w_H)} = \frac{1}{v'(w_L)}. \tag{2.6}$$

- **Step 3. Invert.** Equal reciprocals give $v'(w_H) = v'(w_L)$, and since v' is strictly decreasing ($v'' \leq 0$) it is one-to-one, so

$$w_H = w_L.$$

Full insurance – The risk-sharing result

- **Full-insurance / risk-sharing result:**
  - The risk-neutral principal offers a contract that pays the *same wage* regardless of the outcome.
  - The risk-averse worker faces **no income risk**.
  - The principal bears all the payoff fluctuations.
- **Intuition:** the less risk-averse party (principal) should bear all the risk in equilibrium.
- This is a Pareto-improving rearrangement: shifting risk from a risk-averse to a risk-neutral party makes the risk-averse party strictly better off without making the risk-neutral one worse off (in expectation).

Full insurance – Why the risk-neutral principal absorbs it

- The risk-neutral principal values only the *mean* of her net return, so she is willing to swap a variable payoff for its expected value at no utility cost. Absorbing the profit risk is therefore free for her.
- The risk-averse worker, in contrast, would demand a risk premium to bear any wage variation, so loading the risk onto him would only raise the wage bill. The principal avoids that by paying a flat wage.
- In Chapter 3, when effort is unobservable, this result will **break down**: incentive provision requires wage variation, so the worker can no longer be fully insured.

Picture: who bears the risk under Case 1?

- **Before the contract:** profit π is random and, without a contract, both parties are exposed to it.
- **After the contract:** the wage is a flat number w^*, independent of the realized π.
  - Worker's income: w^* with probability one – a certainty.
  - Principal's residual: $\pi - w^*$, still random, with the same variance as π itself.
- Since the principal is risk-neutral, she does not mind bearing that variance – she cares only about the mean. The risk-averse worker, in contrast, strictly prefers a certain wage.

Picture: who bears the risk – Efficient allocation

- **Efficient allocation.** Risk sits with the party that can bear it at zero utility cost. This is the same principle behind hiring managers on fixed salaries when the firm's owners are diversified (and thus effectively risk-neutral).
- The flat wage is not charity: it is the cheapest way to meet the PC, because a risk-averse worker would demand a premium for any wage variation the principal could just as well absorb for free.
- The same logic reappears in insurance, where a diversified insurer offers a risk-averse policyholder full coverage against an idiosyncratic loss.

Pinning down the wage level

- Equation (2.6) gives $w_H = w_L$ but not the wage level itself; the PC pins it down.
- **Step 1. Write the binding PC** (it holds with equality at the optimum):

$$p_H v(w_H) + (1 - p_H)\, v(w_L) - g(e) = \overline{u}.$$

- **Step 2. Use** $w_H = w_L$, so $v(w_H) = v(w_L)$ factors out and $p_H + (1 - p_H) = 1$:

$$[p_H + (1 - p_H)]\, v(w_H) - g(e) = \overline{u} \quad \Rightarrow \quad v(w_H) = \overline{u} + g(e).$$

- **Step 3. Invert** v (strictly increasing, so v^{-1} exists):

$$w_H = v^{-1}\left(\overline{u} + g(e)\right).$$

Figure 2.1 – Salary inducing high effort

- Read $\overline{u} + g(e)$ up the vertical axis, then follow v^{-1} across to the wage w_H.

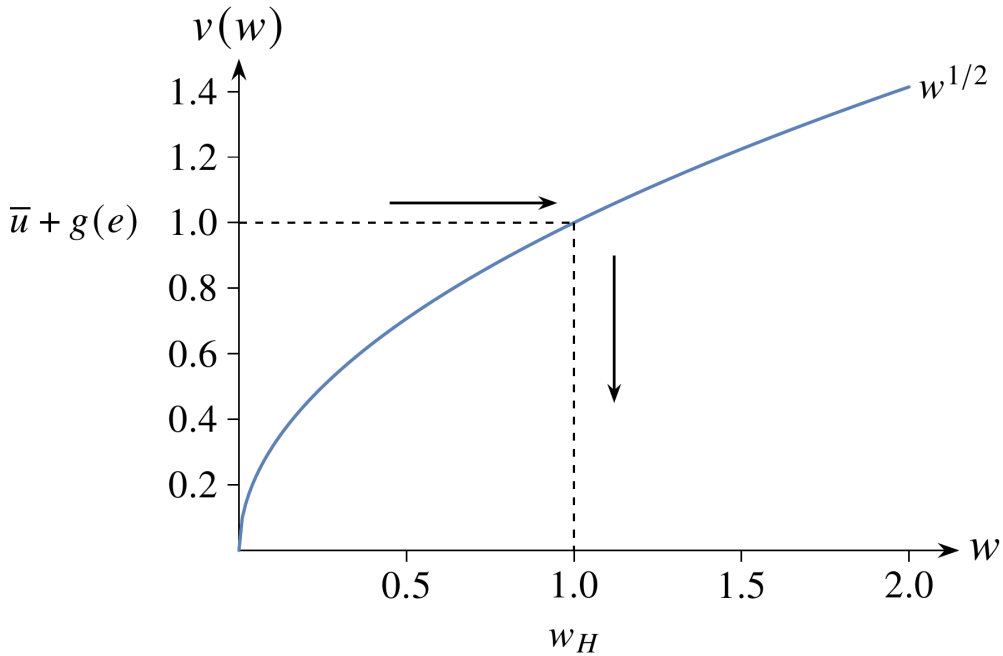


Figure 2.1. Salary w_H inducing high effort under symmetric information. Take the vertical-axis value $\overline{u} + g(e)$ and map it to the horizontal axis through the inverse $v^{-1}(\cdot)$.

Comparative statics of w_H

- From $w_H = v^{-1}(\overline{u} + g(e))$, three comparative statics (each uses the inverse-function rule, $(v^{-1})' = 1/v'$):
- **1. Reservation utility $\overline{u}$:** $\partial w_H / \partial \overline{u} = 1/v'(\overline{u} + g(e)) > 0$, positive since $v' > 0$.
  - Higher outside option $\Rightarrow$ higher wage.
- **2. Cost of effort $g(e)$:** $\partial w_H / \partial g(e) = 1/v'(\overline{u} + g(e)) > 0$.
  - Higher cost of effort $\Rightarrow$ higher wage.

Comparative statics of w_H (cont.)

- **3. Concavity of $v(w)$:** more concave $v \Rightarrow$ **lower** wage to induce the same effort.
  - Intuition: a more risk-averse worker is "easier to please" because his utility function is steep near low wages, so a modest wage already clears $\overline{u} + g(e)$.
- We make this precise in Example 2.5, where $v(w) = w^{1/\alpha}$ and the wage $(\overline{u} + e_i)^{\alpha}$ falls in α whenever $\overline{u} + e_i < 1$.

Figure 2.2a – Effect of $\overline{u}$ and $g(e)$

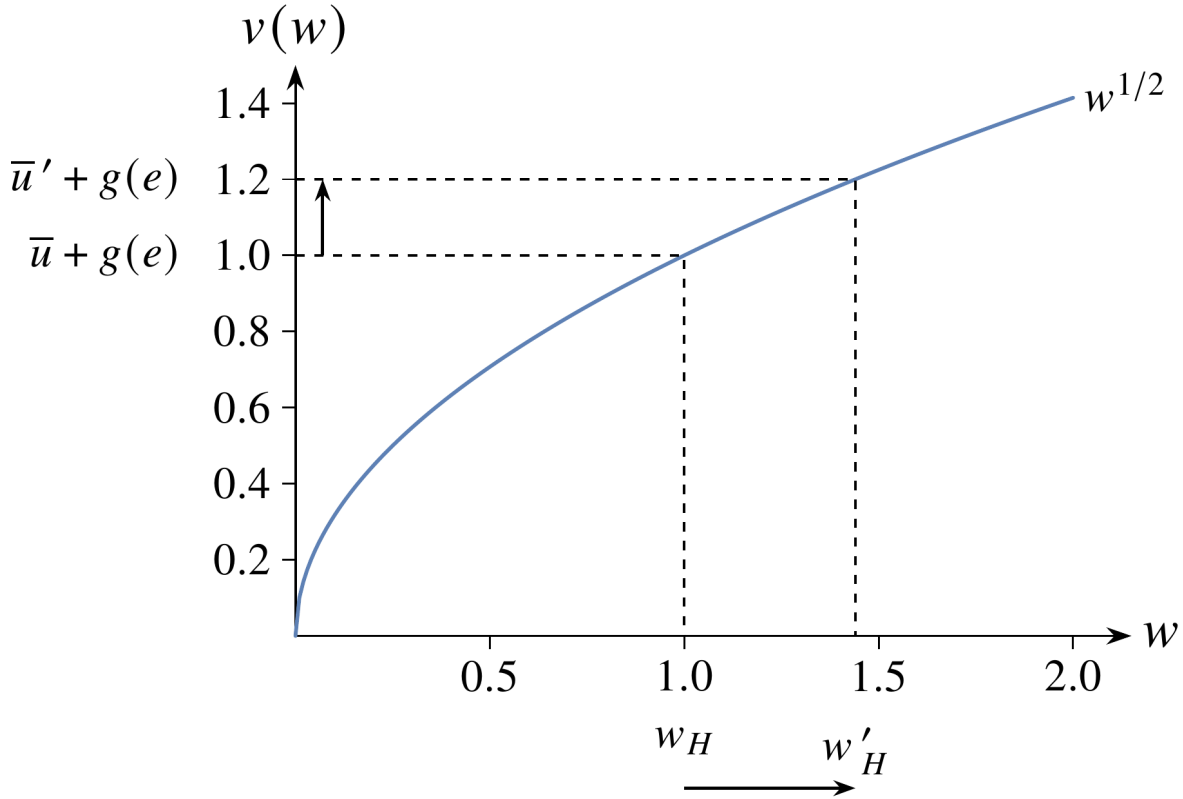


Figure 2.2a. A higher reservation utility $\overline{u}$ or higher cost of effort $g(e)$ raises the vertical-axis value $\overline{u} + g(e)$, which maps into a higher salary w'_H on the horizontal axis.

Figure 2.2b – Effect of concavity of $v(w)$

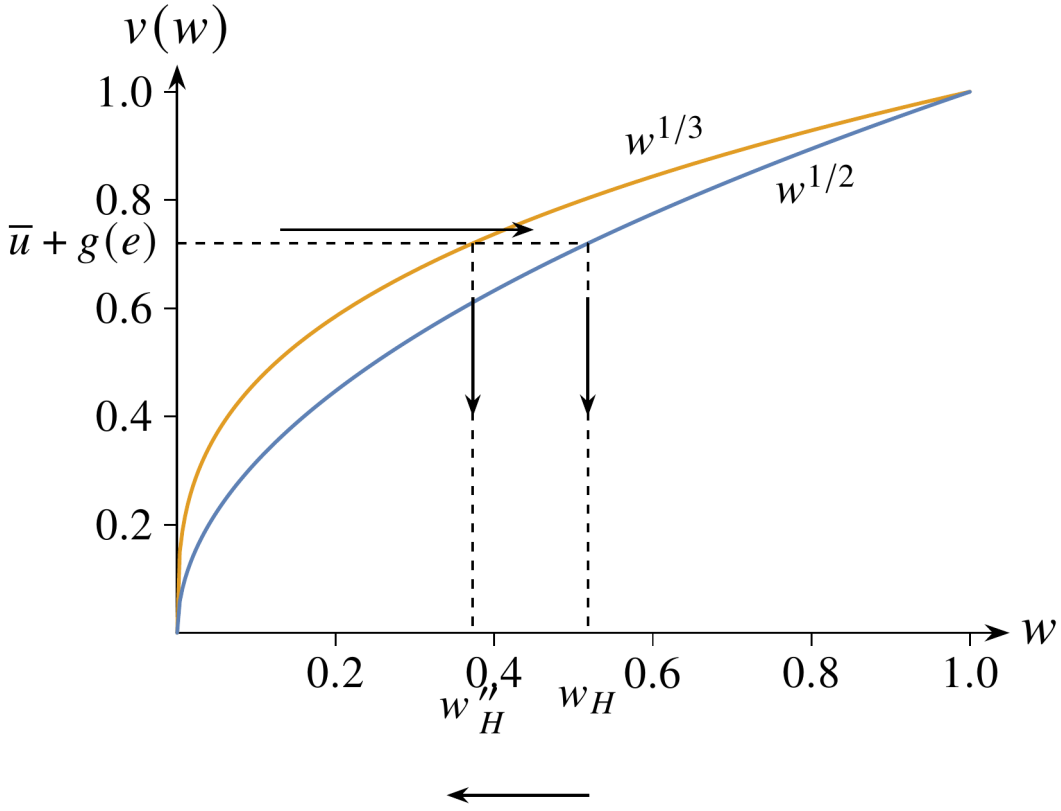


Figure 2.2b. A more concave utility function $v(w)$ maps the same vertical-axis value into a *lower* salary w''_H on the horizontal axis – a more risk-averse worker is "easier to please".

Principal's reduced problem

- Inserting the optimal wage $w_H = w_L = v^{-1}(\overline{u} + g(e))$ into the objective (the common wage factors out and $p_H + p_L = 1$):

$$\begin{aligned} & \max_{e \geq 0} \; p_H \left[\pi_H - w_H\right] + p_L \left[\pi_L - w_L\right] \\ = \quad & p_H \pi_H + p_L \pi_L - v^{-1}\left(\overline{u} + g(e)\right) \end{aligned}$$

- With wages gone from the objective, the principal solves a single-variable problem in e:

$$\max_{e \geq 0} \underbrace{p_H \pi_H + p_L \pi_L}_{E[\pi]} - \underbrace{v^{-1}\left(\overline{u} + g(e)\right)}_{\text{Salary}} \tag{2.7}$$

- Trade-off: higher e raises expected profits $E[\pi]$ but also raises the wage needed to satisfy PC.

First-order condition in effort

- Differentiating (2.7) with respect to e:

$$\frac{\partial p_H}{\partial e}\left(\pi_H - \pi_L\right) - \frac{\partial v^{-1}\left(\overline{u} + g(e)\right)}{\partial e} g'(e) = 0.$$

- Substitute $\partial v^{-1}(\overline{u} + g(e))/\partial e = g'(e)/v'(\overline{u} + g(e))$ (inverse-function rule and chain rule) and move the salary term to the right:

$$\frac{\partial p_H}{\partial e}\left(\pi_H - \pi_L\right) = \frac{g'(e)}{v'\left(\overline{u} + g(e)\right)}. \tag{2.8}$$

- **Interpretation:** the marginal profit from one more unit of effort (LHS) equals the marginal cost of compensating the worker for the extra disutility (RHS). The principal pushes effort to this balance point.

Second-order condition (Eq. 2.9)

- Differentiating (2.8) gives the second-order condition:

$$\frac{\partial^2 p_H}{\partial e^2}(\pi_H - \pi_L) \leq \underbrace{\frac{g''(e)}{v'(\overline{u} + g(e))} - \frac{[g'(e)]^2 v''(\overline{u} + g(e))}{[v'(\overline{u} + g(e))]^2}}_{+} \tag{2.9}$$

- The RHS is unambiguously **positive** because:
  - $g''(e) \geq 0$ and $v' > 0$.
  - $v'' \leq 0$ makes the second term positive.
- So the SOC holds if the LHS is nonpositive, which occurs when p_H is weakly concave in e ($\partial^2 p_H / \partial e^2 \leq 0$); if p_H is convex, that convexity must not be too strong.
- Intuitively, this rules out a runaway where extra effort keeps paying for itself, so the FOC in (2.8) identifies a maximum rather than a minimum.

Example 2.2 – Setup

- Risk-neutral principal hires a risk-averse worker with:
  - $v(w) = \sqrt{w}$, $g(e) = e$, $\overline{u} = 9$.
  - Effort levels: $e_H = 5$, $e_L = 0$.
  - Outcomes: $\pi_H = \$1{,}200$, $\pi_L = \$300$.
  - Probabilities: $p_H = f(\pi_H|e_H) = 0.5$, $p_L = f(\pi_H|e_L) = 0.25$.
- Expected profit, weighting each profit by its probability at that effort:

$$\begin{aligned} E[\pi|e_H] &= 0.5(1{,}200) + 0.5(300) = \$750, \\ E[\pi|e_L] &= 0.25(1{,}200) + 0.75(300) = \$525. \end{aligned}$$

- High effort raises expected profit by $\$750 - \$525 = \$225$, since it tilts probability toward π_H.

Example 2.2 – Wages from PC

- At the optimum the PC binds, so $v(w_i) = \bar{u} + g(e_i)$. With $v(w) = \sqrt{w}$ and $g(e) = e$, this reads $\sqrt{w_i} = \bar{u} + e_i$.
- **High effort** $e_H = 5$. Substitute and isolate the square root, then square both sides:

$$\sqrt{w_H} = \bar{u} + e_H = 9 + 5 = 14 \quad \Rightarrow \quad w_H = 14^2 = \$196.$$

- **Low effort** $e_L = 0$. The same two steps give:

$$\sqrt{w_L} = \bar{u} + e_L = 9 + 0 = 9 \quad \Rightarrow \quad w_L = 9^2 = \$81.$$

- Inducing more effort costs more, $w_H = \$196 > \$81 = w_L$: the worker must be compensated for the extra disutility $g(e_H) - g(e_L) = 5$ he bears.
- Check: $\sqrt{196} = 14 = \bar{u} + e_H$ and $\sqrt{81} = 9 = \bar{u} + e_L$, so the PC holds with equality at both wages.

Example 2.2 – Which effort does the principal induce?

- Principal's net profit from inducing **high effort**:

$$E[\pi|e_H] - w_H = \$750 - \$196 = \$554.$$

- Principal's net profit from inducing **low effort**:

$$E[\pi|e_L] - w_L = \$525 - \$81 = \$444.$$

- Since $\$554 > \$444$, the principal optimally induces $e_H = 5$; her net gain from the high-effort contract is $\$554 - \$444 = \$110$.
- **Key:** even though inducing e_H requires a higher wage (\$196 vs. \$81, a \$115 increase), the \$225 gain in expected profit more than compensates.

Example 2.3 – Effect of reservation utility

- Same setting as Example 2.2 but with general $\overline{u} \geq 0$ and general effort levels e_H, e_L.
- Binding PC: $\sqrt{w_i} - e_i = \overline{u}$; isolate the root and square both sides:

$$w_i = (\overline{u} + e_i)^2 \quad \text{for } i = \{H, L\}.$$

- Comparative statics (chain rule on the square):

$$\frac{\partial w_i}{\partial \overline{u}} = 2(\overline{u} + e_i) > 0, \qquad \frac{\partial w_i}{\partial e_i} = 2(\overline{u} + e_i) > 0.$$

- Both effects are unambiguously **positive**. Check: at $\overline{u} = 9$, $e_H = 5$ we recover $w_H = (9 + 5)^2 = \$196$, exactly as in Example 2.2.

Example 2.4 – Effect of disutility of effort

- Now consider $g(e) = a + be$, $a \geq 0$, $b \geq 1$.
  - $a = 0, b = 1$: collapses to $g(e) = e$ as in Example 2.2.
  - $a > 0$: shifts the cost function upward.
  - $b > 1$: makes the cost function steeper in e.
- Binding PC: $\sqrt{w_i} = \overline{u} + a + be_i$; square both sides:

$$w_i = \left(\overline{u} + a + be_i\right)^2 .$$

- Comparative statics:

$$\begin{aligned} \frac{\partial w_i}{\partial a} &= 2\left(\overline{u} + a + be_i\right) > 0, \\ \frac{\partial w_i}{\partial b} &= 2e_i\left(\overline{u} + a + be_i\right) > 0. \end{aligned}$$

- Intuitively, a higher fixed cost a or a steeper marginal cost b both raise the disutility of a given effort, so the principal must pay more to keep the worker participating.

Example 2.5 – Effect of risk aversion

- Now $v(w) = w^{1/\alpha}$, $\alpha \geq 1$:
  - $\alpha = 1$: $v(w) = w$, risk-neutral.
  - $\alpha = 2$: $v(w) = \sqrt{w}$ (Example 2.2).
  - Larger α: more concave, more risk-averse.
- Binding PC: $w_i^{1/\alpha} - e_i = \overline{u}$; raise both sides to the power α:

$$w_i = (\overline{u} + e_i)^{\alpha}.$$

- Comparative statics (differentiate x^{α} in α, giving $x^{\alpha} \ln x$):

$$\frac{\partial w_i}{\partial \alpha} = (\overline{u} + e_i)^{\alpha} \ln(\overline{u} + e_i),$$

  which is **negative** whenever $\overline{u} + e_i < 1$, since $\ln(x) < 0$ for $x < 1$.
- **Interpretation:** a more risk-averse worker is "easier to please," so the wage falls; e.g. $(\overline{u} + e_i)^3 < (\overline{u} + e_i)^2$ exactly when $\overline{u} + e_i < 1$.

Case-comparison teaser – What flips in Case 2?

- **Case 1** (just finished): risk-neutral principal, risk-averse worker $\Rightarrow$ constant wage, principal absorbs all risk.
- **Case 2** (next): the roles are swapped – risk-averse principal, risk-neutral worker.
- **What should we expect?** By the same risk-sharing logic, the *worker* (now the less risk-averse party) should bear the risk. So the wage should *move with* profits.

Case-comparison teaser – Franchise contracts and Case 3

- **Familiar instance:** a **franchise contract** – the worker pays the principal a fixed fee and keeps the residual profit. We will derive this shortly.
- **Case 3** (after that): both parties are risk-averse – the wage moves with profits, but only partially, depending on the ratio of risk-aversion coefficients.

2.4.2 Case 2: Risk-averse principal, risk-neutral agent

Case 2 – The setup

- Principal: risk-averse, $B(\pi - w)$ is strictly increasing and concave, $B' > 0$, $B'' < 0$.
- Agent: risk-neutral, $v(w) = w$, $u(w, e) = w - g(e)$.
- Equation (2.4) becomes

$$\lambda = B'(\pi_H - w_H) = B'(\pi_L - w_L),$$

since $v'(w_H) = v'(w_L) = 1$.

- Strict concavity makes B' one-to-one, so equal marginal benefits force equal arguments:

$$\pi_H - w_H = \pi_L - w_L = K.$$

- That is, the principal's net return is the same constant K in both outcomes.

The franchise contract

- The risk-averse principal receives the same payoff K regardless of profit realization.
- The worker's wage is:

$$w_H = \pi_H - K \quad \text{when profits are high,}$$

$$w_L = \pi_L - K \quad \text{when profits are low.}$$

- **Now the risk-neutral worker bears all the risk.**

The franchise contract – Deriving the fee

- This is a **franchise contract**:
  - Principal "delegates" production to the worker.
  - Worker pays a fixed franchise fee K to the principal.
  - Worker keeps the realized profit (π_H or π_L).
- Real-world example: McDonald's franchisees pay a franchise fee and keep the operating profits.

Why a franchise contract? – Economic reading

- In a **franchise contract** the worker (risk-neutral) buys the right to the firm's profit for a fixed fee K and keeps whatever remains.
- So the worker's residual is $\pi - K$; the principal's payoff is just K, independent of the realized profit.
- **Two things this delivers at once:**
  - *Risk-sharing.* The risk-averse principal's payoff is now certain (K); all the profit variance sits with the risk-neutral worker.
  - *Effort provision.* The worker is the residual claimant, so he internalizes the full marginal return to effort – there is no wedge between his effort choice and the principal's ideal effort.

Why a franchise contract? – Examples and Chapter 3 tension

- **Real-world instances beyond retail franchising:** taxi medallions rented to drivers, farm sharecropping with a fixed rent, independent-contractor arrangements, some management buyouts.
- **Tension in Chapter 3.** Once effort is hidden and the worker is risk-averse, the same contract has to trade risk-sharing against effort provision – the two goals stop being compatible.

Finding the franchise fee K

- **Step 1. Insert the wages $w_i = \pi_i - K$ into the binding PC** (the worker's wage is his residual profit):

$$p_H(\pi_H - K) + p_L(\pi_L - K) - g(e) = \overline{u}.$$

- **Step 2. Collect terms**; the K's add up since $p_H + p_L = 1$, leaving expected profit minus K:

$$\underbrace{p_H\pi_H + p_L\pi_L}_{E[\pi]} - K - g(e) = \overline{u}.$$

- **Step 3. Solve for K:**

$$K = E[\pi] - \overline{u} - g(e). \tag{2.10}$$

- **Comparative statics:**
  - K *increases* in expected profits $E[\pi]$.
  - K *decreases* in $\overline{u}$ and $g(e)$.

Principal's reduced problem (Case 2)

- Plugging wages $w_H = \pi_H - K$ and $w_L = \pi_L - K$ into the objective, each net return is K, so B is the same in both outcomes:

$$\begin{aligned} & p_H B(\pi_H - w_H) + (1 - p_H)\, B(\pi_L - w_L) \\ = \quad & p_H B(K) + (1 - p_H)\, B(K) = B(K), \end{aligned}$$

  where the last step uses $p_H + (1 - p_H) = 1$.

- With her benefit now certain, the principal's problem reduces to:

$$\max_{e \geq 0} \; B(K) = B\left(E[\pi] - \overline{u} - g(e)\right). \tag{2.11}$$

Effort FOC and SOC (Case 2)

- Differentiating (2.11) with respect to e (chain rule through B):

$$B'(\cdot)\left(\frac{\partial p_H}{\partial e}(\pi_H - \pi_L) - g'(e)\right) = 0.$$

- Cancel $B' > 0$, so the FOC simplifies to:

$$\frac{\partial p_H}{\partial e}(\pi_H - \pi_L) = g'(e). \tag{2.12}$$

- **SOC** (differentiate 2.12; RHS positive since $g'' \geq 0$):

$$\frac{\partial^2 p_H}{\partial e^2}(\pi_H - \pi_L) \leq \underbrace{g''(e)}_{+}. \tag{2.13}$$

- It holds when p_H is linear or weakly concave in effort. Note (2.12) is the Case 1 condition with $v' = 1$: a risk-neutral worker needs no insurance, so only his disutility enters.

Example 2.6 – Risk-averse principal, RN agent

- Setup of Example 2.2 but now principal is risk-averse: $B(\pi_i - w_i) = \sqrt{\pi_i - w_i}$, $v(w_i) = w_i$.
- Expected profits unchanged: $E[\pi|e_H] = \$750$, $E[\pi|e_L] = \$525$, with $\overline{u} = 9$, $g(e_H) = 5$, $g(e_L) = 0$.
- Franchise fee from (2.10):

$$\begin{aligned} K &= E[\pi] - \overline{u} - g(e) = \$750 - 9 - 5 = \$736 \text{ (high effort)}, \\ K &= E[\pi] - \overline{u} - g(e) = \$525 - 9 - 0 = \$516 \text{ (low effort)}. \end{aligned}$$

- Since B is increasing, maximizing $B(K)$ means choosing the higher K, so she induces high effort and sets $K = \$736$.
- Check: under high effort the worker keeps $\pi_i - K$, with expected value $E[\pi] - K = \$750 - \$736 = \$14 = \overline{u} + g(e_H)$, so his PC binds.

2.4.3 Case 3: Principal and agent are both risk-averse

Case 3 – Both risk-averse

- Now both parties are risk-averse, so B and v are both concave.
- With neither party risk-neutral, the wage in outcome i moves with the profit it is paid on. Write it as $w_i(\pi_i)$ and ask how responsive it is, $w_i'(\pi_i)$.
- **Step 1. Start from the wage FOC** in outcome i:

$$-B'(\pi_i - w_i) + \lambda v'(w_i) = 0.$$

- **Step 2. Differentiate with respect to π_i.** The argument $\pi_i - w_i$ changes by $1 - w_i'$, and w_i by w_i', so the chain rule gives

$$-B''\left(1 - w_i'(\pi_i)\right) + \lambda v''\, w_i'(\pi_i) = 0.$$

Case 3 – Solving for the wage slope

- **Step 3. Collect the w_i' terms.** Expanding the previous line and grouping,

$$-B'' + \left(B'' + \lambda v''\right) w_i'(\pi_i) = 0 \quad \Rightarrow \quad w_i'(\pi_i) = \frac{B''}{B'' + \lambda v''}.$$

- **Step 4. Substitute $\lambda = B'/v'$** (from the wage FOC in Case 1):

$$w_i'(\pi_i) = \frac{B''}{B'' + \dfrac{B'}{v'} v''} = \frac{B''}{B'' + B' \dfrac{v''}{v'}}.$$

- Both B'' and v'' are negative, so the slope is a ratio of negative terms: it lies between 0 and 1, as we read next.

Wage responsiveness and Arrow-Pratt

- Divide numerator and denominator by $B' > 0$ (legitimate, and it makes the Arrow-Pratt ratios appear):

$$w_i'(\pi_i) = \frac{\frac{B''}{B'}}{\frac{B''}{B'} + \frac{v''}{v'}} = \frac{r_P}{r_P + r_W} \tag{2.14}$$

- Where:

$$r_P \equiv -\frac{B''}{B'}, \qquad r_W \equiv -\frac{v''}{v'},$$

are the **Arrow-Pratt coefficients of absolute risk aversion** of the principal and worker.

Arrow-Pratt – Interpretation

- r captures the **curvature** of the utility function: a larger r means a more risk-averse party, who demands a larger **risk premium** to bear a given gamble.
- A more risk-averse worker ($r_W \uparrow$) makes the wage less responsive to profit, while a more risk-averse principal ($r_P \uparrow$) makes it more responsive.
- The slope $w_i'(\pi_i) \in [0, 1]$ measures how much the wage moves with each \$1 change in profit: 0 is full insurance, 1 is the worker bearing every dollar.

Comparative statics of $w_i'(\pi_i)$

- Compute partials of (2.14) by the quotient rule; the signs follow because $r_P, r_W > 0$:

$$\begin{aligned} \frac{\partial w_i'(\pi_i)}{\partial r_P} &= \frac{r_W}{(r_P + r_W)^2} > 0, \\ \frac{\partial w_i'(\pi_i)}{\partial r_W} &= -\frac{r_P}{(r_P + r_W)^2} < 0. \end{aligned}$$

- Wages fluctuate **more** when the principal is more risk-averse (higher r_P).
- Wages fluctuate **less** when the worker is more risk-averse (higher r_W).
- Intuition: the contract shifts risk onto whichever party is less risk-averse, since that party bears it at lower utility cost.

Figure 2.3 – Salary response to profits

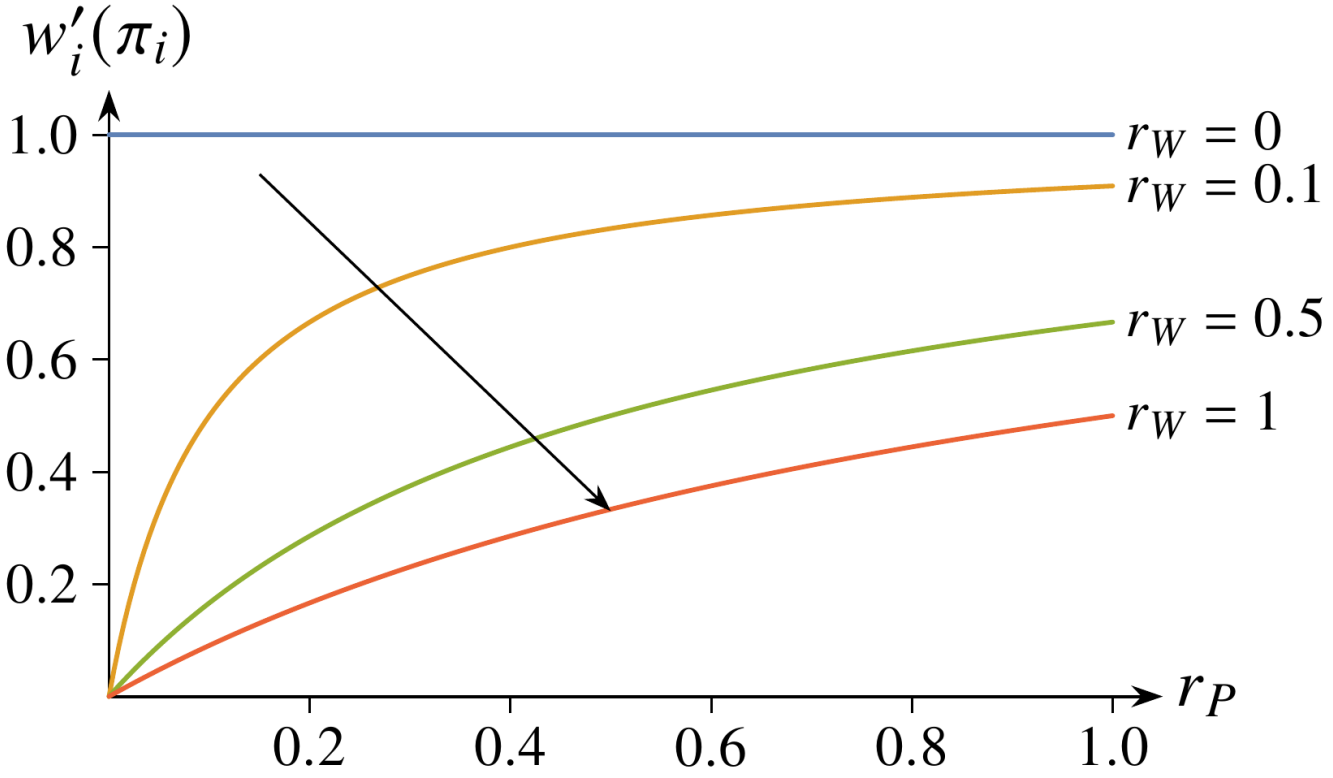


Figure 2.3. Salary response $w_i'(\pi_i) = r_P/(r_P + r_W)$ plotted against r_P on the horizontal axis. Curves for $r_W = 0, 0.1, 0.5, 1$. When $r_W = 0$, response is constant at 1 (worker bears all the risk).

Reading Figure 2.3

- **Top curve ($r_W = 0$):** worker is risk-neutral, so $w_i'(\pi_i) = 1$ regardless of r_P.
  - Worker bears all volatility (confirming Case 2 – franchise).
- **Lower curves ($r_W > 0$):** as r_W rises, the curve shifts down.
  - For each level of r_P, a more risk-averse worker has wages that move less with profits.
- **Origin ($r_P = 0$):** $w_i'(\pi_i) = 0$ for any $r_W > 0$ – worker gets a constant wage (Case 1).

Reading Figure 2.3 (cont.)

- Between the extremes, both parties share the risk: the wage moves partway with profits, by a fraction $r_P/(r_P + r_W)$ strictly between 0 and 1.
- Moving right along any curve (higher r_P) raises that fraction, so a more risk-averse principal pushes more of the risk onto the worker.
- The height of each curve is set by r_W: the four curves are simply the formula $r_P/(r_P + r_W)$ read off at $r_W = 0, 0.1, 0.5, 1$.

Risk-aversion asymmetry – Worker more averse

- Case (a): worker is more risk-averse than principal, $r_W > r_P > 0$.
- Then:

$$w_i'(\pi_i) = \frac{r_P}{r_P + r_W} < \frac{1}{2}.$$

- Worker's wage faces only **small** variations when profits change.
- **Concrete interpretation:** a \$1 decrease in profits reduces the worker's wage by less than 50 cents; the principal absorbs more than 50 cents of the loss.
- Example: $r_P = 0.1$, $r_W = 0.3$ give $w_i'(\pi_i) = 0.1/0.4 = 0.25$, so the worker bears only a quarter of each dollar swing.
- Result: the principal, being less risk-averse, shoulders most of the income volatility.

Risk-aversion asymmetry – Principal more averse

- Case (b): principal is more risk-averse than worker, $r_P > r_W > 0$.
- Then:

$$w_i'(\pi_i) = \frac{r_P}{r_P + r_W} > \frac{1}{2}.$$

- A \$1 decrease in profits reduces the worker's wage by more than 50 cents; e.g. $r_P = 0.3$, $r_W = 0.1$ give $w_i'(\pi_i) = 0.3/0.4 = 0.75$.
- Worker – being less risk-averse – absorbs most of the volatility.
- **Symmetric case:** $r_P = r_W = r \Rightarrow$

$$w_i'(\pi_i) = \frac{r}{r + r} = \frac{1}{2}.$$

Both parties bear equal shares of the risk.

General lesson on risk-sharing

- **Key insight:** the *less* risk-averse player bears *more* payoff volatility in equilibrium.
- Corner cases recover Cases 1 and 2:
  - $r_P = 0$ (Case 1, risk-neutral principal): $w_i'(\pi_i) = 0$ – worker gets full insurance.
  - $r_W = 0$ (Case 2, risk-neutral worker): $w_i'(\pi_i) = 1$ – worker bears all risk (franchise).
- Interior solutions: both parties share risk proportionally to their relative tolerance.

General lesson on risk-sharing (cont.)

- This is the **first-best risk-sharing** arrangement.
- In Chapter 3, this principle breaks down when effort is hidden – risk-sharing must be sacrificed to provide incentives.

Why go beyond two outcomes?

- Two outcomes (π_H, π_L) kept the algebra clean and let every mechanism show up in its simplest form.
- But real profit is not binary – it takes many values.
  - Sales revenue can hit any dollar amount.
  - Insurance claims can be small, medium, or large.
  - R&D projects can succeed, partially succeed, or fail.
- **What survives with N outcomes?**
  - The Lagrangian setup is identical – just add more terms.
  - The MRS condition still equates marginal rates of substitution across every pair of outcomes.
  - Case 1 still delivers a flat wage (full insurance).

Why go beyond two outcomes? – What survives and what is new

- **What is new?** With three or more outcomes the wage schedule need not be monotone in profit – a hint of the *informativeness principle* (Chapters 3–4), where the wage rewards the outcome that best signals effort, not simply the highest profit.
- For now, under symmetric information and a risk-neutral principal the schedule stays flat, as we confirm in Example 2.7: adding outcomes does not disturb full insurance.
- The payoff of the general setup is that it carries straight into the asymmetric-information chapters with no change of notation.

Section 2.5
More than Two Outcomes

The general problem with N outcomes

- Now allow $N \geq 2$ possible profit realizations: $\pi_1, \pi_2, \ldots, \pi_N$ with associated probabilities $p_i = f(\pi_i|e)$.
- Properties carry over: $p_i \in [0,1]$ and $\sum_{i=1}^{N} p_i = 1$.
- Principal's problem, now summing the benefit over all N outcomes:

$$\max_{e, w_1, \ldots, w_N \geq 0} \sum_{i=1}^{N} p_i B(\pi_i - w_i)$$

$$\text{s.t.} \quad \sum_{i=1}^{N} p_i v(w_i) - g(e) \geq \overline{u}. \quad \text{(PC)}$$

- The principal now chooses $N+1$ variables (effort e and N wages), but there is still a single PC, since effort stays observable.

Lagrangian with N outcomes

- Lagrangian:

$$\mathcal{L} = \sum_{i=1}^{N} p_i B(\pi_i - w_i) + \lambda \left[\sum_{i=1}^{N} p_i v(w_i) - g(e) - \overline{u} \right].$$

- Differentiating with respect to w_i for each i (chain rule puts the -1 on B'):

$$-p_i B'(\pi_i - w_i) + \lambda p_i v'(w_i) = 0.$$

- Cancel $p_i > 0$ and divide by $v'(w_i) > 0$ to solve for λ:

$$\lambda = \frac{B'(\pi_i - w_i)}{v'(w_i)} \quad \text{for every } i.$$

- **Same MRS condition as before**, now imposed across all N outcomes: every pair of wages satisfies $MRS_i = MRS_j$, so nothing qualitative changes as N grows.

Example 2.7 – Two efforts, three outcomes

- Worker exerts $e_H = e$ or $e_L = 0$.
- Outcomes: $\pi_H > \pi_M > \pi_L$.

		Profits		
		π_H	π_M	π_L
Effort	$e_H = e$	p_1	p_2	p_3
	$e_L = 0$	q_1	q_2	q_3

Table 2.4. Conditional probabilities with three outcomes.

- Worker utility $v(w) = \sqrt{w}$, disutility $g(e) = e$, reservation $\overline{u}$.

Example 2.7 – Principal's problem (high effort)

- Inducing high effort:

$$\max_{w_H, w_M, w_L} (p_1\pi_H + p_2\pi_M + p_3\pi_L) - (p_1 w_H + p_2 w_M + p_3 w_L)$$

$$\text{s.t.} \quad p_1\sqrt{w_H} + p_2\sqrt{w_M} + p_3\sqrt{w_L} - e \geq \overline{u}. \quad \text{(PC)}$$

- PC binds, so set it to equality, isolate $p_1\sqrt{w_H}$, and square:

$$w_H = \left(\frac{\overline{u} + e - p_2\sqrt{w_M} - p_3\sqrt{w_L}}{p_1}\right)^2.$$

- Substituting this w_H into the objective removes one variable. Since expected profit is fixed once effort is set, maximizing profit is the same as minimizing the expected wage bill in w_M, w_L.

Example 2.7 – Solving the FOCs

- Differentiating with respect to w_M and w_L yields:

$$\begin{aligned} \overline{u} + e - p_2\sqrt{w_M} - p_3\sqrt{w_L} &= p_1\sqrt{w_M}, \\ \overline{u} + e - p_2\sqrt{w_M} - p_3\sqrt{w_L} &= p_1\sqrt{w_L}. \end{aligned}$$

- The two FOCs share the same left-hand side, so $p_1\sqrt{w_M} = p_1\sqrt{w_L}$, giving $\sqrt{w_M} = \sqrt{w_L}$, i.e.

$$w_M = w_L.$$

- Substitute back into the binding PC and use $p_1 + p_2 + p_3 = 1$:

$$\overline{u} + e = (p_1 + p_2 + p_3)\sqrt{w_L} = \sqrt{w_L}.$$

- Hence $w_L = (\overline{u} + e)^2$, $w_M = (\overline{u} + e)^2$, and feeding these into w_H gives $w_H = (\overline{u} + e)^2$ as well.

Example 2.7 – Constant wage across outcomes

- All three wages are equal:

$$w_H = w_M = w_L = (\overline{u} + e)^2.$$

- This confirms the **full-insurance result** for Case 1: with risk-neutral principal and risk-averse agent, wages do not vary with outcomes – regardless of how many outcomes there are.
- The constant wage exactly offsets the worker's reservation utility $\overline{u}$ plus disutility $g(e) = e$.
- **General lesson:** the MRS condition (2.4) holding across all outcomes pins down a single optimal wage when the principal is risk-neutral.

Example 2.7 – Inducing high vs. low effort

- Net profit from high effort:

$$p_1\pi_H + p_2\pi_M + p_3\pi_L - (\overline{u} + e)^2.$$

- Net profit from low effort:

$$q_1\pi_H + q_2\pi_M + q_3\pi_L - (\overline{u})^2.$$

- Principal induces high effort iff

$$(p_1 - q_1)\pi_H + (p_2 - q_2)\pi_M + (p_3 - q_3)\pi_L > e(2\overline{u} + e).$$

- LHS: extra expected profit from inducing high effort.
- RHS: extra wage cost (simplified using $(\overline{u} + e)^2 - \overline{u}^2 = e(2\overline{u} + e)$).

Example 2.7 – Numerical example (Table 2.5)

- Parameters: $e = \frac{1}{2}$, $\overline{u} = 1$, $\pi_H = 100$, $\pi_M = 50$, $\pi_L = 10$.

		Profits		
		π_H	π_M	π_L
Effort	$e_H = e$	0.6	0.3	0.1
	$e_L = 0$	0.1	0.2	0.7

Table 2.5. Conditional probabilities – numerical example.

- Wages: $w_i = (\overline{u} + e)^2 = \left(\frac{3}{2}\right)^2 = \$2.25$ for every i, flat across outcomes by the full-insurance result.

Example 2.7 – Numerical results

- $E[\pi|e_H] = 0.6(100) + 0.3(50) + 0.1(10) = \$76.$
- $E[\pi|e_L] = 0.1(100) + 0.2(50) + 0.7(10) = \$27.$
- The wage is $(\overline{u} + e)^2 = \$2.25$ under high effort but $(\overline{u})^2 = \$1$ under low effort ($e_L = 0$), so net profits are $\$76 - \$2.25 = \$73.75$ versus $\$27 - \$1 = \$26$.
- Since $\$73.75 > \$26$, the principal optimally induces e_H: the \$49 gain in expected profit swamps the \$1.25 extra wage.

Section 2.6
Summary

Table 2.6 – Comparative statics of wages

	Parameter	Effect on wages
Reservation utility	$\overline{u}$	$+$
Cost of effort	$g(e)$	$+$
Agent's risk aversion	α	$-$
Convexity of cost of effort	β	$+$
Probabilities of outcomes	π_H and π_L	$=$

Table 2.6. Summary of comparative statics – Symmetric information.

- Wage **increases** in $\overline{u}$, $g(e)$, and convexity β.
- Wage **decreases** in agent's risk aversion α.
- Wage **unaffected** by the probability distribution over outcomes – a key feature that changes under asymmetric information (Chapter 3).

Summary – Moral hazard examples

- Moral hazard arises whenever an agent's actions are imperfectly monitored or whenever risks borne by one party are transferred to another.
- **Labor markets:** workers may shirk; addressed by performance pay and monitoring.
- **Health insurance:** overconsumption of services; addressed by co-pays and VBID.

Summary – Case studies at scale

- **Financial markets:** excessive risk-taking; addressed by capital requirements and regulation.
- Four case studies illustrate moral hazard at scale:
  - 2007–08 financial crisis ("too big to fail").
  - U.S. health insurance 1990s (low co-pays).
  - Israel mandatory auto insurance 1975.
  - Tanking in professional sports (hidden action).

Summary – The model

- **Principal** maximizes $B(\pi - w)$; **agent** maximizes $u(w, e) = v(w) - g(e)$.
- Effort e affects the conditional probability $f(\pi|e)$ of each profit realization.
- Properties: $f(\pi_i|e) \in [0, 1]$, $\sum_i f(\pi_i|e) = 1$.
- Under symmetric information, the principal solves:

$$\max_{e,\{w_i\}} \sum_i p_i B(\pi_i - w_i) \quad \text{s.t. PC.}$$

- The PC **binds** in equilibrium (otherwise the principal would extract more surplus).

Summary – Key results across cases

- **MRS condition** (Eq. 2.5):

$$\frac{B'(\pi_H - w_H)}{v'(w_H)} = \frac{B'(\pi_L - w_L)}{v'(w_L)}.$$

- Three risk-attitude configurations yield distinct optimal contracts:
  - **Case 1** (RN principal, RA agent): $w_H = w_L = v^{-1}(\bar{u} + g(e))$ **– full insurance**.
  - **Case 2** (RA principal, RN agent): $\pi_i - w_i = K$ **– franchise contract**.
  - **Case 3** (both RA): $w_i'(\pi_i) = \dfrac{r_P}{r_P + r_W}$ **– risk-sharing rule**.
- **Universal lesson:** the less risk-averse party bears more of the income volatility.

Summary – Generalizations and extensions

- **More than two outcomes:** the principal's problem extends naturally with $N \geq 2$ outcomes.
- Same MRS condition applies across every pair of outcomes.
- For risk-neutral principal: wage remains **constant** across all N outcomes (full insurance generalizes).
- These results are the **first-best benchmark**.
- In Chapter 3, asymmetric information **breaks** all three clean risk-attitude results:
  - No more full insurance for the risk-averse agent.
  - Wages must vary with output to induce effort – the risk-sharing-vs-incentives trade-off.

Looking ahead – Chapter 3

- Chapter 2 established the **symmetric-information benchmark**.
- Chapter 3 relaxes this: the principal can no longer observe the agent's effort.
- This introduces a **new constraint**:

$$\text{Incentive Compatibility (IC): } p_H v(w_H) + p_L v(w_L) - g(e_H)$$

$$\geq (1 - p_H) v(w_H) + (1 - p_L) v(w_L) - g(e_L).$$

Looking ahead – Chapter 3 (cont.)

- Key new results:
  - Wages must spread across outcomes ($w_H > w_L$) to provide incentives.
  - Risk-sharing vs. incentives **trade-off**.
  - Wages are **higher** under asymmetric than symmetric info.

Common student mistakes to avoid

- **Signing errors in FOCs.** When you differentiate $B(\pi - w)$ with respect to w, the chain rule gives $B'(\pi - w) \times (-1)$. The -1 is not the derivative of B.
- **Reading PC as an inequality after solving.** The PC binds in equilibrium ($\lambda > 0$); if you leave it as a strict inequality you are throwing away one equation and typically fail to pin down the wage level.
- **Confusing $w_H = w_L$ with $w = \overline{u}$.** Under Case 1 the wage is flat across outcomes, but its *level* depends on $\overline{u}$, $g(e)$, and the shape of v – it is not just $\overline{u}$.

Common student mistakes to avoid (cont.)

- **Applying Case 1 conclusions everywhere.** Full insurance holds because the principal is risk-neutral, not because information is symmetric. In Chapter 3, symmetric information disappears and so does full insurance.
- **Treating p_H as a constant.** Probabilities depend on the induced effort ($p_H = f(\pi_H|e)$); when the principal chooses e she also changes p_H.

Key takeaways from Chapter 2

1. Under symmetric information, the only binding constraint is the **participation constraint** (PC).
2. Optimal contracts are characterized by the **MRS condition**:

$$\frac{B'(\pi_H - w_H)}{v'(w_H)} = \frac{B'(\pi_L - w_L)}{v'(w_L)}.$$

3. Three risk-attitude configurations give three distinct contract shapes (full insurance, franchise, sharing rule).
4. In all three, the **less risk-averse party bears more risk**.
5. Wages are unaffected by the probability distribution over outcomes – a property that will fail in Chapter 3.
6. This chapter is the **first-best benchmark** against which we compare second-best contracts under asymmetric information.

End of Chapter 2

Chapter 3

Moral Hazard under Asymmetric Information

Instructor Slides

Contract Theory: An Introduction with Applications

Ana Espinola-Arredondo, Felix Muñoz-Garcia, and Pak-Sing Choi

Outline of Chapter 3

- This chapter extends Chapter 2 by allowing the principal to be unable to observe the agent's effort – **asymmetric information**.
- Sections:
  - 3.1 Introduction.
  - 3.2 Remodeling the principal's problem – IC constraint, MLRP, roadmap.
  - 3.3 More effort levels and more outcomes.
  - 3.4 Applications – credit rationing, insurance, efficiency wages.
  - 3.5 Summary. 3.6 Exercises. Appendix 3.A.

What you will learn

- How to add the **incentive compatibility** (IC) constraint to the principal's problem.
- Why wages must be **higher under asymmetric information** for high outcomes and **lower** for low outcomes.
- The **Monotone Likelihood Ratio Property** (MLRP) – when optimal wages are increasing in profits.
- A step-by-step **roadmap** for solving moral-hazard problems with hidden action.
- How to handle settings with more than two effort levels or more than two outcomes.
- Three applications: **credit rationing**, **insurance design**, and **efficiency wages**.

Section 3.1
Introduction

Asymmetric information – The hidden action problem

- Chapter 2 assumed the principal could perfectly observe the agent's effort.
- In most real-world settings, this is not realistic:
  - The manager cannot watch every minute of the worker's day.
  - Cameras and logs are imperfect (cannot measure concentration or focus).
  - Sales outcomes may reflect luck, not effort.
- **Hidden action problem:** the agent chooses effort *after* the contract is signed, and the principal cannot verify it.

Asymmetric information – The hidden action problem (cont.)

- Contracts must be designed so that the agent *voluntarily* chooses the desired effort – the **Incentive Compatibility** (IC) constraint.
- Real-world examples:
  - Insurance: drivers, patients.
  - Lending: borrowers' use of funds.
  - Labor: shirking.

New ingredient – The IC constraint

- In Chapter 2, the only constraint was the **Participation Constraint** (PC), worker's expected utility at least $\overline{u}$:

$$p_H v(w_H) + (1 - p_H) v(w_L) - g(e_H) \geq \overline{u}.$$

- Under asymmetric information, we add the **Incentive Compatibility constraint** (IC): the agent must prefer the desired effort over deviating.
- For two effort levels and two outcomes:

$$\begin{aligned} & p_H v(w_H) + (1 - p_H) v(w_L) - g(e_H) \\ \geq \quad & p_L v(w_H) + (1 - p_L) v(w_L) - g(e_L). \end{aligned}$$

The left-hand side is the worker's payoff from the target effort e_H; the right-hand side, his payoff from the best deviation e_L.

New ingredient – The IC constraint (cont.)

- **New trade-off:** wages now serve two purposes – meet PC *and* provide incentives via IC.
- Under symmetric information the principal only had to keep the worker participating. Now she must also keep him from quietly switching to the cheaper effort once the contract is signed.
- One channel for both goals: the **wage spread** $w_H - w_L$. A wider spread makes high effort more attractive (helps IC) but exposes the risk-averse worker to income risk (makes PC more expensive).
- Example: a salesperson paid the same whether sales are high or low has no reason to work hard; tying pay to sales restores the incentive, at the cost of a riskier paycheck.

IC constraint – Reading it in one line

- In words: the agent, given the offered wage schedule, must find high effort at least as good as any alternative effort.
- **Two ways it can be written:**
  - As an *inequality*, comparing expected utility from the target effort against every deviation (as we do here).
  - As a *first-order condition* of the agent's own problem (the **first-order approach**, used in Chapter 4).

IC constraint – Why not use the FOC now?

- **Why not use the FOC now?** Because with two discrete efforts, calculus does not apply – we compare payoffs from e_H against payoffs from e_L.
- **What does IC add to the principal's problem?** A second constraint. So the Lagrangian will now carry *two* multipliers, one for PC and one for IC, instead of the single one we had in Chapter 2.

Section 3.2
Remodeling the Principal's Problem

The principal's problem under asymmetric information

- Setting: two effort levels $e_H > 0$ and $e_L = 0$; two outcomes π_H and π_L.
- Principal seeks to induce high effort:

$$\max_{w_H, w_L \geq 0} \underbrace{p_H \pi_H + (1 - p_H)\pi_L}_{\text{expected profit}} - \underbrace{(p_H w_H + (1 - p_H) w_L)}_{\text{expected salary}}$$

$$\text{s.t.} \quad p_H v(w_H) + (1 - p_H) v(w_L) - g(e_H) \geq \overline{u}, \quad \text{(PC)}$$

$$\begin{aligned} & p_H v(w_H) + (1 - p_H) v(w_L) - g(e_H) \\ \geq \; & p_L v(w_H) + (1 - p_L) v(w_L) - g(e_L). \quad \text{(IC)} \end{aligned}$$

- Rewriting the objective this way shows the principal cares about wages only through the **expected salary** she pays.
- New: **IC constraint** ensures the worker prefers e_H over e_L.

An economist's view before the algebra

- The principal ultimately chooses *which* effort to induce. So ask, for each effort separately, what is the cheapest contract that induces it, and only then compare profits.
- **Cheapest contract for $e_L = 0$.** The worker has no reason to deviate from a costless action, so IC is slack and only PC binds. The cheapest way to deliver $\overline{u}$ is a *flat* wage $w_H = w_L = w$ solving $v(w) = \overline{u} + g(e_L)$: no risk, no risk premium.
- **Cheapest contract for e_H.** A flat wage now fails: paid the same in both outcomes, the worker saves $g(e_H) - g(e_L) > 0$ by shirking, so IC breaks. The wage must respond to the outcome, and both PC and IC bind.

An economist's view before the algebra (cont.)

- **Why both bind for e_H.** PC binds because the principal never leaves the worker more utility than necessary. IC binds because any spread between w_H and w_L beyond what deters shirking is pure extra risk, and so pure extra cost for the firm.
- Economically, the shape of the answer is already clear: the wage is flat when effort is cheap to induce, and spreads with the outcome when it is not, with the size of the spread set by how much the worker must be paid to bear that risk.
- The Lagrangian below confirms this, and is the route we will need once there are more than two efforts or more than two outcomes.

Intuition of the IC constraint

- If $w_H > w_L$ (which we will verify below), the worker faces a trade-off:
  - **High effort:** more costly (higher disutility) but raises the probability of π_H – so the worker earns $w_H > w_L$ more often.
  - **Low effort:** cheaper but raises the probability of π_L – so the worker earns w_L more often.
- The IC inequality balances these two forces:
  - Left-hand side: expected utility from high effort.
  - Right-hand side: expected utility from low effort.
- The principal designs the wage spread so that the high-effort side is at least as attractive as the low-effort side.

When does IC bind, and why does it matter?

- If IC held with strict slack, the principal could *lower* w_H (or raise w_L) without losing the worker's incentive to choose e_H, and thereby save on the wage bill.
- So the principal drives IC to *equality* in equilibrium – the worker is indifferent between the target effort and the best deviation.

When does IC bind – The limited-liability caveat

- **Same logic as PC.** Both constraints bind because slack means money left on the table.
- **One caveat.** If the required w_L turns negative, we add a **limited-liability** floor $w_L \geq 0$. Then IC still binds, but PC may not – the worker gets a rent. We come back to this below.

The Lagrangian with two multipliers

- Build the Lagrangian as objective + $\lambda\times$(PC written as "slack ≥ 0") + $\mu\times$(IC written as "slack ≥ 0"):

$$\begin{aligned}\mathcal{L} \;=\; & p_H(\pi_H - w_H) + (1-p_H)(\pi_L - w_L) \\ & +\lambda\left[p_H v(w_H) + (1-p_H)v(w_L) - g(e_H) - \overline{u}\right] \\ & +\mu\left[p_H v(w_H) + (1-p_H)v(w_L) - g(e_H)\right. \\ & \left. -p_L v(w_H) - (1-p_L)v(w_L) + g(e_L)\right].\end{aligned}$$

  Each bracket is the left-hand side minus the right-hand side of its constraint, so it is nonnegative at any feasible contract.

- Multipliers:
  - $\lambda \geq 0$ on the PC.
  - $\mu \geq 0$ on the IC.
- We will show both bind in equilibrium ($\lambda > 0$, $\mu > 0$).

Two multipliers – Reading the Lagrangian

- Two constraints $\Rightarrow$ two multipliers, one for each:
  - $\lambda \geq 0$ on PC – the shadow price of a marginal increase in the worker's outside option.
  - $\mu \geq 0$ on IC – the shadow price of a marginal tightening in the incentive to choose e_H.
- **The multipliers tell us where the cost is.**
  - If $\lambda > 0$ and $\mu = 0$: the participation constraint is what pins wages down – the case we saw in Chapter 2.
  - If both $\lambda > 0$ and $\mu > 0$: participation *and* incentives are jointly active. This is the typical case here.
- **Preview.** We will show below that under asymmetric information both multipliers are strictly positive – and this is exactly why the wage schedule must vary with the profit realization.

First-order condition with respect to w_H

- **Step 1. Differentiate $\mathcal{L}$ in w_H.** The profit term gives $-p_H$, the PC term $\lambda p_H v'(w_H)$, and the IC term $\mu[p_H - p_L]v'(w_H)$ (the $-p_L v'$ comes from the deviation payoff on the IC's right-hand side):

$$-p_H + \lambda p_H v'(w_H) + \mu \left[p_H v'(w_H) - p_L v'(w_H)\right] = 0. \quad (3.1)$$

- **Step 2. Divide by $p_H > 0$** (it multiplies every term), which leaves

$$-1 + \lambda v'(w_H) + \mu v'(w_H)\left(1 - \frac{p_L}{p_H}\right) = 0.$$

- **Step 3. Collect the $v'(w_H)$ terms and divide by $v'(w_H) > 0$** to isolate the shadow prices:

$$\lambda + \mu\left(1 - \frac{p_L}{p_H}\right) = \frac{1}{v'(w_H)}. \quad (3.2)$$

First-order condition – Reading equation (3.2)

- Under **symmetric information** (Chapter 2) the IC was absent, so $\mu = 0$ and the condition collapses to $\lambda = 1/v'(w_H)$.
- Under **asymmetric information** the extra term $\mu\left(1 - \frac{p_L}{p_H}\right) > 0$, because $\mu \geq 0$ and $p_H > p_L$ by assumption, so the left-hand side is strictly larger:

$$\lambda + \underbrace{\mu\left(1 - \frac{p_L}{p_H}\right)}_{+} > \lambda.$$

- Since $1/v'(w_H)$ is increasing in w_H, a larger left-hand side is matched at a **higher** wage w_H.
- Intuitively, the IC forces the principal to raise w_H above its symmetric-information level to make high effort worth the worker's while.

3.2.1 Comparing symmetric and asymmetric information

- Equation (3.2) on the LHS: $\lambda + \mu(p_H - p_L)/p_H > \lambda$ when $p_H > p_L$.
- Equation (3.2) on the RHS: $1/v'(w_H)$ is increasing in w_H (since v' is weakly decreasing).
- Therefore, the crossing of LHS and RHS happens at a higher w_H under AI than under SI:

$$w_H^{AI} > w_H^{SI}.$$

- Example: with $v(w_H) = \sqrt{w_H}$, $v'(w_H) = 1/(2\sqrt{w_H})$, so $1/v'(w_H) = 2\sqrt{w_H}$, clearly increasing and concave in w_H.

Figure 3.1 – Wages under SI vs. AI

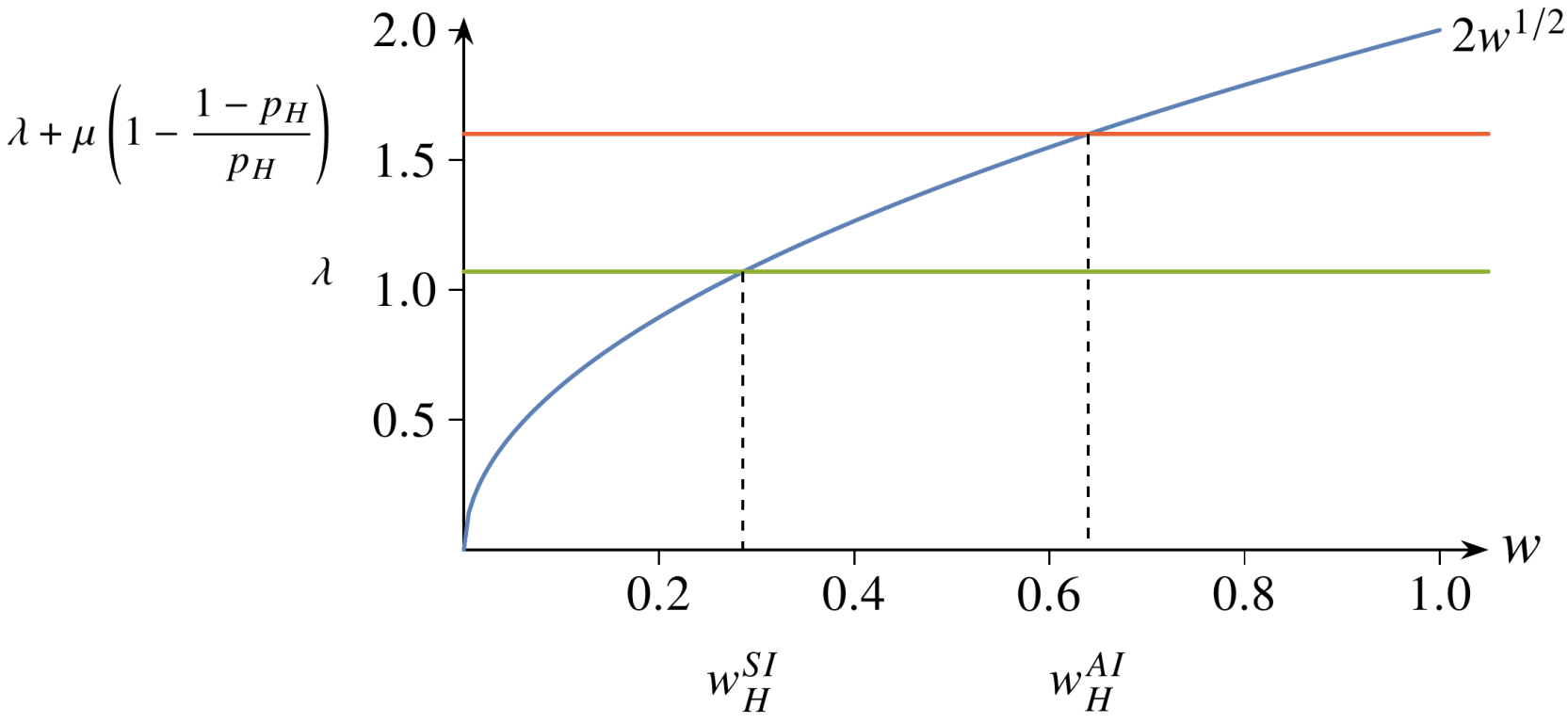


Figure 3.1. Wages under symmetric and asymmetric information. Curve $1/v'(w_H)$ is increasing in w_H. The horizontal line for AI lies above the SI line, so it crosses the curve at a higher wage.

3.2.2 The IC condition holds with equality

- **Step 1. Differentiate $\mathcal{L}$ in w_L**, symmetric to the w_H condition (here the low outcome carries weight $1 - p_H$, and the deviation weight $1 - p_L$):

$$\begin{aligned} &-(1-p_H) + \lambda(1-p_H)v'(w_L) \\ &+\mu\left[(1-p_H)v'(w_L) - (1-p_L)v'(w_L)\right] = 0. \end{aligned}$$

- **Step 2. Divide by $1 - p_H > 0$ and by $v'(w_L) > 0$**, then solve for λ:

$$\lambda = \frac{1}{v'(w_L)} - \mu\left(1 - \frac{1-p_L}{1-p_H}\right). \tag{3.3}$$

- Equations (3.2) and (3.3) both equal λ, so we can set them equal and solve for μ on the next slide.

3.2.2 The IC condition holds with equality (cont.)

- **Step 3. Equate (3.2) and (3.3)** (both give λ):

$$\frac{1}{v'(w_H)} - \mu\left(1 - \frac{p_L}{p_H}\right) = \frac{1}{v'(w_L)} - \mu\left(1 - \frac{1-p_L}{1-p_H}\right).$$

- **Step 4. Collect the μ terms and solve**, using $\frac{1-p_L}{1-p_H} - \frac{p_L}{p_H} = \frac{p_H - p_L}{p_H(1-p_H)}$:

$$\mu = \frac{\frac{1}{v'(w_H)} - \frac{1}{v'(w_L)}}{\frac{1-p_L}{1-p_H} - \frac{p_L}{p_H}} = \frac{\frac{1}{v'(w_H)} - \frac{1}{v'(w_L)}}{\frac{p_H - p_L}{p_H(1-p_H)}}. \tag{3.4}$$

- We read off the sign of μ from (3.4) next.

Why $\mu > 0$ – The IC binds

- Numerator of (3.4): if $w_H > w_L$, then $v'(w_H) < v'(w_L)$, so $1/v'(w_H) > 1/v'(w_L)$ – positive.
- Denominator of (3.4): $p_H > p_L$ by assumption – positive.
- Hence $\mu > 0 \Rightarrow$ **IC binds**.

Why $\mu > 0$ – The IC binds (cont.)

- **Intuition:** if IC held with strict inequality, the principal could lower w_H slightly, still inducing high effort but saving on wages.
- "Money on the table" – leaving slack in IC means the principal could extract more surplus.
- **Both PC and IC bind** at the optimum.

Interpreting the multiplier μ

- Recall λ: shadow value of relaxing PC (allowing the agent to require less utility).
- μ: shadow value of relaxing IC (allowing the agent to require less incentive).
- Equation (3.4) says μ is **larger** when:
  - The wage spread $\frac{1}{v'(w_H)} - \frac{1}{v'(w_L)}$ is large – agent already faces strong incentives.
  - Probabilities p_H and p_L are similar – effort doesn't shift the distribution much.

Interpreting the multiplier μ (cont.)

- Intuition: when effort barely changes outcome probabilities, satisfying IC requires a very large wage spread, which is costly.
- $\mu \to 0$ recovers the symmetric-information case (no IC binding).

The risk-sharing / incentives trade-off

- **Two forces pull in opposite directions:**
  - Risk-sharing wants the wage as *flat* as possible – ideally $w_H = w_L$.
  - Incentives want a *wide* gap $w_H - w_L$, so the worker cares whether the good state occurs.
- **The optimal contract balances the two:** pays some performance-linked variation, but only as much as IC requires – every extra unit of variation is costly to the risk-averse worker.
- **Second-best.** Pareto-inferior to Chapter 2's first-best: the worker faces income risk that would not exist under symmetric information.

Limited liability – A modeling note

- In many real settings, wages must satisfy $w \geq 0$ (limited liability).
- Implications:
  - The principal cannot punish bad outcomes with negative wages.
  - Spread is constrained from below.
  - PC may not bind – the worker may earn rents above $\overline{u}$.
- Common in:
  - Tenant farming (Exercise 3.1).
  - Lending (borrowers cannot owe more than they earn after default).
  - Employment law (minimum wage as a soft form of limited liability).

Limited liability – Consequences for the contract

- In Chapters 3–4, we generally allow w unrestricted, deferring limited-liability analysis to specific exercises.

3.2.3 Roadmap – A 5-step recipe

1. **Use IC binding**: solve for w_H to obtain a function $w_H = f(w_L, p_H, p_L)$.
2. **Use PC binding**: solve for w_H to obtain a function $w_H = g(w_L, p_H, p_L)$.
3. **Equate** $f(w_L, p_H, p_L) = g(w_L, p_H, p_L)$ and solve for w_L^{AI}.

- With two effort levels and two outcomes, we have 2 equations (binding PC and IC) and 2 unknowns (w_H, w_L) – so the principal's objective is not even needed in steps 1–4.

Roadmap – Steps 4–5 and applications

1. **Back-substitute** w_L^{AI} into either binding constraint to find w_H^{AI}.
2. **Check** that the principal prefers to induce high effort by comparing with the low-effort net profit.

3.2.4 Monotonic salaries – When is $w_H > w_L$?

- From equation (3.2): the RHS $1/v'(w_H)$ depends on w_H, the LHS depends on p_H.
- For $w_H > w_L$, the LHS evaluated at high profits must exceed the LHS evaluated at low profits:

$$\lambda + \mu\left(1 - \frac{p_L}{p_H}\right) > \lambda + \mu\left(1 - \frac{1-p_L}{1-p_H}\right).$$

- Simplifying:

$$1 - \frac{p_L}{p_H} > 1 - \frac{1-p_L}{1-p_H} \iff p_L < p_H.$$

- Recall: $p_H = f(\pi_H|e_H)$, $p_L = f(\pi_H|e_L)$.
- $p_L < p_H$ means a high profit is more likely under high than under low effort.

Figure 3.2 – Wages increasing in profit realizations

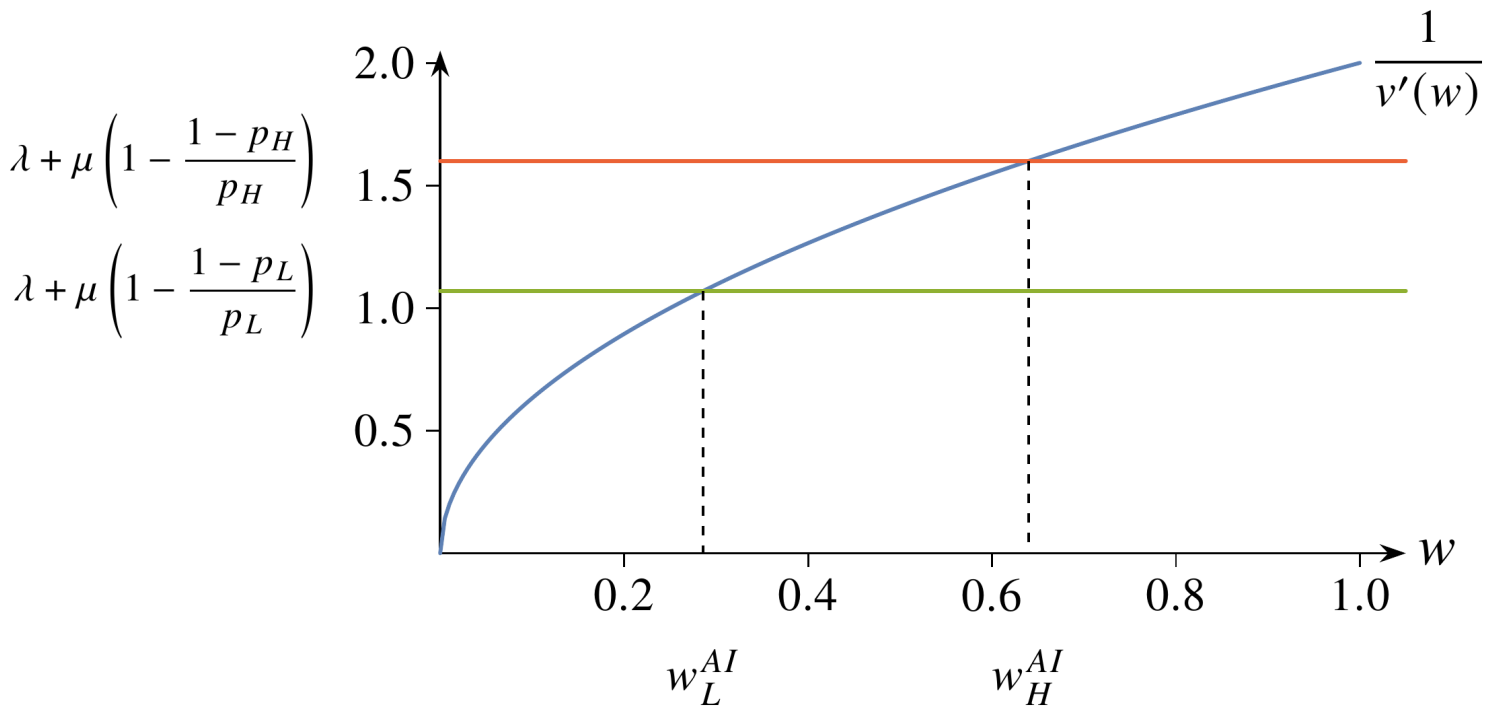


Figure 3.2. Curve $1/v'(w)$ crosses the (higher) horizontal line $\lambda + \mu(1 - (1 - p_H)/p_H)$ at the higher wage w_H^{AI}, and the (lower) horizontal line $\lambda + \mu(1 - (1 - p_L)/p_L)$ at the lower wage w_L^{AI}.

Monotone Likelihood Ratio Property (MLRP)

- The ranking $1 - \frac{1-p_H}{p_H} > 1 - \frac{1-p_L}{p_L}$ is known as the **MLRP**.
- Using $p_H = f(\pi_H|e_H)$ and $p_L = f(\pi_L|e_H)$ (and similarly for e_L), it can be rewritten as:

$$\frac{f(\pi_H|e_L)}{f(\pi_H|e_H)} < \frac{f(\pi_L|e_L)}{f(\pi_L|e_H)}. \quad \text{(MLRP)}$$

- **Interpretation:** as profits rise from π_L to π_H, the relative likelihood of high effort to low effort **rises**.
- In other words, higher profits are a more credible signal of high effort.
- Under MLRP, optimal wages are monotonically increasing in observed profits.

Reading MLRP – Good news vs. bad news

- MLRP says: higher outcomes are *more* evidence that the worker chose high effort.
- Imagine the principal sees π_H. She asks: which effort was more likely to have produced this? MLRP guarantees the answer is "high effort."
- So the principal rewards π_H with a high wage – not because π_H mechanically means good effort, but because π_H is *more informative* of good effort than π_L is.

Reading MLRP – Why the ordering matters

- **Consequence for the contract.** Under MLRP, the optimal wage schedule is *monotone* in profit: $w_H > w_L$. Without MLRP, the ranking can flip.
- **Not obvious!** The result is not "bigger profit, bigger wage." It is "more-informative-of-effort profit, bigger wage," and these coincide only when MLRP holds.

Distributions satisfying MLRP

- MLRP is a mild condition – many standard families satisfy it:
  - Normal distributions $N(\mu(e), \sigma^2)$ with $\mu(e)$ strictly increasing in e.
  - Exponential families with monotone parameter $\theta(e)$.
  - Beta distributions $\text{Beta}(\alpha(e), \beta)$ with $\alpha(e)$ increasing in e.
- **When MLRP fails:** optimal wages may be *non-monotonic* – the principal might pay more after a medium outcome than a high outcome.
- Practical example: when the highest profit might come from luck rather than effort, large bonuses for top performance can be inefficient.

Distributions satisfying MLRP (cont.)

- In economic terms: MLRP $\Leftrightarrow$ "higher profit is better news about effort."

MLRP in our two-outcome example

- In Example 3.1: $f(\pi_H|e_H) = \frac{1}{2}$, $f(\pi_H|e_L) = \frac{1}{4}$, $f(\pi_L|e_H) = \frac{1}{2}$, $f(\pi_L|e_L) = \frac{3}{4}$.
- Check MLRP:

$$\frac{f(\pi_H|e_L)}{f(\pi_H|e_H)} = \frac{1/4}{1/2} = \frac{1}{2}$$

$$\frac{f(\pi_L|e_L)}{f(\pi_L|e_H)} = \frac{3/4}{1/2} = \frac{3}{2}.$$

- Since $\frac{1}{2} < \frac{3}{2}$, MLRP holds.
- Therefore $w_H^{AI} > w_L^{AI}$ – which is exactly what we found ($\$576 > \$16$).

Example 3.1 – Setup

- Recall Example 2.2: risk-neutral principal, risk-averse worker $v(w) = \sqrt{w}$, $g(e) = e$, $\overline{u} = 9$, $e_H = 5$, $e_L = 0$, $\pi_H = \$1{,}200$, $\pi_L = \$300$.
- Conditional probabilities: $p_H = \frac{1}{2}$, $p_L = \frac{1}{4}$.
- Now the principal cannot observe effort – she solves:

$$\max_{w_H, w_L \geq 0} \ \tfrac{1}{2}(1{,}200 - w_H) + \tfrac{1}{2}(300 - w_L)$$

$$\text{s.t.} \quad \tfrac{1}{2}\sqrt{w_H} + \tfrac{1}{2}\sqrt{w_L} - 5 \geq 9, \quad \text{(PC)}$$

$$\tfrac{1}{2}\sqrt{w_H} + \tfrac{1}{2}\sqrt{w_L} - 5 \geq \tfrac{1}{4}\sqrt{w_H} + \tfrac{3}{4}\sqrt{w_L}. \quad \text{(IC)}$$

Example 3.1 – Simplified problem

- Simplify the PC: move the -5 across and multiply by 2, giving $\sqrt{w_H} + \sqrt{w_L} \geq 28$.
- Simplify the IC: move the deviation terms to the left, so $\left(\frac{1}{2} - \frac{1}{4}\right)\sqrt{w_H} + \left(\frac{1}{2} - \frac{3}{4}\right)\sqrt{w_L} \geq 5$, i.e., $\frac{1}{4}\left(\sqrt{w_H} - \sqrt{w_L}\right) \geq 5$.
- Problem becomes:

$$\max_{w_H, w_L \geq 0} \; 750 - \tfrac{1}{2}(w_H + w_L)$$

$$\text{s.t.} \quad \sqrt{w_H} + \sqrt{w_L} \geq 28, \quad \text{(PC)}$$

$$\sqrt{w_H} - \sqrt{w_L} \geq 20. \quad \text{(IC)}$$

- Both PC and IC bind. Now apply the 5-step roadmap.

Example 3.1 – Steps 1 & 2

- **Step 1 – IC binds:** $\sqrt{w_H} - \sqrt{w_L} = 20$, so

$$w_H = \left(20 + \sqrt{w_L}\right)^2 .$$

- **Step 2 – PC binds:** $\sqrt{w_H} + \sqrt{w_L} = 28$, so

$$w_H = \left(28 - \sqrt{w_L}\right)^2 .$$

- Now equate the two expressions in Step 3.

Example 3.1 – Steps 3 & 4

- **Step 3 – Equate:**

$$(20 + \sqrt{w_L})^2 = (28 - \sqrt{w_L})^2.$$

- Taking square roots (both positive): $20 + \sqrt{w_L} = 28 - \sqrt{w_L}$.
- Rearranging: $2\sqrt{w_L} = 8$, so $\sqrt{w_L} = 4$.
- Therefore $w_L^{AI} = \$16$.
- **Step 4 – Back-substitute:** from Step 1,

$$w_H^{AI} = (20 + 4)^2 = 24^2 = \$576.$$

- **Check:** $\sqrt{576} - \sqrt{16} = 24 - 4 = 20$ (binding IC) and $\sqrt{576} + \sqrt{16} = 24 + 4 = 28$ (binding PC). Both hold.

Example 3.1 – Step 5 (verify high effort)

- Principal's expected profit from high effort:

$$\tfrac{1}{2}(1{,}200 - 576) + \tfrac{1}{2}(300 - 16) = \$454.$$

- Principal's expected profit from low effort (constant wage $w = \$81$, from Chapter 2):

$$\tfrac{1}{4}(1{,}200 - 81) + \tfrac{3}{4}(300 - 81) = \$444.$$

- Since $\$454 > \$444$, the principal optimally induces e_H with the wage pair:

$$(w_H^{AI}, w_L^{AI}) = (\$576, \$16).$$

- **Wage spread** $w_H^{AI} > w_L^{AI}$ provides the incentive to exert high effort.

3.2.5 Asymmetric vs. symmetric wages

- Compare with the symmetric-information results from Chapter 2:

$$\begin{aligned} w_H^{AI} &= \$576 > \$196 = w_H^{SI}, \\ w_L^{AI} &= \$16 < \$196 = w_L^{SI}. \end{aligned}$$

- Worker gets **more** when profits are high, **less** when profits are low.
- Worker's expected utility is unchanged (both PC bindings give $u = \overline{u} = 9$):

$$u_H^{AI} = \tfrac{1}{2}\sqrt{576} + \tfrac{1}{2}\sqrt{16} - 5 = 12 + 2 - 5 = 9.$$

- Same as $u_H^{SI} = \sqrt{196} - 5 = 14 - 5 = 9$.

Profit loss due to moral hazard

- Principal's profit is **lower** under asymmetric information:

$$\pi_H^{AI} = \$454 < \$554 = \pi_H^{SI}.$$

- **Loss from moral hazard:**

$$\pi_H^{SI} - \pi_H^{AI} = \$554 - \$454 = \$100.$$

- Why? Wage volatility is **inefficient risk-sharing** – the risk-averse worker dislikes uncertain wages.
- To deliver the same expected utility $\overline{u}$ while spreading wages, the principal must **pay more in expectation**.
- This \$100 loss is the **cost of providing incentives**.

How large is the moral-hazard loss?

- The gap between the principal's first-best profit (Chapter 2) and second-best profit (this chapter) is a *welfare cost* of asymmetric information – the price the principal pays for not observing effort.
- **Two components:**
  - A *risk premium* paid to compensate the worker for carrying income risk.
  - In some parametrizations, a *lower induced effort* – the principal may find high effort no longer worth inducing.
- The size of the loss grows with the worker's risk aversion and with the noise in the effort-profit relationship (small $p_H - p_L$).

How large is the moral-hazard loss? (cont.)

- **Bottom line.** Asymmetric information is not just a mathematical inconvenience – it makes the whole relationship less valuable. The empirical evidence from Chapter 2 (Bandiera-Barankay-Rasul, Lazear) shows that this loss is real and measurable.

Section 3.3
More Effort Levels and More Outcomes

More profit realizations – A preview

- With three profit levels and two efforts, write $p_i = f(\pi_i|e_H)$ and $q_i = f(\pi_i|e_L)$ for $i \in \{H, M, L\}$, each set summing to one. The PC keeps its shape:

$$p_H v(w_H) + p_M v(w_M) + p_L v(w_L) - g(e_H) \geq \overline{u}.$$

- The IC compares high effort against low effort, now over three outcomes:

$$\begin{aligned} & p_H v(w_H) + p_M v(w_M) + p_L v(w_L) - g(e_H) \\ \geq \quad & q_H v(w_H) + q_M v(w_M) + q_L v(w_L) - g(e_L). \end{aligned}$$

- Same shape as the two-outcome IC, but with N wage terms and the low-effort probabilities q_i on the right.

More profit realizations – Why the roadmap is not enough

- **Key difference:** the roadmap (steps 1–4) does not pin down all wages – only relations between them.
- Why? With N outcomes there are N wages but still only 2 binding constraints (PC + IC), so $N - 2$ degrees of freedom remain.
- Resolution: substitute the binding PC and IC into the principal's objective and minimize over the remaining $N - 2$ wages.
- Intuitively, with only two constraints the principal can shift pay across the extra outcomes however is cheapest, so her objective, not the constraints, decides those wages.
- Worked through in Example 3.2 below ($N = 3$).

3.3.1 More than two effort levels

- Now three effort levels: e_H, e_M, e_L but still two profit realizations.
- Probabilities (Table 3.1):

		Profits	
		π_H	π_L
Effort	e_H	$p_1 = \frac{5}{6}$	$1 - p_1 = \frac{1}{6}$
	e_M	$q_1 = \frac{1}{4}$	$1 - q_1 = \frac{3}{4}$
	e_L	$r_1 = \frac{1}{12}$	$1 - r_1 = \frac{11}{12}$

Table 3.1. Three effort levels, two outcomes.

Three effort levels – Two IC constraints

- Principal's problem (assuming risk-neutral, $v(w) = \sqrt{w}$, inducing e_H):

$$\max_{w_H, w_L} \; p_1 \pi_H + (1 - p_1)\pi_L - (p_1 w_H + (1 - p_1) w_L)$$

$$\text{s.t. } p_1 \sqrt{w_H} + (1 - p_1)\sqrt{w_L} - e_H \geq \overline{u}, \quad \text{(PC)}$$

$$p_1 \sqrt{w_H} + (1 - p_1)\sqrt{w_L} - e_H \geq q_1 \sqrt{w_H} + (1 - q_1)\sqrt{w_L} - e_M, \; (IC_M)$$

$$p_1 \sqrt{w_H} + (1 - p_1)\sqrt{w_L} - e_H \geq r_1 \sqrt{w_H} + (1 - r_1)\sqrt{w_L} - e_L. \; (IC_L)$$

- Now **two** IC constraints – one for each alternative effort.

Simplified form – Minimization

- Equivalently, minimize expected wage:

$$\min_{w_H, w_L} \; p_1 w_H + (1 - p_1) w_L$$

$$\text{s.t. } p_1 \sqrt{w_H} + (1 - p_1)\sqrt{w_L} \geq \bar{u} + e_H, \quad \text{(PC)}$$

$$\sqrt{w_H} - \sqrt{w_L} \geq \frac{e_H - e_M}{p_1 - q_1}, \quad (IC_M)$$

$$\sqrt{w_H} - \sqrt{w_L} \geq \frac{e_H - e_L}{p_1 - r_1}. \quad (IC_L)$$

- Each IC divides by its probability gap ($p_1 - q_1$ or $p_1 - r_1$) to isolate the spread $\sqrt{w_H} - \sqrt{w_L}$: same left-hand side, different right – two possible binding cases.

Two cases – Which IC binds?

- **Case 1:** $\dfrac{e_H - e_M}{p_1 - q_1} \geq \dfrac{e_H - e_L}{p_1 - r_1}$.
  - IC_M binds, IC_L is slack.
  - Worker is more tempted to deviate to e_M than to e_L.
  - Principal only needs to prevent the e_M deviation.
- **Case 2:** $\dfrac{e_H - e_M}{p_1 - q_1} < \dfrac{e_H - e_L}{p_1 - r_1}$.
  - IC_L binds, IC_M is slack.
  - Worker is more tempted to deviate to e_L.
  - Principal only needs to prevent the e_L deviation.
- In either case: 2 binding constraints (PC + one IC), 2 unknowns – same approach as Chapter 3.2.

Three effort levels – Case 2 ($e_H = 5$)

- Effort levels $e_L = 0$, $e_M = 3$, $e_H = 5$; reservation $\overline{u} = 9$.
- **Step 1. Classify the case.** The Case-2 cutoff is $e_H < \frac{9e_M - 7e_L}{2}$. Here

$$5 < \frac{(9 \times 3) - (7 \times 0)}{2} = 13.5,$$

  so we are in Case 2, where IC_L binds and IC_M is slack.
- **Step 2. Binding PC.** From $\frac{5}{6}\sqrt{w_H} + \frac{1}{6}\sqrt{w_L} = \overline{u} + e_H = 14$, multiply by 6 to get $5\sqrt{w_H} + \sqrt{w_L} = 84$, hence

$$\sqrt{w_L} = 84 - 5\sqrt{w_H}.$$

Three effort levels – Case 2 ($e_H = 5$) (cont.)

- **Step 3. Binding** IC_L. The spread is $\sqrt{w_H} - \sqrt{w_L} = \frac{e_H - e_L}{p_1 - r_1} = \frac{5}{3/4} = \frac{20}{3}$, i.e., $\sqrt{w_L} = \sqrt{w_H} - \frac{20}{3}$.
- **Step 4. Equate and solve.** Setting the two expressions for $\sqrt{w_L}$ equal,

$$84 - 5\sqrt{w_H} = \sqrt{w_H} - \frac{20}{3} \;\Rightarrow\; \sqrt{w_H} = \frac{136}{9},$$

  so $w_H = \frac{18{,}496}{81} \simeq \$228.35$ and, from Step 2, $\sqrt{w_L} = \frac{76}{9}$, giving $w_L = \left(\frac{76}{9}\right)^2 \simeq \$71.31$.
- **Check:** $\sqrt{w_H} - \sqrt{w_L} = \frac{136}{9} - \frac{76}{9} = \frac{20}{3}$, matching binding IC_L.

Three effort levels – Case 1 ($e_H = 15$)

- Now $e_L = 0$, $e_M = 3$, $e_H = 15$.
- **Step 1. Classify the case.** The cutoff now fails, $15 > \frac{(9\times 3)-(7\times 0)}{2} = 13.5$, so we are in Case 1, where IC_M binds and IC_L is slack.
- **Step 2. Binding PC.** From $\frac{5}{6}\sqrt{w_H} + \frac{1}{6}\sqrt{w_L} = \overline{u} + e_H = 24$, we get $5\sqrt{w_H} + \sqrt{w_L} = 144$, hence

$$\sqrt{w_L} = 144 - 5\sqrt{w_H}.$$

(It differs from Case 2 because e_H rose from 5 to 15.)

Three effort levels – Case 1 ($e_H = 15$) (cont.)

- **Step 3. Binding** IC_M. The spread is $\sqrt{w_H} - \sqrt{w_L} = \frac{e_H - e_M}{p_1 - q_1} = \frac{12}{7/12} = \frac{144}{7}$, i.e., $\sqrt{w_L} = \sqrt{w_H} - \frac{144}{7}$.
- **Step 4. Equate and solve.**

$$144 - 5\sqrt{w_H} = \sqrt{w_H} - \frac{144}{7} \;\Rightarrow\; \sqrt{w_H} = \frac{192}{7},$$

so $w_H = \left(\frac{192}{7}\right)^2 \simeq \$752.33$ and $\sqrt{w_L} = \frac{48}{7}$, giving $w_L = \left(\frac{48}{7}\right)^2 \simeq \$47.02$.

- **Check:** $\sqrt{w_H} - \sqrt{w_L} = \frac{192}{7} - \frac{48}{7} = \frac{144}{7}$, matching binding IC_M.

Table 3.2 – Comparing Cases 1 and 2

		Case 1	Case 2
Salaries	w_H	\$752.33	\$228.35
	w_L	\$47.02	\$71.31

Table 3.2. Salaries in Cases 1 and 2.

- Case 1 has a **steeper wage spread**: a higher w_H after a good outcome, but a lower w_L after a bad outcome.
- e_H is three times larger (15 vs. 5) $\Rightarrow$ a larger spread is needed to deter deviation to e_M.

Three effort levels – Does the principal induce e_H?

- Assign $\pi_H = \$1{,}200$ and $\pi_L = \$300$. Inducing $e_L = 0$ needs only a flat wage $w = \overline{u}^2 = \$81$, so

$$E[\pi|e_L] = \tfrac{1}{12}(1{,}200) + \tfrac{11}{12}(300) - 81 = \$294.$$

- Inducing e_H with the Table 3.2 wages yields, in Case 1,

$$E[\pi|e_H] = \tfrac{5}{6}(1{,}200 - 752.33) + \tfrac{1}{6}(300 - 47.02) = \$415.22,$$

and in Case 2,

$$E[\pi|e_H] = \tfrac{5}{6}(1{,}200 - 228.35) + \tfrac{1}{6}(300 - 71.31) = \$847.82,$$

both well above \$294.

- So the principal induces high effort in either case. Intuitively, a higher effort raises the chance of high profit by enough to justify its more generous salary.

3.3.2 More than two outcomes – Example 3.2 setup

- Two efforts $e_H = 5, e_L = 0$, three outcomes (Table 3.3):

		Profits		
		$\pi_H = \$1{,}200$	$\pi_M = \$300$	$\pi_L = \$0$
Effort	$e_H = 5$	0.5	0.25	0.25
	$e_L = 0$	0.2	0.25	0.55

Table 3.3. Conditional probabilities – Example 3.2.

- Expected profits:

$$\begin{aligned} E\pi(e_H) &= 0.5(1{,}200) + 0.25(300) + 0.25(0) = \$675, \\ E\pi(e_L) &= 0.2(1{,}200) + 0.25(300) + 0.55(0) = \$315. \end{aligned}$$

Example 3.2 – Principal's problem

- Inducing high effort (expected profit $E\pi(e_H) = \$675$, less expected salary):

$$\max_{w_H, w_M, w_L} \; 675 - (0.5w_H + 0.25w_M + 0.25w_L)$$

$$\text{s.t. } 0.5\sqrt{w_H} + 0.25\sqrt{w_M} + 0.25\sqrt{w_L} - 5 \geq 9, \quad \text{(PC)}$$

$$0.5\sqrt{w_H} + 0.25\sqrt{w_M} + 0.25\sqrt{w_L} - 5$$

$$\geq 0.2\sqrt{w_H} + 0.25\sqrt{w_M} + 0.55\sqrt{w_L}. \quad \text{(IC)}$$

Example 3.2 – Simplified problem

- Maximizing 675 minus expected salary is the same as minimizing expected salary, and the $0.25\sqrt{w_M}$ terms cancel from the IC:

$$\min_{w_H, w_M, w_L} \ \frac{2w_H + w_M + w_L}{4}$$

$$\text{s.t. } 2\sqrt{w_H} + \sqrt{w_M} + \sqrt{w_L} \geq 56, \quad \sqrt{w_H} - \sqrt{w_L} \geq \frac{50}{3}.$$

- The PC multiplies through by 4; the IC, after cancelling w_M and dividing by $0.3 = p_H - q_H$, leaves the spread $\sqrt{w_H} - \sqrt{w_L} \geq \frac{50}{3}$.
- Notice w_M drops out of the IC entirely – it is pinned down only by the objective, as previewed for the N-outcome case.

Example 3.2 – Steps 1 & 2

- **Step 1.** IC binds: $\sqrt{w_H} = \sqrt{w_L} + \frac{50}{3}$. Writing the right side over 3 and squaring both sides,

$$w_H = \frac{(3\sqrt{w_L} + 50)^2}{9}.$$

- **Step 2.** Substitute into principal's objective and into PC:

$$2 \cdot \frac{3\sqrt{w_L} + 50}{3} + \sqrt{w_M} + \sqrt{w_L} = 56.$$

- Rearranging: $\sqrt{w_M} + 3\sqrt{w_L} = \frac{68}{3}$, so $w_M = \left(\frac{68}{3} - 3\sqrt{w_L}\right)^2$.
- Substituting $w_M = \frac{(68-9\sqrt{w_L})^2}{9}$ into the objective leaves a single-variable minimization in w_L.

Example 3.2 – Steps 3 & 4

- **Step 3.** Minimize over w_L:

$$\min_{w_L} \frac{2(3\sqrt{w_L}+50)^2+(68-9\sqrt{w_L})^2+9w_L}{36}.$$

- Differentiating and setting to zero yields

$$\frac{9\sqrt{w_L}-26}{3\sqrt{w_L}}=0 \Rightarrow \sqrt{w_L}=\frac{26}{9},$$

since a fraction is zero only when its numerator is (and $3\sqrt{w_L}>0$), so $w_L=\frac{676}{81}\simeq \$8.35$.

- **Step 4.** Back-substitute:

$$\begin{aligned} w_M &= \left(\frac{68-26}{3}\right)^2 = 14^2 = \$196, \\ w_H &= \frac{(176/3)^2}{9} = \frac{30{,}976}{81} \simeq \$382.42. \end{aligned}$$

Example 3.2 – Step 5 (check principal's choice)

- Triplet: $(w_L, w_M, w_H) = (\$8.35, \$196, \$382.42)$. Check binding IC: $\sqrt{w_H} - \sqrt{w_L} = \frac{176}{9} - \frac{26}{9} = \frac{150}{9} = \frac{50}{3}$.
- Expected profit from high effort:

$$E\pi(e_H) = 675 - \frac{2 \cdot \frac{30{,}976}{81} + 196 + \frac{676}{81}}{4} = \frac{11{,}683}{27} \simeq \$432.7.$$

- Expected profit from low effort (constant wage $w = 81$):

$$E\pi(e_L) = 315 - 81 = \$234.$$

- Principal optimally induces high effort.
- **Observation:** wages are increasing in the profit realization: $w_H > w_M > w_L$ (MLRP holds).

Comparative statics under asymmetric information

	Measured by	Effect on w_H	Effect on w_L
Reservation utility	$\overline{u}$	$+$	$+$
Effort level	e	$+$	$-$
Prob. high profit (high effort)	p	$-$	$+$
Prob. high profit (low effort)	q	$+$	$-$
Agent's risk aversion	r_W	$-$	$+$
Signal responsiveness	$p - q$	$-$	$+$

Table 3.4. Comparative statics under asymmetric information.

Reading the comparative statics table

- $\overline{u} \uparrow$: principal must pay more in both states to keep the worker.
- $e \uparrow$: higher effort demanded $\Rightarrow$ higher w_H; but worker punished more for low outcome ($w_L \downarrow$).
- $p \uparrow$ (high profit becomes more likely under high effort): spread can shrink – $w_H \downarrow$, $w_L \uparrow$.

Reading the comparative statics table (cont.)

- $q \uparrow$ (high profit also likely under low effort): high profit becomes less informative of effort – need a larger spread, $w_H \uparrow$, $w_L \downarrow$.
- Worker more risk-averse: $w_H \downarrow$ but $w_L \uparrow$.
- Responsiveness $p - q \uparrow$: signal more informative, spread shrinks, $w_H \downarrow$, $w_L \uparrow$.

Section 3.4
Moral Hazard Applications

3.4.1 Credit rationing – Setup

- Bank (lender) and $N \geq 1$ firms; all risk-neutral.
- Each firm can invest $\$I$ in either a **safe** project (S) or a **risky** project (R).
- Returns: $X_R > X_S$ but $p_S X_S > p_R X_R$ since $p_S > p_R$.

Credit rationing – Setup (cont.)

- Both projects profitable in expectation: $p_i X_i > I$ for all $i \in \{R, S\}$.
- Default occurs with probability $1 - p_i$ – firm makes no money, defaults on principal and interest.
- Bank charges interest $r \in (0, 1)$; total revenue $Rev = (1 + r)I$.
- Loan capacity: $L > I$ but $N \cdot I > L$ – bank cannot fund all firms.

Credit rationing – Objective functions

- **Bank's profit** from project i:

$$\pi_i = p_i Rev - I = p_i(1+r)I - I.$$

- Bank earns the loan back (with interest) only if the project succeeds.
- **Firm's expected payoff** from project i:

$$u_i = p_i \left[X_i - (1+r)I\right].$$

- Term in brackets: net return after paying back the loan.

Credit rationing – Symmetric information

- Benchmark: bank observes which project the firm chooses.
- Bank's problem:

$$\max \ \pi_i = p_i Rev - I, \quad \text{s.t. } p_i(X_i - Rev) \geq 0. \ \ \text{(PC)}$$

- PC binds: $X_i = Rev$. Unconstrained: max $p_i X_i - I$.

Credit rationing – Symmetric information (cont.)

- Bank has enough funds for one loan – choice is **binary**.
- Since $p_S X_S > p_R X_R$, the bank funds the **safe** project with $Rev = p_S X_S$.
- Firm earns zero surplus; **no credit rationing** – every firm that asks for a loan to fund a safe project gets one.

Credit rationing – Asymmetric info, firm's stage

- Bank cannot observe project choice – moral hazard arises.
- Solve by backward induction; firm moves last.
- Firm chooses safe over risky iff

$$p_S(X_S - Rev) \geq p_R(X_R - Rev).$$

Credit rationing – Asymmetric info, firm's stage (cont.)

- Expanding $p_S(X_S - Rev) \geq p_R(X_R - Rev)$ gives $p_S X_S - p_R X_R \geq (p_S - p_R)Rev$, and dividing by $p_S - p_R > 0$ yields

$$Rev \leq \frac{p_S X_S - p_R X_R}{p_S - p_R} \equiv \overline{Rev}.$$

- If Rev is too high, firm switches to the risky project.
- Comparative statics: $\overline{Rev}$ rises in p_S and X_S (safe more attractive), falls in p_R and X_R (risky more attractive).

Credit rationing – Bank's stage

- Anticipating firm's choice:

$$\pi = \begin{cases} p_S Rev - I & \text{if } Rev \leq \overline{Rev}, \\ p_R Rev - I & \text{if } \overline{Rev} < Rev \leq X_R. \end{cases}$$

- Two cases:
  1. If $p_R X_R > p_S \overline{Rev}$: bank sets $Rev = X_R$, induces risky project, firm gets 0. **No credit rationing.**
  2. Otherwise: bank sets $Rev = \overline{Rev}$, induces safe project. Firm's expected profit:

$$p_S(X_S - \overline{Rev}) = \frac{p_S p_R (X_R - X_S)}{p_S - p_R} > 0.$$

- In Case 2, all firms want a loan but $L < NI \Rightarrow$ **credit rationing**.

Example 3.3 – Credit rationing numerical

- $N = 10$ firms, $I = 0.2$, $X_S = 2$, $X_R = 3$, $p_S = 0.8$, $p_R = 0.3$, $L = 1$.
- Check $p_S X_S > p_R X_R$: $0.8 \times 2 = 1.6 > 0.9 = 0.3 \times 3$. ✓
- Cutoff $\overline{Rev} = \frac{0.8 \times 2 - 0.3 \times 3}{0.8 - 0.3} = \frac{0.7}{0.5} = 1.4$.

Example 3.3 – Credit rationing numerical (cont.)

- Bank profit at $Rev = \overline{Rev}$: $0.8 \times 1.4 - 0.2 = 0.92$.
- Bank profit at $Rev = X_R$: $0.3 \times 3 - 0.2 = 0.7$.
- Bank prefers $Rev = \overline{Rev}$ to induce safe project.
- Firms still accept, since $p_S(X_S - \overline{Rev}) = 0.8(2 - 1.4) = 0.48 > 0$.
- Total loan demand $= N \cdot I = 2 > 1 = L$, so only **5 of the 10 firms** get funded – **credit rationing** emerges.

Credit rationing – Comparative statics of $\overline{Rev}$

- Cutoff $\overline{Rev} = \dfrac{p_S X_S - p_R X_R}{p_S - p_R}$.
- Partial derivatives:

$$\begin{aligned}
\frac{\partial \overline{Rev}}{\partial p_S} &= \frac{p_R(X_R - X_S)}{(p_S - p_R)^2} > 0, \\
\frac{\partial \overline{Rev}}{\partial p_R} &= -\frac{p_S(X_R - X_S)}{(p_S - p_R)^2} < 0, \\
\frac{\partial \overline{Rev}}{\partial X_S} &= \frac{p_S}{p_S - p_R} > 0, \quad \frac{\partial \overline{Rev}}{\partial X_R} = -\frac{p_R}{p_S - p_R} < 0.
\end{aligned}$$

- **Intuition:** as the safe project becomes *more attractive* (higher p_S or X_S), or the risky project becomes less attractive (lower p_R or X_R), the bank can charge higher revenue and still induce the firm to pick safe.

Credit rationing – Policy implications

- Excess demand for credit at the equilibrium interest rate is a **persistent feature**, not a temporary disequilibrium.
- Banks rationally refuse to raise rates because doing so would push firms toward risky projects.
- **Stiglitz-Weiss (1981)** formalized this insight: credit rationing in markets with imperfect information.

Credit rationing – Policy implications (cont.)

- Policy implications:
  - Collateral requirements can substitute for monitoring.
  - Government loan guarantees can crowd in private lending to projects that would otherwise be rationed.
  - Bank capital regulation: limits incentive to make risky loans.

Credit rationing – Why the model matters

- The credit-rationing story is the earliest and cleanest application of moral hazard outside the labor market.
- **Empirical anchor.** Stiglitz and Weiss (1981) showed that banks may *refuse* to raise interest rates even when demand for loans exceeds supply – because a higher rate attracts riskier borrowers and induces safer borrowers to take riskier projects.

Credit rationing – Reading it as a contract

- **Reading it as a contract.** The interest rate is the principal's wage schedule; the borrower's project choice is the hidden action; collateral is a design tool that partially converts the hidden action into a payoff.
- **Same math as Chapter 2, different clothing.** A risk-neutral bank contracting with a risk-averse borrower reproduces the incentives-vs-risk-sharing trade-off we already know – only with loans instead of wages and default instead of low effort.

3.4.2 Moral hazard in insurance – Setup

- Insurance company sets premium p and compensation z paid in event of accident.
- Insured client exerts effort $e \in \{e_H, e_L\}$, with accident probabilities $q_H < q_L$.
- Initial wealth w_0; loss L if accident.

Moral hazard in insurance – Setup (cont.)

- Agent's utility: $u(w) - c(e)$ with $c(e_H) > c(e_L) \geq 0$; reservation utility $\overline{u}$.
- Wealth in each state:
  - No accident: $w_0 - p$.
  - Accident: $w_0 - p - L + z$.
- Question: can the insurer set $z = L$ (full coverage)?

Insurance – PC and IC for high effort

- **PC (inducing high effort):** careful driving must beat the outside option,

$$q_H u(w_0 - p - L + z) + (1 - q_H) u(w_0 - p) - c(e_H) \geq \overline{u}.$$

- **IC (high vs. low effort):** careful driving must beat careless driving,

$$\begin{aligned} & q_H u(w_0 - p - L + z) + (1 - q_H) u(w_0 - p) - c(e_H) \\ \geq \quad & q_L u(w_0 - p - L + z) + (1 - q_L) u(w_0 - p) - c(e_L). \end{aligned}$$

- Both sides share the same two wealth levels; only the accident probabilities and the effort costs differ.

Insurance – Rearranging the IC

- Collect the accident term $u(w_0 - p - L + z)$ and the no-accident term $u(w_0 - p)$ on each side. Their coefficients combine into $q_L - q_H$, giving

$$(q_L - q_H)\left[u(w_0 - p) - u(w_0 - p - L + z)\right] \geq c(e_H) - c(e_L).$$

- **LHS: marginal benefit of high effort.** A careful driver lowers the accident probability by $q_L - q_H > 0$, and so more often keeps the better (no-accident) wealth.
- **RHS: marginal cost of high effort,** $c(e_H) - c(e_L) > 0$.
- The contract induces care only if the wealth gap in brackets is wide enough to cover the extra effort cost.

Insurance – Result: $L > z$ (less than full coverage)

- $q_L - q_H > 0$ and $c(e_H) - c(e_L) > 0$ by assumption.
- Therefore the bracket must be positive:

$$u(w_0 - p) > u(w_0 - p - L + z).$$

- Since $u(\cdot)$ is strictly increasing, this requires

$$-L + z < 0 \iff L > z.$$

- **Less than full coverage.** The deductible $L - z > 0$ exposes the driver to some residual loss, preserving the incentive for careful driving.
- If $z = L$ (full coverage), the IC fails – the driver has no private gain from careful driving.

Insurance – Real-world implications

- Deductibles and co-insurance: standard features of insurance contracts – a direct response to moral hazard.
- Auto insurance: usage-based premiums (telematics), no-claims bonuses.
- Health insurance: co-payments, deductibles.

Insurance – Real-world implications (cont.)

- Mortgage lending: down-payment requirements.
- **Trade-off:** too small a deductible $\Rightarrow$ moral hazard. Too large a deductible $\Rightarrow$ insurance loses value to risk-averse clients.
- Optimal deductible balances incentive provision and risk-sharing.

Insurance – The full-insurance trap

- Suppose the insurer ignored moral hazard and offered $z = L$ (full coverage).
- Then in both states the worker's wealth is the same: $w_0 - p$ (absent the loss).
- Worker's expected utility is the same regardless of effort, so the worker chooses the **cheaper** effort – e_L.

Insurance – The full-insurance trap (cont.)

- Higher accident probability $q_L > q_H$ means more claims.
- Insurer must raise premium p to break even, which raises expected loss – a vicious cycle.
- **Conclusion:** optimal contracts must leave the agent *exposed* to part of the loss to align incentives.

The full-insurance trap – A design menu

- Every observed feature of health, auto, and disability policies is a response to this trap:
  - *Deductibles.* The insured pays the first dollars of any loss.
  - *Co-insurance.* The insured pays a fixed fraction of each dollar of loss.
  - *Experience rating and no-claims bonuses.* Move next year's premium in response to this year's claims.
  - *Coverage limits and exclusions.* Cap the transfer of risk.
- Each tool trades some risk-sharing against some incentive provision – the same trade-off from the wage-contract problem, with premiums and claims in place of wages and profits.

3.4.3 Efficiency wages – Shapiro & Stiglitz (1984)

- Question: why do firms pay above the competitive wage in equilibrium?
- Risk-neutral worker, $u(w, e) = w - e$, binary effort $e \in \{0, 1\}$.
- Firm's product: αm, where $m =$ number of workers choosing $e = 1$, $\alpha > 2$.
- Firm monitors each worker with equal probability; pays w if $e = 1$ observed, 0 otherwise.
- Fixed monitoring cost $F > 0$; $n > 2$ workers.

Efficiency wages – Timing and solution method

- Two-stage game:
  1. Firm sets wage w.
  2. Worker chooses effort e.
- Solve by **backward induction**: first find the worker's best effort given a wage, then plug back into the firm's problem.
- The firm's monitoring turns the worker's effort into a signal – what would be a hidden action becomes observed with some probability.
- That probability, together with the wage spread between working and being fired, will set the size of the required wage premium.

Efficiency wages – Perfect monitoring

- Slacking ($e = 0$) is detected with certainty $\Rightarrow$ worker is fired (payoff 0).
- Working ($e = 1$) yields utility $w - 1$.
- Worker exerts effort iff $w - 1 \geq 0$, i.e., $w \geq 1$.
- Firm sets $w = 1$ (just induces effort).
- Firm's profit-max (with $m = n$ when $w = 1$):

$$\max_{0 \leq n \leq \overline{n}} \alpha n - n - F.$$

- FOC: $\alpha - 1 > 0$. Since $\alpha > 2$, the firm hires the entire pool, $n^* = \overline{n}$.

Efficiency wages – Imperfect monitoring

- Now: if worker exerts $e = 0$, he is caught with probability $\frac{1}{n}$ only (firm monitors one worker at a time).
- Worker's payoff if $e = 0$:

$$\underbrace{\tfrac{1}{n} \cdot 0}_{\text{caught}} + \underbrace{\tfrac{n-1}{n} w}_{\text{not caught}} = \frac{n-1}{n} w.$$

- Worker exerts $e = 1$ iff

$$w - 1 \geq \frac{n-1}{n} w.$$

- Multiply both sides by $n > 0$: $nw - n \geq (n-1)w = nw - w$, and cancelling nw leaves $-n \geq -w$, i.e., $w \geq n$.
- Firm needs to pay $w \geq n$ to deter shirking: a larger workforce (each monitored less often) requires a more generous wage.

Efficiency wages – Firm's optimal hiring

- Firm sets $w = n$ (just enough to induce effort).
- Firm's profit (with $m = n$, $w = n$):

$$\max_{n \geq 0} \ \alpha n - n^2 - F.$$

- FOC: $\alpha - 2n = 0 \Rightarrow n^* = \frac{\alpha}{2}$.
- Equilibrium wage: $w^* = n^* = \frac{\alpha}{2} > 1$.
- **Wage premium:** $w^* - 1 = \frac{\alpha}{2} - 1 > 0$.
- The firm pays **above the competitive wage** to deter shirking when monitoring is imperfect.

Efficiency wages – Implications

- **Result:** imperfect monitoring $\Rightarrow$ higher wages *and* fewer workers.
- Worker hires under perfect monitoring: $n^* = \overline{n}$ (whole pool).
- Worker hires under imperfect monitoring: $n^* = \frac{\alpha}{2}$ (finite).

Efficiency wages – Implications (cont.)

- Pay premium $w^* - 1$ measures the firm's willingness to pay for better monitoring technology.
- Macro implication (Shapiro & Stiglitz):
  - Efficiency wages can sustain **involuntary unemployment** in equilibrium.
  - Workers don't want to risk losing high-paying jobs.
  - Consistent with empirical evidence of wage dispersion across similar workers.

Efficiency wages – Same math, different label

- An **efficiency wage** is a wage above the market-clearing level, paid to keep the worker's IC binding when the deterrent for shirking is dismissal rather than a low-wage state.
- Compare with the two-outcome wage contract:
  - Two-outcome contract: w_H if profit is high, w_L if profit is low – the wage *gap* deters shirking.
  - Efficiency-wage contract: $w > \overline{w}$ if not caught shirking, $\overline{w}$ (outside option) if caught – the wage *premium* deters shirking.

Efficiency wages – Same math (cont.)

- **Same idea, different variable.** In both cases the IC constraint prescribes a positive spread; only the labeling of the two outcomes differs.
- Because efficiency wages exceed the market-clearing wage, they generate involuntary unemployment in equilibrium – a macroeconomic consequence of a microeconomic incentive problem.

Section 3.5
Summary

Efficiency wages – Numerical illustration ($\alpha = 4$)

- With $\alpha = 4$:
  - Perfect monitoring: $w = 1$, hire all available workers.
  - Imperfect monitoring: $n^* = 2$, $w^* = 2$.
  - Wage premium: $w^* - 1 = 1$.
- Firm's profit under perfect monitoring (with n workers): $(\alpha - 1)n - F = 3n - F$ – grows linearly.
- Firm's profit under imperfect monitoring: $\alpha n - n^2 - F = 4n - n^2 - F$, maximized at $n = 2$ with profit $4 - F$.

Efficiency wages – Numerical illustration (cont.)

- Imperfect monitoring is doubly costly:
  - Fewer workers hired.
  - Higher wage per worker.

Efficiency wages – The role of n as a discipline device

- The firm can adjust n to make shirking less attractive.
- Smaller $n \Rightarrow$ higher monitoring intensity per worker $\Rightarrow$ catch rate $\frac{1}{n}$ is higher.
- But smaller n also reduces output αn.

Efficiency wages – Role of n (cont.)

- Optimal $n^* = \frac{\alpha}{2}$ balances:
  - Output gain from more workers.
  - Wage cost (which equals n).
- Higher α (more productive technology) $\Rightarrow$ higher n^* and higher w^*.
- Connecting to data: high-productivity industries pay higher efficiency wages.

Two-period moral hazard – A brief mention

- In multi-period settings, contracts can use **dynamic incentives**:
  - Deferred compensation (vesting schedules).
  - Promotion tournaments.
  - Reputation building.
- Government policies:
  - Unemployment insurance with declining benefits over time.
  - Payroll taxes adjusted to incentivize job search.
- These dynamic schemes can sometimes **relax** the static moral hazard problem by drawing on future payoffs.

Two-period moral hazard – A brief mention (cont.)

- Detailed analysis: see Holmström & Milgrom (1987), Spear & Srivastava (1987), and the literature on dynamic principal-agent models.

Summary – The principal's problem

- Under asymmetric information the principal solves:

$$\max_{w_H, w_L} \; p_H(\pi_H - w_H) + (1 - p_H)(\pi_L - w_L)$$

  subject to PC and IC.
- Lagrangian has **two multipliers** (λ on PC, μ on IC).
- Both constraints **bind** in equilibrium ($\lambda, \mu > 0$).
- Result: wages spread across outcomes, $w_H > w_L$.
- Compared with symmetric info:

$$w_H^{AI} > w_H^{SI}, \quad w_L^{AI} < w_L^{SI}.$$

Summary – Roadmap and MLRP

- **5-step roadmap:**
  1. Use binding IC to express w_H as a function of w_L.
  2. Use binding PC to express w_H as a function of w_L.
  3. Equate to find w_L^{AI}.
  4. Back-substitute for w_H^{AI}.
  5. Verify high effort beats low effort for the principal.
- **MLRP** guarantees monotonic wages ($w_H > w_L$):

$$\frac{f(\pi_H|e_L)}{f(\pi_H|e_H)} < \frac{f(\pi_L|e_L)}{f(\pi_L|e_H)}.$$

- Profit loss from moral hazard = cost of providing incentives.

Summary – Generalizations and applications

- **Three effort levels:** only one IC binds (the binding one identifies which deviation is most tempting – e_M or e_L).
- **Three outcomes:** same roadmap with one extra wage; results satisfy MLRP.
- **Applications:**
  - **Credit rationing**: bank caps revenue at $\overline{Rev}$ to induce safe project; not all firms get loans.
  - **Insurance moral hazard**: $L > z$ (less than full coverage) preserves driver's incentive for care.
  - **Efficiency wages**: $w^* = \frac{\alpha}{2} > 1$ paid to deter shirking under imperfect monitoring; involuntary unemployment.

Appendix 3.A

Moral Hazard with Two Efforts and Two Outcomes – A More General Model

Appendix 3.A – General conditional probabilities

		Profits	
		π_H	π_L
Effort	$e_H = e$	p	$1-p$
	$e_L = 0$	q	$1-q$

Table 3.A1. Conditional probabilities.

- Probabilities p and q satisfy $p > q$ (high effort makes high profits more likely).
- Allows for general parametric analysis – not tied to the specific $p_H = \frac{1}{2}$, $p_L = \frac{1}{4}$ values used in Example 3.1.
- Here $p \equiv f(\pi_H|e_H)$ and $q \equiv f(\pi_H|e_L)$, and we keep $v(w) = \sqrt{w}$ with $g(e) = e$.
- We now solve for the general wage pair (w_H, w_L) and read its comparative statics.

Appendix 3.A – Problem and binding constraints

- Inducing high effort $e_H = e$, the principal solves

$$\min_{w_H, w_L} \; p w_H + (1-p) w_L$$

subject to a binding PC and a binding IC:

$$\begin{aligned} p\sqrt{w_H} + (1-p)\sqrt{w_L} - e &= \bar{u}, \quad \text{(PC)} \\ p\sqrt{w_H} + (1-p)\sqrt{w_L} - e &= q\sqrt{w_H} + (1-q)\sqrt{w_L}. \quad \text{(IC)} \end{aligned}$$

- Subtracting the right side of IC from its left side, the common terms cancel and the IC collapses to the **spread condition**

$$(p-q)\left(\sqrt{w_H} - \sqrt{w_L}\right) = e.$$

- So the spread is $\sqrt{w_H} - \sqrt{w_L} = \dfrac{e}{p-q}$: a more responsive signal (larger $p-q$) needs a smaller spread.

Appendix 3.A – Solving for the wages

- **Step 1.** Substitute the spread condition into the binding PC, which reduces to

$$\sqrt{w_L} + q\left(\sqrt{w_H} - \sqrt{w_L}\right) = \overline{u}.$$

- **Step 2.** Solve for $\sqrt{w_H}$ and return to the spread condition; it reduces to $(p-q)\dfrac{\overline{u} - \sqrt{w_L}}{q} = e$, hence

$$w_L = \left(\frac{(p-q)\overline{u} - eq}{p-q}\right)^2.$$

- **Step 3.** Add the spread $\dfrac{e}{p-q}$ to $\sqrt{w_L}$:

$$w_H = \left(\frac{(p-q)\overline{u} + e(1-q)}{p-q}\right)^2.$$

Appendix 3.A – Checking against Example 3.1

- Example 3.1 has $p = \frac{1}{2}$, $q = \frac{1}{4}$, $e = 5$, $\overline{u} = 9$, so $p - q = \frac{1}{4}$.
- Low wage:

$$w_L = \left(\frac{\frac{1}{4}(9) - 5 \cdot \frac{1}{4}}{\frac{1}{4}} \right)^2 = (9-5)^2 = \$16.$$

- High wage:

$$w_H = \left(\frac{\frac{1}{4}(9) + 5 \cdot \frac{3}{4}}{\frac{1}{4}} \right)^2 = (9+15)^2 = \$576.$$

- These reproduce $(w_H^{AI}, w_L^{AI}) = (\$576, \$16)$ exactly, a check on the general formulas.

Appendix 3.A – Comparative statics (Table 3.A2)

	w_H	w_L
Reservation utility, $\bar{u}$	$+$	$+$
Effort, e	$+$	$-$
Prob. of high profit under high effort, p	$-$	$+$
Prob. of high profit under low effort, q	$+$	$-$

Table 3.A2. Comparative statics of w_H and w_L.

- p measures the **efficacy** of effort; q measures the worker's **luck** (high profit despite low effort).
- Intuitively, a higher q makes shirking look better, so the principal widens the spread ($w_H \uparrow$, $w_L \downarrow$); a higher p makes effort more telling, so the spread narrows.

Information rents – A first look

- Under symmetric information, the agent earns exactly $\overline{u}$ (PC binds, no surplus).
- Under asymmetric information, the agent *still* earns exactly $\overline{u}$ in expectation (PC still binds).
- But ex-post, the agent receives wages that vary with outcome – a **lottery** rather than a sure thing.

Information rents – A first look (cont.)

- For the risk-averse agent, this lottery is less valuable than the constant wage that delivers the same expected utility.
- **Information rent** (formally introduced in Chapter 6): the extra expected payment the principal must make to compensate the agent for bearing risk.
- In our examples, the principal's profit loss \$100 captures this inefficiency.

Linking back to risk attitudes

- Chapter 2 showed: optimal risk-sharing gives the risk-averse agent a **constant** wage when the principal is risk-neutral.
- Chapter 3 shows: under hidden action, this constant wage destroys incentives.
- Optimal contract under AI **trades off** two desiderata:
  - **Risk sharing** – prefer constant wages.
  - **Incentive provision** – prefer wages that vary with output.

Linking back to risk attitudes (cont.)

- Equilibrium: a compromise – wages spread enough to satisfy IC, but not more (to minimize the agent's risk burden).
- **Welfare loss** relative to SI captures this fundamental trade-off.
- Chapter 4 quantifies it with a single bonus parameter b when effort is continuous.

Empirical evidence on moral hazard contracts

- **Performance pay:** Lazear (2000) – Safelite Glass switch to piece rates raised productivity by 44%.
- **Sharecropping:** Burchardi et al. (2019) – exogenous variation in sharecropping vs. fixed-rent contracts in India showed substantial moral hazard effects on effort.
- **Health insurance:** RAND HIE confirms moral hazard in healthcare consumption.

Empirical evidence on moral hazard contracts (cont.)

- **Executive compensation:** Bebchuk & Fried (2004) – CEO pay structure responds to moral-hazard considerations but also reflects managerial power.
- Empirically, contracts in the field exhibit features predicted by moral hazard theory: deductibles, bonuses, share contracts, deferred pay.

Caveats and extensions

- Our model assumes:
  - Single agent – no team production.
  - Discrete effort – relaxed in Chapter 4.
  - One-shot interaction – no reputation effects.
  - Full commitment by both parties.
- Extensions covered later or in the literature:
  - **Teams** (Holmström 1982): moral hazard among multiple agents; budget-breaking and free-riding.
  - **Career concerns** (Holmström 1982/1999): implicit incentives from future labor market.
  - **Limited liability** and risk-shifting in finance.
  - **Continuous time models** (Holmström & Milgrom 1987, Sannikov 2008).

Looking ahead – Chapter 4

- Chapter 3 dealt with **discrete** effort (2 or 3 levels).
- Chapter 4 generalizes to **continuous** effort $e \in [0, 1]$.
- Two new tools:
  - The **first-order approach** (FOA): replaces the discrete IC with the agent's FOC.
  - **Linear contracts** $w(x) = a + bx$, which have closed-form solutions under CARA utility and normal noise.

Looking ahead – Chapter 4 (cont.)

- Key result with CARA utility:

$$b^* = \frac{1}{1 + \rho\sigma^2}.$$

- Trade-off between risk-sharing and incentives is captured by a single bonus parameter b.
- Plus the **informativeness principle** and **multitask** moral hazard.

Common student mistakes to avoid (Chapter 3)

- **Skipping the IC constraint.** IC is not optional – forgetting it collapses the problem back to Chapter 2 and delivers full insurance again.
- **Assuming $w_H > w_L$ from the outset.** Monotonicity is a *consequence* of MLRP, not an assumption; without MLRP, the ranking can flip.
- **Setting the multipliers to zero.** Both PC and IC bind ($\lambda > 0$, $\mu > 0$); dropping either leaves the problem underdetermined.

Common student mistakes to avoid (Chapter 3) (cont.)

- **Forgetting limited liability.** If $w_L < 0$, add $w_L \geq 0$ – PC then becomes slack and the worker gets a rent.
- **Reading "moral hazard" as "bad behavior."** The agent is optimizing given the contract; the contract itself has to be redesigned.

Key takeaways from Chapter 3

1. Asymmetric information adds an **Incentive Compatibility** constraint; **both PC and IC bind**.
2. Wages spread across outcomes ($w_H > w_L$) under **MLRP**.
3. Wages are **higher** (for high outcomes) and **lower** (for low outcomes) than under symmetric information.
4. Moral hazard generates a **profit loss** for the principal – the cost of providing incentives.
5. The 5-step **roadmap** solves a 2-equation, 2-unknown system (binding PC + binding IC) – without needing the principal's objective function for wages.
6. Applications: **credit rationing**, **insurance deductibles**, and **efficiency wages** all emerge from the same hidden-action logic.

End of Chapter 3

Chapter 4

Moral Hazard with Continuous Effort – The First-Order Approach

Instructor Slides

Contract Theory: An Introduction with Applications

Ana Espinola-Arredondo, Felix Muñoz-Garcia, and Pak-Sing Choi

Outline of Chapter 4

- Extends Chapter 3 by letting the agent's effort be a **continuous** variable $e \in [0, 1]$.
- Sections:
  - 4.1 Introduction.
  - 4.2 Finding equilibrium effort – the first-order approach (FOA).
  - 4.3 The first-order approach with linear contracts.
  - 4.4 Observing output or performance.
  - 4.5 Linear contracts with informative signals.
  - 4.6 Moral hazard with multiple tasks.
  - 4.7 Summary. 4.8 Exercises. Appendix 4.A.

What you will learn

- The **first-order approach** (FOA): replace the discrete IC with the agent's first-order condition over a continuum of effort levels.
- How to model output as $x = e + \varepsilon$ with normal noise.
- How **linear contracts** $w(x) = a + bx$ emerge naturally with CARA utility and normal output.

What you will learn (cont.)

- Closed-form solution: $b^* = \dfrac{1}{1+\rho\sigma^2}$.
- Three extensions:
  - Wages based on **performance** ($p = \beta e + \phi$) rather than output.
  - **Informative signals**: $\alpha^* = -\sigma_{\varepsilon s}/\sigma_s^2$.
  - **Multitask moral hazard**: how multiple effort dimensions interact.

Section 4.1
Introduction

From discrete to continuous effort

- Chapters 2–3 let the worker pick between two or three discrete effort levels.
- A worker who only decides whether to make an extra sales call is well described by two actions, so we do not switch to a continuum for realism.
- We let effort be **continuous**, $e \in [0, 1]$, because a continuum makes **marginal analysis** available.
- That buys us three questions the two-action model cannot pose:
  - *how much* effort the contract induces, not merely which of two actions it selects;
  - what the **interior optimum** looks like, where the marginal benefit of more effort equals its marginal cost;
  - how the fixed payment and bonus respond to risk aversion, output variance, and the cost of effort (**comparative statics**).

From discrete to continuous effort (cont.)

- **Challenge.** The worker now picks effort from an interval, so his IC is an **argmax** over a continuum:

$$e^* \in \arg\max_{e \in [0,1]} \; f(\pi_H|e)u(w_H) + f(\pi_L|e)u(w_L) - g(e),$$

  and we can no longer just compare two utility numbers, as in Chapter 3.
- **Solution.** Replace the global argmax with the worker's **first-order condition** in e – the **first-order approach** (FOA).
- Intuitively, we put ourselves in the worker's shoes: the effort he chooses sets the slope of his expected utility to zero, and that single equation stands in for the whole argmax.

Why linear contracts?

- Under continuous effort the optimal contract can in principle be **nonlinear** (Mirrlees, Rogerson).
- Holmström and Milgrom (1987) give a clean benchmark: with **CARA** utility and **normally distributed** output, the optimal contract is **linear**,

$$w(x) = a + bx,$$

  where a is the fixed payment (unaffected by profit) and $b \in [0, 1]$ is the bonus rate, the worker's share of output.
- Economically, the two assumptions reduce the problem to a mean and a variance, so the principal needs only a slope and an intercept rather than a full nonlinear schedule.

Why linear contracts? (cont.)

- The linear form $w(x) = a + bx$ nests three familiar wage arrangements:
  1. **Constant salary**: $a > 0$, $b = 0$ – pay is flat, insulated from profit.
  2. **Pure bonus**: $a = 0$, $b \in (0, 1]$ – pay is a share of output, with no guaranteed floor.
  3. **Salary plus bonus**: $a > 0$, $b \in (0, 1]$ – a guaranteed floor plus a performance share.
- This one functional form yields clean closed-form solutions for (a^*, b^*, e^*) and intuitive comparative statics, which is why we adopt it for the rest of the chapter.

Roadmap for Chapter 4

- **Section 4.2.** Continuous effort with just *two* outcomes – the smallest jump beyond Chapter 3. Wage-schedule spirit kept, but effort now lives on the real line.
- **Section 4.3.** Continuous *output* plus **linear contracts** $w = a + b\pi$ and **CARA** utility – the workhorse model, closed-form solutions, and clean comparative statics.
- **Section 4.4.** Performance and output split – the wage depends on a performance measure, not just the profit level.

Roadmap for Chapter 4 (cont.)

- **Section 4.5. Informative signals.** Extra signals about effort enter the contract via Holmström's informativeness principle.
- **Section 4.6. Multitask** moral hazard – Holmström and Milgrom (1991); rewarding one task can distort another.

Section 4.2
Finding Equilibrium Effort

Setup – Continuous effort, two outcomes

- Effort is continuous, $e \in [0, 1]$, with higher e meaning more effort.
- Two profit outcomes $\pi_H > \pi_L$ occur with conditional probabilities $f(\pi_H|e)$ and $f(\pi_L|e)$, so effort shifts the *odds* of each outcome rather than the outcome itself.
- The principal's expected profit nets wages out of each profit level, weighted by its probability:

$$f(\pi_H|e)[\pi_H - w_H] + f(\pi_L|e)[\pi_L - w_L].$$

Setup – Continuous effort, two outcomes (cont.)

- The principal chooses **three** variables: the effort e she wants induced and the two wages w_H, w_L.
- The worker's **PC** (he participates) and **IC** (he exerts the desired effort) must both hold.
- The IC is now an **argmax**: the worker picks the effort level that maximizes his own expected utility, given the wages.
- Economically, the principal sets wages anticipating the whole best-response effort $e^*(w_H, w_L)$, not a single number she can dictate.

Example 4.1 – A well-behaved conditional probability

- A well-behaved $f(\pi_H|e)$ should have three properties:
  - (i) **increasing in** e, so more effort makes high profit more likely;
  - (ii) attaining its **largest** value at full effort, $e = 1$;
  - (iii) attaining its **smallest** value at zero effort, $e = 0$.
- Note (ii) does *not* say the probability reaches one: even full effort cannot guarantee high profit, since profit also depends on factors outside the worker's control.
- Likewise (iii) does *not* say it reaches zero: even zero effort may yield high profit by luck, as when a vibrant market fills orders through repeat refills rather than the worker's outreach.

Example 4.1 – A well-behaved conditional probability (cont.)

- One functional form satisfying (i)–(iii) is a linear combination:

$$f(\pi_H|e) = e \cdot f_H(\pi_H) + (1-e) \cdot f_L(\pi_H),$$

  where $f_H(\pi_H) > f_L(\pi_H)$ are the probabilities of π_H under full and zero effort.
- Here effort e is a **weight**: raising e puts more weight on the high-effort distribution, so high profit becomes more likely; slacking shifts the weight the other way.
- The low-profit probability is symmetric,

$$f(\pi_L|e) = e \cdot f_H(\pi_L) + (1-e) \cdot f_L(\pi_L),$$

  with $f_L(\pi_L) > f_H(\pi_L)$, so a low profit is more likely under no effort.

The principal's problem with continuous effort

- The principal maximizes expected profit over (e, w_H, w_L):

$$\max_{e, w_H, w_L} \; f(\pi_H|e)[\pi_H - w_H] + f(\pi_L|e)[\pi_L - w_L]$$

- subject to the worker's **PC**, so he participates only if the contract clears his reservation utility $\overline{u}$:

$$f(\pi_H|e)u(w_H) + f(\pi_L|e)u(w_L) - g(e) \geq \overline{u}.$$

The principal's problem with continuous effort (cont.)

- and subject to the worker's **IC**, now an argmax because effort is continuous:

$$e^* \in \arg\max_{e \in [0,1]} \; f(\pi_H|e)u(w_H) + f(\pi_L|e)u(w_L) - g(e).$$

- This is a **two-stage game**: first the principal offers the contract, then the worker chooses effort.
- We solve it by **backward induction**: read off the worker's effort in stage two, then pick the contract in stage one anticipating that response.

The agent's first-order approach

- **Stage 2.** Taking the contract as given, the worker solves

$$\max_{e \in [0,1]} \; f(\pi_H|e)u(w_H) + f(\pi_L|e)u(w_L) - g(e). \qquad (4.1)$$

- Differentiate in e and set the slope to zero (the FOA step):

$$f'(\pi_H|e)u(w_H) + f'(\pi_L|e)u(w_L) - g'(e) = 0. \qquad (4.2)$$

- A prime on f is the derivative in its *second* argument, effort, holding π_i fixed: $f'(\pi_i|e) \equiv \partial f(\pi_i|e)/\partial e$.
- We use (4.2) as the new IC constraint, **IC'**. It is a valid stand-in for the argmax only if the objective is concave in e, i.e., $g''(e) > 0$ and $f''(\pi_i|e) < 0$.

When is the first-order approach valid?

- The FOA replaces the worker's IC (an inequality that must hold at *every* effort level) with his first-order condition, a single equality.
- **Cost of the shortcut.** The FOC is only *necessary*: it locates a stationary point, which need not be the worker's global optimum unless his objective is concave in effort.
- The catch is that the utilities $u(w_H)$, $u(w_L)$ are not primitives – the principal *chooses* them – so whether the objective is concave depends on the very contract being designed. The FOA is therefore not valid in general.

When is the first-order approach valid? (cont.)

- **Sufficient conditions** (Rogerson, 1985): the conditional-probability family satisfies the **monotone likelihood-ratio property** (higher profit relatively more likely as effort rises) and the **convexity of the distribution function** (CDFC).
- Under MLRP + CDFC the relaxed problem we solve is equivalent to the original one.
- In this chapter CARA utility, normal noise, and linear contracts all satisfy these conditions, so we apply the FOA throughout.
- Intuitively, these conditions keep the worker's payoff single-peaked in effort, so the one FOC pins down his true best response.

The reformulated principal's problem

- Substituting IC' for the argmax, the principal's problem is

$$\max_{e, w_H, w_L} \; f(\pi_H|e)[\pi_H - w_H] + f(\pi_L|e)[\pi_L - w_L]$$

$$\text{s.t. } f(\pi_H|e)u(w_H) + f(\pi_L|e)u(w_L) - g(e) \geq \overline{u}, \quad \text{(PC)}$$

$$f'(\pi_H|e)u(w_H) + f'(\pi_L|e)u(w_L) - g'(e) = 0. \quad \text{(IC')}$$

The reformulated principal's problem (cont.)

- The IC is now a single **equation**, not an argmax over an interval.
- That makes the program a standard Lagrangian problem with two multipliers, λ on the PC and μ on IC'.
- Economically, the FOA converts an infinite-dimensional constraint (optimality at every effort) into one tractable marginal condition we can attach a shadow price to.

Lagrangian and FOC for w_H

- Attach λ to the PC and μ to IC':

$$\begin{aligned}\mathcal{L} \;=\;& f(\pi_H|e)[\pi_H - w_H] + f(\pi_L|e)[\pi_L - w_L] \\ &+\lambda\left[f(\pi_H|e)u(w_H) + f(\pi_L|e)u(w_L) - g(e) - \overline{u}\right] \\ &+\mu\left[f'(\pi_H|e)u(w_H) + f'(\pi_L|e)u(w_L) - g'(e)\right].\end{aligned}$$

- **FOC in w_H.** Only the w_H terms survive; the profit term contributes $-f(\pi_H|e)$ and each utility term an $f\,u'(w_H)$:

$$-f(\pi_H|e) + \lambda f(\pi_H|e)u'(w_H) + \mu f'(\pi_H|e)u'(w_H) = 0.$$

Solving the FOC for w_H – Step by step

- **Step 1. Divide by** $f(\pi_H|e) > 0$ and carry the profit term to the right:

$$\lambda u'(w_H) + \mu \frac{f'(\pi_H|e)}{f(\pi_H|e)} u'(w_H) = 1.$$

- **Step 2. Factor out** $u'(w_H)$**:**

$$\left(\lambda + \mu \frac{f'(\pi_H|e)}{f(\pi_H|e)}\right) u'(w_H) = 1.$$

- **Step 3. Divide by** $u'(w_H) > 0$ to isolate the multiplier combination:

$$\lambda + \mu \frac{f'(\pi_H|e)}{f(\pi_H|e)} = \frac{1}{u'(w_H)}. \tag{4.3}$$

Figure 4.1 – Wages under SI vs. AI

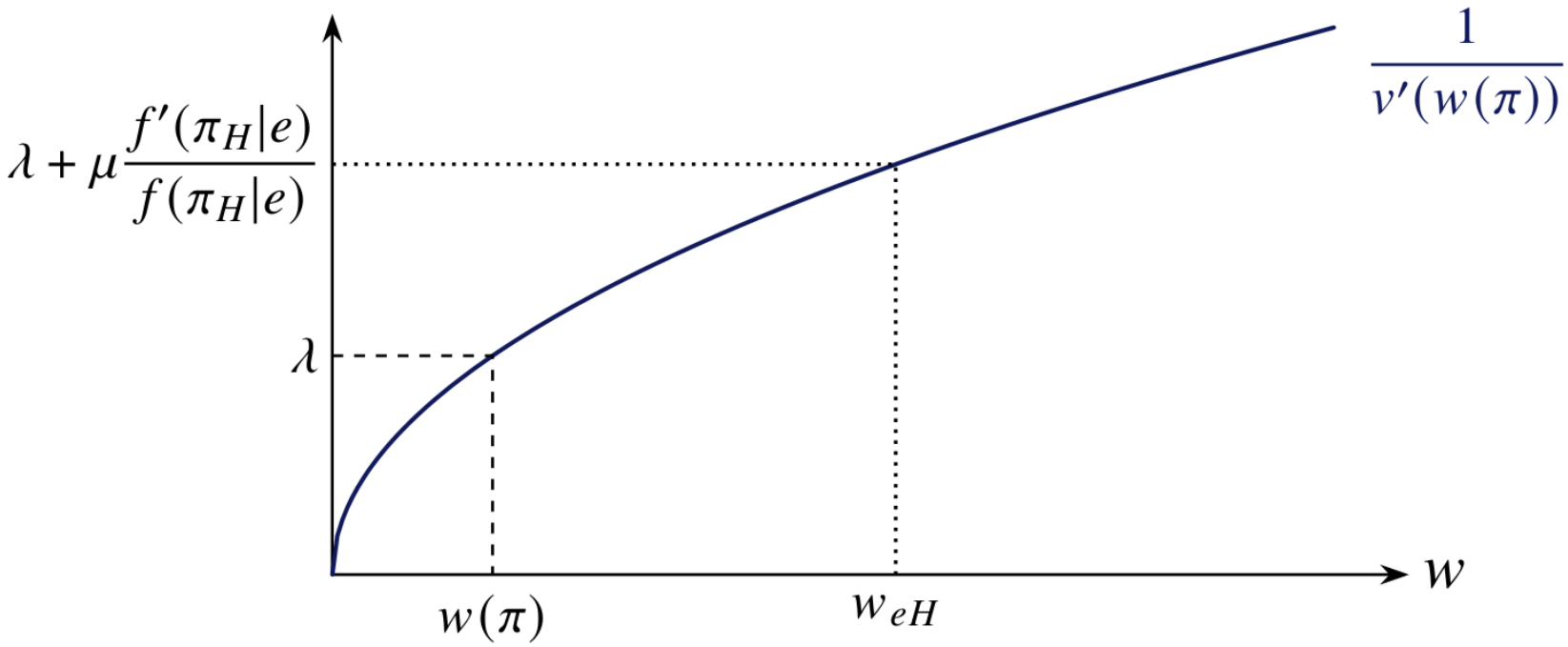


Figure 4.1. Wages under symmetric and asymmetric information. Equation (4.3) for AI plots above the SI line $\lambda = 1/u'(w_H)$, crossing the increasing curve $1/u'(w_H)$ at a higher wage $w_H^{AI} > w_H^{SI}$.

Interpreting equation (4.3) – An elasticity

- The term that separates AI from SI is $\mu \dfrac{f'(\pi_H|e)}{f(\pi_H|e)}$.

  Multiplying and dividing by e,

$$\mu \frac{f'(\pi_H|e)}{f(\pi_H|e)} = \frac{\mu}{e} \cdot \underbrace{\frac{f'(\pi_H|e)\, e}{f(\pi_H|e)}}_{\text{elasticity of } f \text{ in effort}}.$$

- The brace is the **elasticity** ε_H: the percentage rise in the probability of π_H from a one-percent rise in effort.
- The more elastic high profit is in effort, the larger ε_H, and the more generous the salary the principal offers in that state.

Interpreting equation (4.3) – Why not a likelihood ratio

- When effort does not move π_H, $f'(\pi_H|e) = 0$, the elasticity is nil and the AI wage equals the SI wage: paying for effort buys nothing because π_H is no more likely.
- We avoid calling $\dfrac{f'(\pi_H|e)}{f(\pi_H|e)}$ a "likelihood ratio." It is not a ratio of two likelihoods but the derivative of $\log f(\pi_H|e)$ in effort.
- We reserve "likelihood ratio" for the monotone likelihood-ratio property of Chapter 3.

Interpreting equation (4.3) – An elasticity (cont.)

- When effort matters strongly, the elasticity is large, so the wage spread under AI is large.
- The symmetric result for w_L follows the same three steps:

$$\lambda + \mu \frac{f'(\pi_L|e)}{f(\pi_L|e)} = \frac{1}{u'(w_L)}.$$

- Compactly: $\lambda + \mu \frac{f'(\pi_i|e)}{f(\pi_i|e)} = \frac{1}{u'(w_i)}$ for $i \in \{H, L\}$.

FOC with respect to effort e

- **Step 1. Differentiate $\mathcal{L}$ in e.** Each probability and the cost pick up a derivative:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial e} &= f'(\pi_H|e)[\pi_H - w_H] + f'(\pi_L|e)[\pi_L - w_L] \\ &\quad +\lambda[f'(\pi_H|e)u(w_H) + f'(\pi_L|e)u(w_L) - g'(e)] \\ &\quad +\mu[f''(\pi_H|e)u(w_H) + f''(\pi_L|e)u(w_L) - g''(e)].\end{aligned}$$

- **Step 2. Kill the λ bracket.** It is exactly IC', equation (4.2), which equals zero, so that whole term drops out.

FOC with respect to effort e (cont.)

- **Step 3. Collect the profit terms on the left** and move the wage and μ terms to the right:

$$\begin{aligned} & f'(\pi_H|e)\pi_H + f'(\pi_L|e)\pi_L \\ = \quad & f'(\pi_H|e)w_H + f'(\pi_L|e)w_L \\ & -\mu[f''(\pi_H|e)u(w_H) + f''(\pi_L|e)u(w_L) - g''(e)]. \end{aligned}$$

- The left-hand side is the marginal profit gain from more effort; the right-hand side is its marginal cost. The principal pushes effort up until the two balance.

Two sources of marginal cost from effort

- Equation balances marginal benefit (LHS) and marginal cost (RHS).
- RHS has two parts:
  1. Higher effort raises probability of $\pi_H \Rightarrow$ more high wages paid ($w_H > w_L$).
  2. Inducing higher effort requires a more generous spread between w_H and w_L – the μ-term.
- Sign of second term:
  $f''(\pi_H|e)u(w_H) + f''(\pi_L|e)u(w_L) - g''(e) < 0$ by concavity, so the $-\mu[\cdot]$ becomes positive cost.

Two sources of marginal cost from effort (cont.)

- The principal trades these two costs off against the expected profit gain from higher effort, which is the left-hand side of the FOC.
- Intuitively, inducing more effort is worthwhile only up to the point where the extra expected profit just covers the extra wage bill and the extra spread the worker needs to stay incentive-compatible.

Example 4.2 – Setup

- Unconditional probabilities (subscript = effort, superscript = outcome):

$$\begin{aligned} f_H(\pi_H) &\equiv f_H^H = 0.8, \quad f_L(\pi_H) \equiv f_L^H = 0.4, \\ f_H(\pi_L) &\equiv f_H^L = 0.2, \quad f_L(\pi_L) \equiv f_L^L = 0.6. \end{aligned}$$

- Profits $\pi_H = 60$, $\pi_L = 30$.
- Worker: $u(w) = \sqrt{w}$ (risk-averse), $\overline{u} = 6$, and $g(e) = \frac{1}{2}e^2$ so $g'(e) = e$, with $e \in [0, 1]$.

Example 4.2 – Conditional probabilities

- Insert the numbers into $f(\pi_i|e) = ef_H(\pi_i) + (1-e)f_L(\pi_i)$:

$$\begin{aligned} f(\pi_H|e) &= 0.8e + 0.4(1-e) = \frac{2(1+e)}{5}, \quad f'(\pi_H|e) = 0.4, \\ f(\pi_L|e) &= 0.2e + 0.6(1-e) = \frac{3-2e}{5}, \quad f'(\pi_L|e) = -0.4. \end{aligned}$$

- The high-profit probability grows in effort ($f' > 0$) and the low-profit one shrinks ($f' < 0$), as a well-behaved family requires.

Example 4.2 – PC and IC equations

- **Step 1. Write the binding PC**, so the worker's expected utility hits $\overline{u} = 6$:

$$\frac{2(1+e)}{5}\sqrt{w_H} + \frac{3-2e}{5}\sqrt{w_L} - \frac{1}{2}e^2 = 6.$$

- **Step 2. Write IC' from the FOA**: differentiate the worker's utility in e, using $f'(\pi_H|e) = 0.4$, $f'(\pi_L|e) = -0.4$, and $g'(e) = e$:

$$\frac{2}{5}\sqrt{w_H} - \frac{2}{5}\sqrt{w_L} - e = 0.$$

Example 4.2 – Clearing the system

- **Step 3. Clear fractions** (multiply the PC through by 10 and the IC by 5):

$$\begin{aligned} 4(1+e)\sqrt{w_H} + 2(3-2e)\sqrt{w_L} &= 60 + 5e^2, \\ 2\sqrt{w_H} - 2\sqrt{w_L} &= 5e. \end{aligned}$$

- **Step 4. Solve the two equations** for $\sqrt{w_H}$ and $\sqrt{w_L}$, then square, to obtain the wages as functions of effort (next frame).

Example 4.2 – Wages as functions of effort

- Solving the Step-3 system (isolate $\sqrt{w_L}$ from the second equation, substitute, and square) gives

$$\begin{aligned} w_H(e) &= \frac{(12+3e-e^2)^2}{4}, \\ w_L(e) &= \frac{(12-2e-e^2)^2}{4}. \end{aligned}$$

- Both wages now depend on effort alone, so the principal's problem collapses to a single-variable choice of e.

Example 4.2 – Reducing to one variable

- Insert $w_H(e)$ and $w_L(e)$ into the principal's expected profit:

$$\max_{e \geq 0} \; [0.8e + 0.4(1-e)](60 - w_H(e)) \\ + [0.2e + 0.6(1-e)](30 - w_L(e)).$$

- After expanding, the objective simplifies to a single polynomial in e:

$$\max_{e \geq 0} \; \frac{24(1+2e) - e(30e + 2e^2 - 3e^3)}{4}.$$

Example 4.2 – Solving for effort

- **Step 5. First-order condition** in e:

$$12 - 15e - \frac{3e^2}{2} + 3e^3 = 0,$$

which factors as

$$\frac{3(2-e)}{2}(4 - 3e - 2e^2) = 0.$$

- **Step 6. Pick the feasible root.** The factor $2 - e = 0$ gives $e = 2$, rejected since $e \in [0, 1]$; the quadratic $4 - 3e - 2e^2 = 0$ gives

$$e = \frac{-3 \pm \sqrt{41}}{4}.$$

- Only the positive root lies in $[0, 1]$:

$$e^* = \frac{\sqrt{41} - 3}{4} \simeq 0.85.$$

Example 4.2 – Equilibrium wages and profits

- Plug $e^* = \frac{\sqrt{41}-3}{4}$ into the wage schedules:

$$w_H^* = \frac{3065 + 477\sqrt{41}}{128} \simeq 47.81,$$
$$w_L^* = \frac{3465 - 83\sqrt{41}}{128} \simeq 22.92.$$

- Expected profit:

$$E[\pi] = \frac{451\sqrt{41} - 1497}{128} \simeq 10.87.$$

- **Summary:** $w_H \simeq 47.81$, $w_L \simeq 22.92$, $e^* \simeq 0.85$, $E[\pi] \simeq 10.87$; the spread $w_H - w_L \simeq 24.89$ supplies the incentive.
- **Sanity check (IC).** $\sqrt{w_H} - \sqrt{w_L} \simeq 6.91 - 4.79 = 2.13$, matching $\frac{5}{2}e^* \simeq 2.13$.

Section 4.3
The first-order approach with linear contracts

Setup – Continuous output and linear contracts

- For effort $e \geq 0$, output is

$$x = e + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

- Expected output: $E[x] = e$; output equals effort *in expectation*, but is noisy.
- Lower $\sigma^2 \Rightarrow$ output is a more precise predictor of effort.
- Intuitively, the principal sees output x, not effort e: output is a noisy signal of effort, and σ^2 measures how blurred that signal is.

Setup – Continuous output and linear contracts (cont.)

- Holmström-Milgrom (1987) result: optimal contract is linear:

$$w(x) = a + bx, \quad b \in [0, 1].$$

- Five-stage timing:
  1. Principal and agent sign $w(x) = a + bx$.
  2. Agent exerts effort e (unobservable).
  3. Shock ε is realized.
  4. Output x observed by both.
  5. Agent receives $w(x) = a + bx$.

4.3.1 Principal's objective function

- Substitute the linear wage $w = a + bx$ and use linearity of expectation:

$$\begin{aligned} E[\pi] &= E[x - w] = E(x) - E(w) \\ &= E(x) - [a + bE(x)] \\ &= (1 - b)E(x) - a. \end{aligned}$$

- Since $E(x) = e$ (output equals effort in expectation):

$$E[\pi] = (1 - b)e - a.$$

- **Intuitively,** the principal keeps fraction $(1 - b)$ of output and pays the fixed amount a.

4.3.2 Agent's CARA utility

- Agent's utility of wage w:

$$u(w) = -\exp(-\rho w),$$

with $\rho \geq 0$ the **Arrow-Pratt coefficient of absolute risk aversion** – constant absolute risk aversion (CARA).

- Properties:
  - Originates at -1 when $w = 0$.
  - Approaches 0 as $w \to \infty$.
  - Strictly increasing and concave.
- Convenient for analysis because it has constant Arrow-Pratt coefficient – expected utility separates into a mean-variance form.

Mean-variance form of expected utility

- Under CARA + normality, the agent's expected utility is:

$$E[u(w,e)] = E(w) - \frac{1}{2}\rho \text{Var}(w) - c(e),$$

  (see Appendix 4.A for derivation).
- For $w(x) = a + bx$ and $x = e + \varepsilon$:
  - $E(w) = a + be$ (shift $E(x) = e$ by a and scale by b).
  - $\text{Var}(w) = b^2\sigma^2$ (only the bx term is random; a is constant).
- Hence:

$$E[u(w,e)] = a + be - \frac{1}{2}\rho b^2\sigma^2 - c(e).$$

Mean-variance form (cont.)

- Cost of effort: $c(e) = \frac{1}{2}e^2$, so $c'(e) = e$ and $c''(e) = 1$ – increasing and convex, each extra unit of effort hurting a little more than the last.
- Reservation utility $\overline{u} \geq 0$: the worker signs only if his certainty-equivalent pay, net of effort cost, clears $\overline{u}$.
- Economically, the term $-\frac{1}{2}\rho b^2\sigma^2$ is the worker's **risk premium**: the pay he gives up to avoid the wage volatility that the bonus b exposes him to.

Why mean-variance? – The CARA/normal shortcut

- Two assumptions are doing heavy lifting:
  - CARA utility, $v(w) = -\exp(-\rho w)$.
  - Normally distributed output, $\pi \sim N(e, \sigma^2)$.
- Together they make the agent's expected utility depend on the wage *only through its mean and variance*. Higher moments drop out.
- So the principal's Lagrangian is a standard mean-variance problem in the fixed payment a and the bonus b.

Why mean-variance? (cont.)

- **What we buy with these assumptions:** closed-form solutions for a^*, b^*, and e^*, and clean comparative statics.
- **What we lose:** the results depend on CARA-plus-normal. In Section 4.6 we relax the cost structure to $c(e) = \theta e^2$ while keeping CARA-plus-normal – the linear contract still works.

4.3.3 The principal's problem

- The principal solves

$$\max_{a,b,e} (1-b)e - a$$

$$\text{s.t. } a + be - \frac{1}{2}\rho b^2\sigma^2 - \frac{1}{2}e^2 \geq \overline{u}, \quad \text{(PC)}$$

$$\max_{e\in[0,1]} \; a + be - \frac{1}{2}\rho b^2\sigma^2 - \frac{1}{2}e^2. \quad \text{(IC)}$$

4.3.3 The principal's problem (cont.)

- Apply the FOA to the IC: differentiate the worker's expected utility in e and set it to zero,

$$\frac{\partial E[u(w,e)]}{\partial e} = b - e = 0 \Rightarrow e = b.$$

- The worker chooses effort **equal to the bonus**: with $c(e) = \frac{1}{2}e^2$ his marginal cost e meets his marginal reward b.
- Economically, the bonus is the only lever that moves effort; the fixed payment a shifts his pay level but not his marginal incentive.

Reformulated problem

- Substitute $e = b$ into the objective, giving $(1-b)b - a$, and into the PC. The PC's effort terms collapse:

$$a + b^2 - \tfrac{1}{2}\rho b^2\sigma^2 - \tfrac{1}{2}b^2 \geq \overline{u} \;\Rightarrow\; a + \tfrac{1}{2}b^2(1-\rho\sigma^2) \geq \overline{u}.$$

- The program now has only **two** choice variables, a and b, and only **one** constraint, the PC.

Reformulated problem (cont.)

- So the principal solves

$$\max_{a,b} (1-b)b - a \quad \text{s.t. } a + \tfrac{1}{2}b^2(1-\rho\sigma^2) \geq \overline{u}.$$

- Attach multiplier λ to the PC:

$$\mathcal{L} = (1-b)b - a + \lambda\left[a + \tfrac{1}{2}b^2(1-\rho\sigma^2) - \overline{u}\right].$$

First-order conditions

- **FOC in a:** $-1 + \lambda = 0 \Rightarrow \lambda = 1$, so the PC binds (its shadow price is positive).
- **FOC in b:**

$$1 - 2b + \lambda b(1 - \rho\sigma^2) = 0.$$

First-order conditions (cont.)

- Plug $\lambda = 1$ into the b-FOC and simplify:

$$1 - 2b + b(1 - \rho\sigma^2) = 0 \;\Rightarrow\; 1 - b - b\rho\sigma^2 = 0.$$

- Solve for the optimal bonus:

$$b^* = \frac{1}{1 + \rho\sigma^2}.$$

- Since $b^* \in (0, 1]$ for any $\rho, \sigma^2 \geq 0$, the bonus is a well-defined share of output.

Equilibrium fixed payment a^* and effort

- The PC binds ($\lambda = 1$), so $a + \frac{1}{2}b^2(1 - \rho\sigma^2) = \overline{u}$.
- Substitute $b^* = \frac{1}{1+\rho\sigma^2}$:

$$a + \frac{1}{2}\left(\frac{1}{1+\rho\sigma^2}\right)^2 (1 - \rho\sigma^2) = \overline{u}.$$

- Solving:

$$a^* = \overline{u} - \frac{1 - \rho\sigma^2}{2\left(1+\rho\sigma^2\right)^2}.$$

- Equilibrium effort, from IC' ($e = b$):

$$e^* = b^* = \frac{1}{1+\rho\sigma^2}.$$

4.3.4 Comparative statics – Extreme cases

- **Case** $\sigma^2 = 0$ (output is a perfect signal of effort):

$$(a^*, b^*) = \left(\overline{u} - \tfrac{1}{2},\, 1\right).$$

- The bonus is full, $b^* = 1$: the worker bears all the risk, but there is none to bear, so this is efficient.
- $a^* \geq 0$ only if $\overline{u} \geq \frac{1}{2}$; otherwise limited liability binds, $a^* = 0$, and the worker is paid by bonus alone – "high-powered incentives."

4.3.4 Comparative statics – Noisy output

- **Case** $\sigma^2 = 1$ (output is a noisy signal, e.g., dominated by market conditions):

$$(a^*, b^*) = \left(\overline{u} - \frac{1-\rho}{2(1+\rho)^2}, \frac{1}{1+\rho} \right).$$

- Relative to $\sigma^2 = 0$, the fixed payment is higher and the bonus lower – "low-powered incentives."
- Economically, when output barely reflects effort, loading pay onto the bonus just dumps risk on the worker without buying much effort.

Comparative statics with respect to σ^2

- Derivatives:

$$\frac{\partial a^*}{\partial \sigma^2} = \frac{\rho(3 - \rho\sigma^2)}{2(1 + \rho\sigma^2)^3} > 0$$

(provided $\rho\sigma^2 < 3$, which holds when $\rho, \sigma^2 \leq 1$),

$$\frac{\partial b^*}{\partial \sigma^2} = -\frac{\rho}{(1 + \rho\sigma^2)^2} < 0.$$

- **Interpretation:** as output becomes noisier:
  - Bonus b^* falls – protecting the agent from risk.
  - Fixed payment a^* rises – offsetting the loss in expected bonus.
- Total expected wage may rise (risk premium for the agent).

Figure 4.2 – (a^*, b^*) vs. σ^2

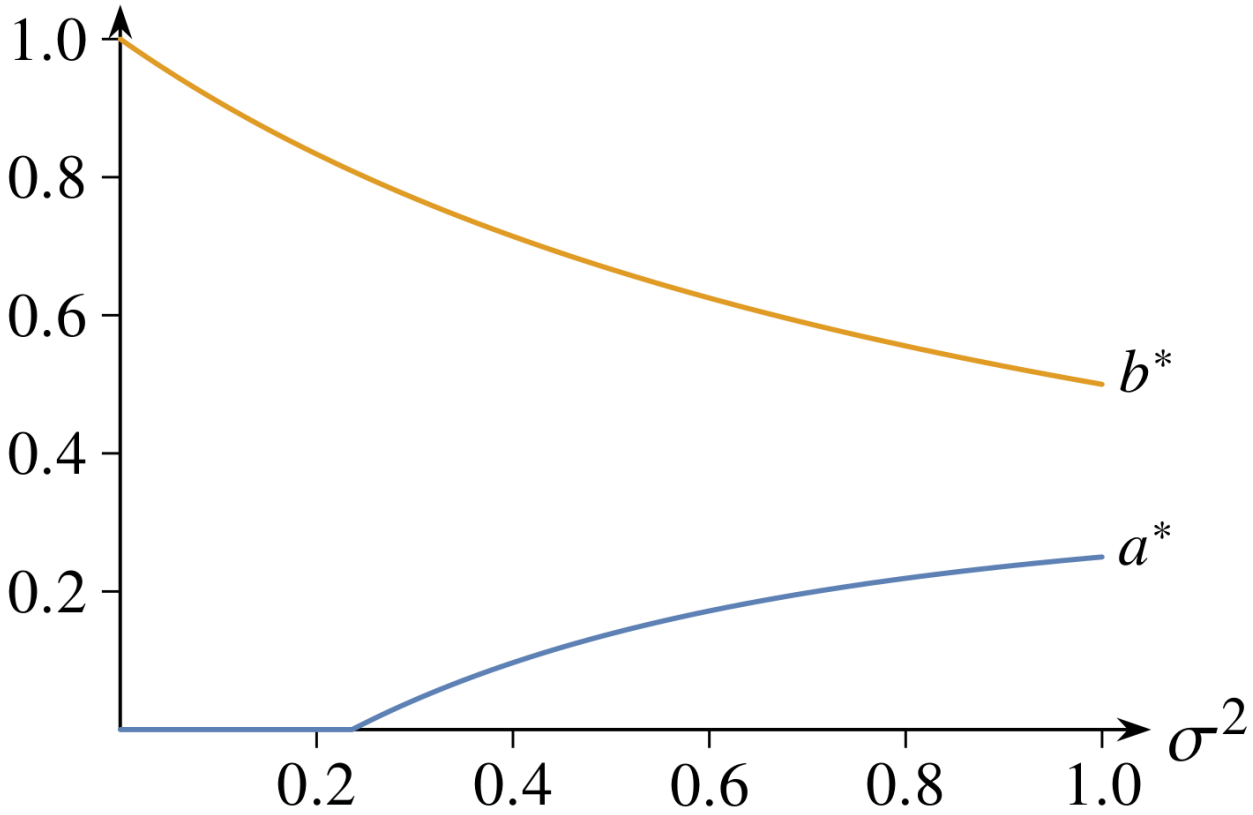


Figure 4.2. Salary pair (a^*, b^*) as a function of profit variance σ^2, evaluated at $\rho = 1$ and $\overline{u} = \frac{1}{4}$. Fixed payment rises in σ^2; bonus falls.

Comparative statics with respect to ρ

- Derivatives:

$$\begin{aligned}\frac{\partial a^*}{\partial \rho} &= \frac{\sigma^2(3-\rho\sigma^2)}{2(1+\rho\sigma^2)^3} > 0, \\ \frac{\partial b^*}{\partial \rho} &= -\frac{\sigma^2}{(1+\rho\sigma^2)^2} < 0.\end{aligned}$$

- More risk-averse agent (higher ρ):
  - Bonus shrinks (agent dislikes risk more).
  - Fixed payment rises (compensating for lower expected bonus).
- More of the salary becomes "guaranteed" as ρ rises – mirrors the risk-sharing result in Chapter 2 (Case 1).

Figure 4.3 – Role of risk aversion ρ

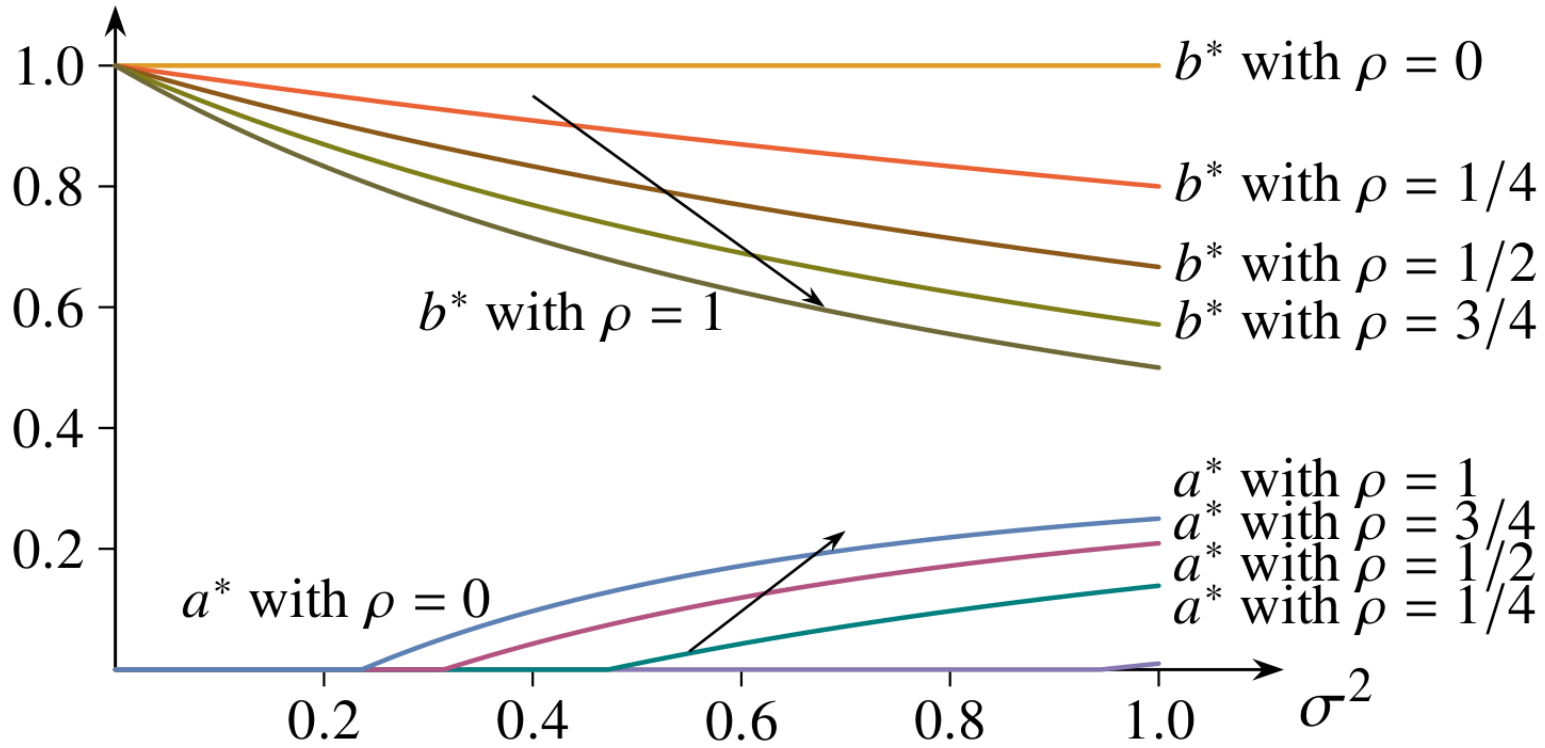


Figure 4.3. The role of risk aversion on salaries. A more risk-averse agent (higher ρ) shifts the a^* curve upward and the b^* curve downward.

Table 4.1 – Comparative statics summary

	Fixed payment, a^*	Bonus, b^*
Effort variance, σ^2	$+$	$-$
Agent's risk aversion, ρ	$+$	$-$

Table 4.1. Equilibrium a^* and b^* – comparative statics.

- Both factors (more noise; more risk aversion) push toward **low-powered incentives**.
- Agent gets more insurance, principal absorbs more risk.
- **Observation:** (a^*, b^*) is unaffected by whether the worker observes ε – linearity leaves the contract unchanged by ex-post information about shocks.

4.3.5 General cost of effort $c(e)$

- Now allow generic $c(e)$ with $c'(e) > 0$, $c''(e) \geq 0$.
- Principal's problem:

$$\max_{a,b,e} (1-b)e - a$$

$$\text{s.t. } a + be - \frac{1}{2}\rho b^2\sigma^2 - c(e) \geq \overline{u}, \quad \text{(PC)}$$

$$\max_{e \geq 0} \; a + be - \frac{1}{2}\rho b^2\sigma^2 - c(e). \quad \text{(IC)}$$

- Same approach: use FOA on IC.

General cost – Solving the principal's problem

- **Step 1. Binding PC**, solve for a:

$$a = \tfrac{1}{2}\rho b^2\sigma^2 - be + c(e) + \overline{u}.$$

- **Step 2. Substitute into the objective** $(1-b)e - a$, which collapses to

$$e - \tfrac{1}{2}\rho b^2\sigma^2 - c(e) - \overline{u}.$$

- **Step 3. Agent's IC (FOA):** $b - c'(e) = 0 \Rightarrow b = c'(e)$.

General cost – First-order condition

- **Step 4. Substitute** $b = c'(e)$, leaving an unconstrained problem in e:

$$\max_{e \geq 0} \; e - \tfrac{1}{2}\rho c'(e)^2\sigma^2 - c(e) - \overline{u}.$$

- **Step 5. Differentiate in** e**:**

$$1 - c'(e^*) - \rho c'(e^*)c''(e^*)\sigma^2 = 0.$$

General cost – Optimal bonus

- Factor $c'(e)$ out of the FOC and solve:

$$c'(e) = \frac{1}{1 + \rho\sigma^2 c''(e^*)}.$$

- Since $b = c'(e)$ from the IC:

$$b^* = \frac{1}{1 + \rho\sigma^2 c''(e^*)}.$$

- **Three drivers** of low bonus:
  - High risk aversion ρ.
  - High noise variance σ^2.
  - High convexity of cost c''.

General cost – Optimal bonus (cont.)

- When $c(e) = \frac{1}{2}e^2$ we have $c'' = 1$, so $b^* = 1/(1 + \rho\sigma^2)$ – exactly the Section 4.3.3 bonus.
- Intuitively, convex cost acts like extra effective risk: a steeper c'' makes each unit of induced effort costlier, so the principal dials the bonus down just as it does when ρ or σ^2 rises.

Example 4.3 – Cost $c(e) = \theta e^2$, $\sigma^2 = 1$

- Here $c'(e) = 2\theta e$ and $c''(e) = 2\theta$, so the Step-5 FOC becomes

$$1 - 2\theta e^* - \rho(2\theta e^*)(2\theta) = 0.$$

- Solve for equilibrium effort:

$$e^* = \frac{1}{2\theta(1 + 2\rho\theta)}.$$

Example 4.3 – Effort and bonus

- Effort decreases in the cost parameter θ:

$$\frac{\partial e^*}{\partial \theta} = -\frac{1+4\rho\theta}{2\theta^2(1+2\rho\theta)^2} < 0,$$

  and likewise in risk aversion ρ.
- The bonus follows from $b^* = 1/(1+\rho\sigma^2 c'')$ with $c'' = 2\theta$ and $\sigma^2 = 1$:

$$b^* = \frac{1}{1+2\rho\theta}.$$

Example 4.3 – Fixed payment a^*

- Evaluate the binding PC $a = \frac{1}{2}\rho b^2 - be + c(e) + \overline{u}$ at (b^*, e^*):

$$a^* = \overline{u} - \frac{1 - 2\rho\theta}{4\theta(1 + 2\rho\theta)^2}.$$

- For small θ the second term can exceed $\overline{u}$, turning a^* negative; then limited liability binds and $a^* = 0$.

Example 4.3 – Fixed payment comparative statics

- a^* rises in risk aversion,

$$\frac{\partial a^*}{\partial \rho} = \frac{3 - 2\rho\theta}{2(1 + 2\rho\theta)^3} > 0 \quad \text{if } \rho\theta < \tfrac{3}{2},$$

- and in the cost parameter,

$$\frac{\partial a^*}{\partial \theta} = \frac{1 + 6\rho\theta - 8\rho^2\theta^2}{4\theta^2(1 + 2\rho\theta)^3} > 0 \quad \text{if } \rho\theta < \tfrac{3+\sqrt{17}}{8}.$$

- Both conditions hold throughout the profitable region, so the fixed payment rises as the worker becomes more risk-averse or effort more costly.

Figures 4.4a & 4.4b – a^* and b^* in θ

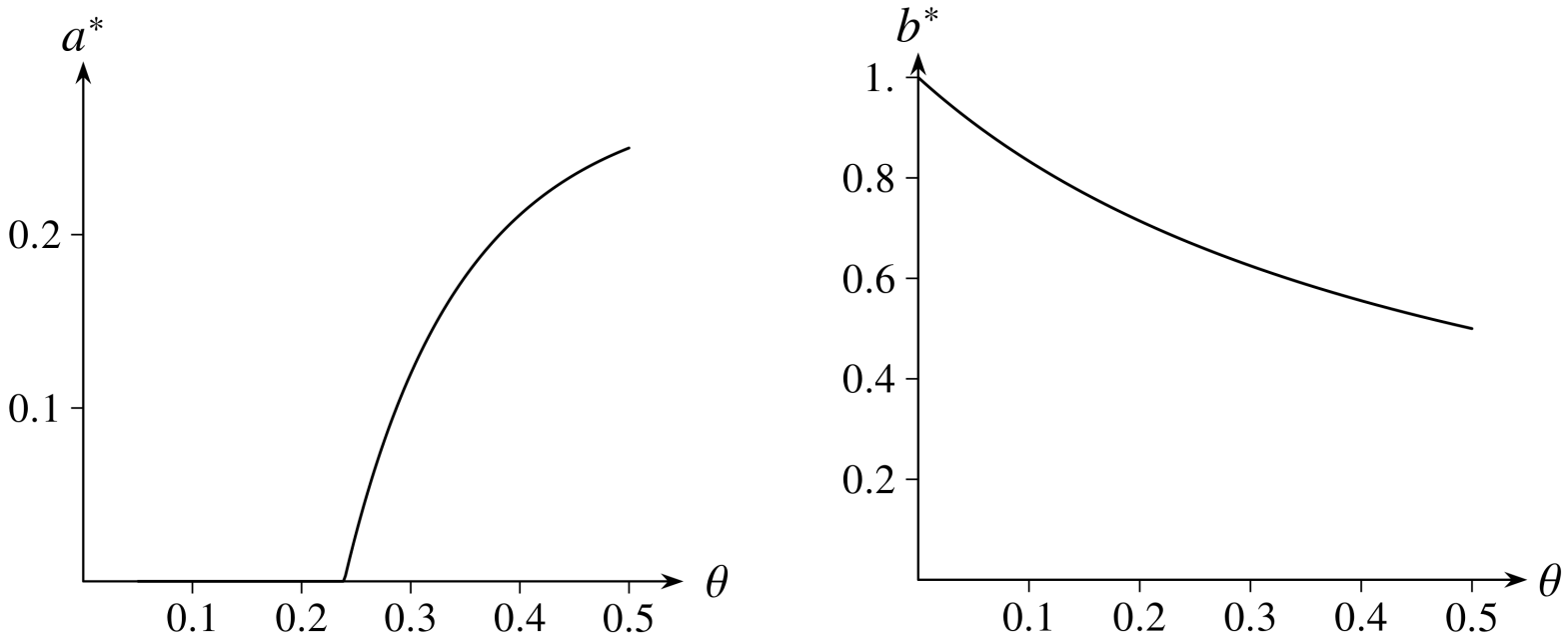


Fig. 4.4a (left). Fixed payment a^* as a function of θ, evaluated at $\rho = 1$.
Fig. 4.4b (right). Bonus b^* in θ.

Table 4.2 – Comparative statics with general cost

	Effort, e^*	Fixed payment, a^*	Bonus, b^*
Agent's risk aversion, ρ	$-$	$+$	$-$
Cost of effort, θ	$-$	$+$	$-$

Table 4.2. Equilibrium effort, fixed payment, and bonus – Example 4.3.

- Higher risk aversion or higher convexity of cost both reduce equilibrium effort and bonus, while raising the fixed payment.
- This makes intuitive sense: when effort is costly or the agent dislikes risk, optimal contracts shift weight from incentive pay to fixed pay.

Example 4.3 – Profitability cutoff on θ

- Equilibrium profit:

$$V^* = \frac{1}{4\theta(1+2\rho\theta)} - \overline{u},$$

non-negative if and only if $\theta \leq \overline{\theta}$, where

$$\overline{\theta} = \frac{\sqrt{\overline{u}^2 + 2\overline{u}\rho} - \overline{u}}{4\overline{u}\rho}.$$

- Comparative statics of the cutoff:

$$\frac{\partial\overline{\theta}}{\partial\rho} < 0 \qquad \text{and} \qquad \frac{\partial\overline{\theta}}{\partial\overline{u}} < 0.$$

- Intuitively, a more risk-averse or more demanding worker shrinks the range of cost parameters θ over which the principal finds it profitable to contract.

Section 4.4
Observing Output or Performance?

Motivation – Performance vs. output

- Often the principal cannot observe output x in time to set the wage; profit may arrive only weeks or months later.
- Instead she observes a **performance** measure p – a proxy correlated with effort but not the same as output x.
- Examples: sales calls made, customer-satisfaction scores, lines of code committed.
- So the contract must be written on p, and the question becomes how well p tracks effort and how the principal should price it.

Performance vs. output – Design implications

- Two random variables:
  - Output: $x = e + \varepsilon$, $E[\varepsilon] = 0$.
  - Performance: $p = \beta e + \phi$, $E[\phi] = 0$, with $\beta \in [0, 1]$.
- β: how much performance reflects effort. $\beta = 1$ means p tracks e perfectly (in expectation); $\beta = 0$ means p is useless.
- Wage based on performance: $w(p) = a + bp$.
- Economically, β is the **signal quality** of the measure: a higher β means p moves more tightly with effort.

4.4.1 Setting up the principal's problem

- With wage written on performance, the principal solves

$$\max_{a,b,e} \; E[x - w(p)]$$

$$\text{s.t. } E[w(p) - c(e)] \geq \overline{u}, \; \text{(PC)}$$

$$\max_{e \geq 0} \; E[w(p) - c(e)]. \; \text{(IC)}$$

4.4.1 Setting up the principal's problem (cont.)

- Substitute $x = e + \varepsilon$, $p = \beta e + \phi$, $w(p) = a + bp$ and use $E[\varepsilon] = E[\phi] = 0$. The objective becomes

$$E[x - w(p)] = e - a - b\beta e = (1 - b\beta)e - a.$$

- The worker's expected utility, in mean-variance form (Appendix 4.A), is

$$E[w(p)] - c(e) = a + b\beta e - \tfrac{1}{2}\rho(\beta b)^2\sigma^2 - c(e),$$

so only the **product** $b\beta$ reaches the worker.

4.4.2 Solving the problem

- Use cost $c(e) = \theta e^2$ for consistency with Section 4.3.
- **Step 1. Agent's FOC** (differentiate the IC in e, with $c'(e) = 2\theta e$): $b\beta - 2\theta e = 0 \Rightarrow b = \dfrac{2\theta e}{\beta}$.
- **Step 2. Binding PC**, solve for a:

$$a = \overline{u} + 2\rho\theta^2 e^2 \sigma^2 - \theta e^2.$$

4.4.2 Solving the problem (cont.)

- **Step 3. Substitute into the objective** $(1 - b\beta)e - a$; the β's cancel because the worker only sees $b\beta = 2\theta e$:

$$\max_{e \geq 0} \left(1 - \frac{2\theta e}{\beta} \cdot \beta\right) e - \left(\overline{u} + 2\rho\theta^2 e^2 \sigma^2 - \theta e^2\right).$$

- **Step 4. Differentiate in e and solve:**

$$e^* = \frac{1}{2\theta(1 + 2\rho\theta\sigma^2)}.$$

Equilibrium bonus and fixed payment

- Equilibrium bonus:

$$b^* = \frac{1}{\beta(1 + 2\rho\theta\sigma^2)}.$$

- Equilibrium fixed payment:

$$a^* = \overline{u} - \frac{1 - 2\rho\theta\sigma^2}{4\theta(1 + 2\rho\theta\sigma^2)^2}.$$

- Effort e^* and fixed payment a^* are **independent of** β; only the bonus rescales as $b^* = \frac{2\theta e^*}{\beta}$.
- At $\beta = 1$, $b^* = \frac{1}{1+2\rho\theta\sigma^2}$, recovering Example 4.3 when $\sigma^2 = 1$.

Figure 4.5 – Effects of β and θ

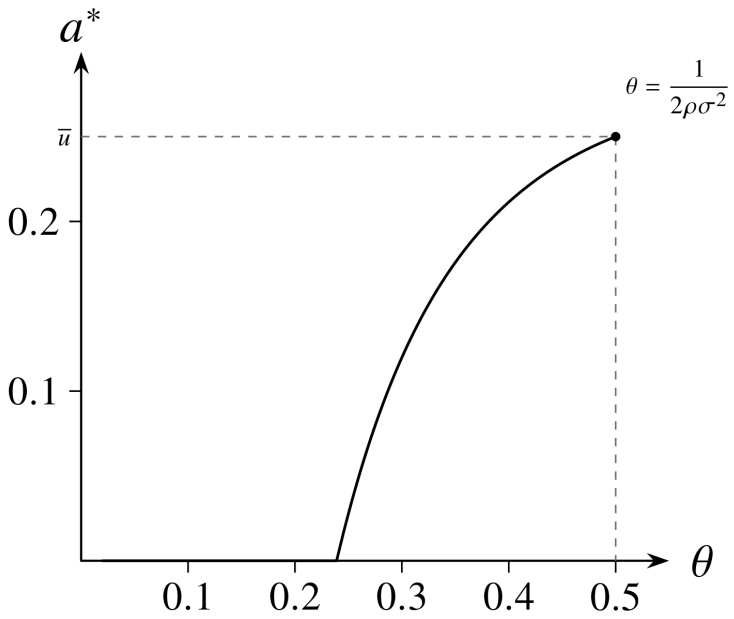


Fig. 4.5a (left). Fixed payment a^*.

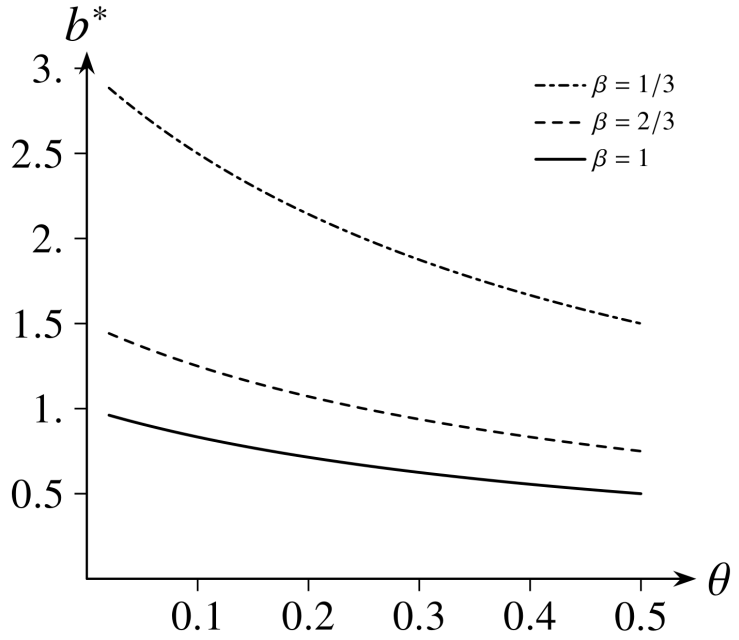


Fig. 4.5b (right). Bonus b^*.

Figure 4.5c – Effort e^* in θ

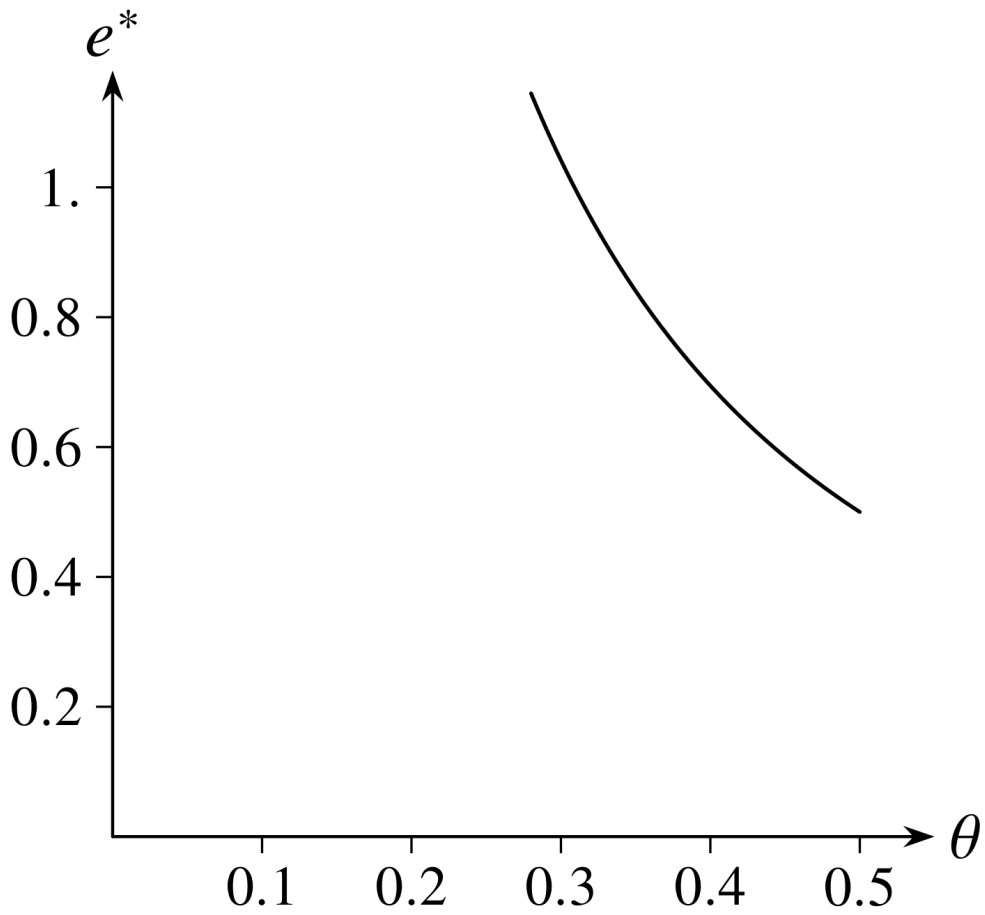


Fig. 4.5c. Effort level e^*, decreasing in θ; independent of β.

Performance vs. output – Intuition

- Higher β (performance better tracks effort):
  - The same incentives need a smaller bonus: b^* falls ($\propto 1/\beta$, Fig. 4.5b).
  - Effort e^* and fixed payment a^* are unchanged.
- Higher θ (effort is more costly):
  - Fixed payment rises, bonus falls, effort falls.
- **Lesson:** the agent responds to βb; a more sensitive measure delivers the same incentives with a smaller bonus, leaving effort and profit unchanged.

Section 4.5
Linear contracts with informative signals

Setup – Informative signals

- As in Section 4.3, output is $x = e + \varepsilon$, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.
- In addition, the principal observes a **signal** s correlated with the noise ε.
- Example: $x =$ this worker's sales, $s =$ another worker's sales – both hit by the same macro shock, so s carries information about ε.

Setup – Informative signals (cont.)

- Joint distribution of the shock and the signal:

$$\begin{pmatrix} \varepsilon \\ s \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon s} \\ \sigma_{\varepsilon s} & \sigma_s^2 \end{bmatrix} \right).$$

- The wage now loads on x and s:

$$w(x, s) = a + b(x + \alpha s),$$

where α is the sensitivity of the bonus to the signal – the parameter of interest here.

Sufficient statistic and informativeness

- Signal s is a **sufficient statistic** (Holmström 1979) if it helps the principal update the probability of effort e given x.
- This occurs whenever $\sigma_{\varepsilon s} \neq 0$ – signal correlates with the noise.
- If $\sigma_{\varepsilon s} = 0$: signal carries no information about effort, principal should ignore it ($\alpha = 0$).

Sufficient statistic and informativeness (cont.)

- Sign of $\sigma_{\varepsilon s}$:
  - Positive: s and ε move together.
  - Negative: s and ε move in opposite directions.
- Question: what is the optimal α^*?

Sufficient statistic – A concrete reading

- A signal s is a **sufficient statistic** for effort given π if, once we know π, s carries no extra information about effort.
- **Contract implication.** If s is a sufficient statistic for π (not for effort), adding s to the contract *is* valuable – s provides orthogonal information.
- **Example – relative performance evaluation.** A CEO's own firm profit is π; the industry average is s. A common sectoral shock hits both, so s helps the principal filter it out.
- **Rule of thumb.** Include every signal that shifts the posterior over effort given the other signals already in the contract.

4.5.1 Setting up the principal's problem

- The principal now also chooses α:

$$\max_{a,b,\alpha,e} \ (1-b)e - a$$

$$\text{s.t. } a + be - \tfrac{1}{2}\text{Var}(w) - c(e) \geq 0, \ \text{(PC)}$$

$$\max_{e \geq 0} \ a + be - \tfrac{1}{2}\text{Var}(w) - c(e). \ \text{(IC)}$$

- The objective is unchanged: s carries information about effort but does not itself add to the firm's output.

4.5.1 Setting up the principal's problem (cont.)

- Since $w = a + b(x + \alpha s)$ and only $x + \alpha s$ is random, the wage variance is

$$\mathrm{Var}(w) = b^2 \mathrm{Var}(\varepsilon + \alpha s) = b^2(\sigma_\varepsilon^2 + \alpha^2 \sigma_s^2 + 2\alpha\sigma_{\varepsilon s}).$$

- Reservation utility is normalized to $\overline{u} = 0$ here, with $\rho = 1$ in the mean-variance form.
- So α enters only through this variance: it is a risk-management dial, not a direct source of output.

4.5.2 Reducing the problem

- **Step 1. Binding PC**, solve for a:
  $$a = c(e) + \frac{b^2 \mathrm{Var}(\varepsilon + \alpha s)}{2} - be.$$
- **Step 2. Substitute into the objective** $(1-b)e - a$:
  $$(1-b)e - a = e - c(e) - \tfrac{1}{2} b^2 \mathrm{Var}(\varepsilon + \alpha s).$$

4.5.2 Reducing the problem (cont.)

- **Step 3. Agent's IC (FOA):** $b = c'(e)$.
- **Step 4. Substitute** $b = c'(e)$, leaving only α and e:

$$\max_{\alpha, e} \; e - \tfrac{1}{2}[c'(e)]^2(\sigma_\varepsilon^2 + \alpha^2\sigma_s^2 + 2\alpha\sigma_{\varepsilon s}) - c(e).$$

- The fixed payment a has dropped out, so the design reduces to the signal weight α and effort e.

4.5.3 Optimal signal weight α^*

- Only the variance term depends on α; differentiate it and set to zero:

$$-\alpha\sigma_s^2 - \sigma_{\varepsilon s} = 0.$$

- Solve for the optimal signal weight:

$$\alpha^* = -\frac{\sigma_{\varepsilon s}}{\sigma_s^2}.$$

- Three cases:
  - **Uncorrelated** ($\sigma_{\varepsilon s} = 0$): $\alpha^* = 0$ – ignore signal.
  - **Positively correlated** ($\sigma_{\varepsilon s} > 0$): $\alpha^* < 0$ – shrink bonus when signal is high (favorable conditions).
  - **Negatively correlated** ($\sigma_{\varepsilon s} < 0$): $\alpha^* > 0$ – increase bonus (more impressive output).

Informativeness principle – The big picture

- Holmström's (1979) **informativeness principle**:
  "A signal should be included in the contract if and only if it provides additional information about the agent's action beyond what is already contained in observed output."
- Practical application: **relative performance evaluation** – CEOs paid based on their firm's return *relative* to peers, filtering out common shocks.
- Reduces risk borne by the agent without sacrificing incentives.

Informativeness principle – The big picture (cont.)

- Variance with optimal α^*:

$$\text{Var}(\varepsilon + \alpha^* s) = \sigma_\varepsilon^2 - \frac{\sigma_{\varepsilon s}^2}{\sigma_s^2} < \sigma_\varepsilon^2.$$

- **Conclusion:** signals reduce salary variance, allowing a **higher bonus** and **more effort**.

4.5.4 Equilibrium bonus

- **FOC in** e, then factor out $c'(e)$:

$$1 - c'(e)c''(e)\mathrm{Var}(\varepsilon + \alpha s) - c'(e) = 0.$$

- Solve and use $b = c'(e)$:

$$b^* = \frac{1}{1 + c''(e)\mathrm{Var}(\varepsilon + \alpha s)}.$$

4.5.4 Equilibrium bonus (cont.)

- Insert α^*, so $\mathrm{Var}(\varepsilon + \alpha^* s) = \sigma_\varepsilon^2 - \sigma_{\varepsilon s}^2/\sigma_s^2$:

$$b^* = \frac{1}{1 + c''(e)\left(\sigma_\varepsilon^2 - \dfrac{\sigma_{\varepsilon s}^2}{\sigma_s^2}\right)}.$$

- Whenever $\sigma_{\varepsilon s} \neq 0$ the bracket shrinks, so b^* is **higher** than without the signal: the informative signal cuts the worker's risk, letting the principal reward output more steeply.

4.5.4 Bonus, effort, and comparative statics

- Write $\Delta \equiv \sigma_\varepsilon^2 \sigma_s^2 - \sigma_{\varepsilon s}^2 > 0$ (positive definiteness). With $c(e) = \theta e^2$,

$$b^* = \frac{\sigma_s^2}{\sigma_s^2 + 2\theta\Delta}, \qquad e^* = \frac{\sigma_s^2}{2\theta\left(\sigma_s^2 + 2\theta\Delta\right)}.$$

- A more correlated signal (higher $\sigma_{\varepsilon s}^2$) raises **both**:

$$\frac{\partial b^*}{\partial \sigma_{\varepsilon s}^2} = \frac{2\theta\sigma_s^2}{\left(\sigma_s^2 + 2\theta\Delta\right)^2} > 0, \qquad \frac{\partial e^*}{\partial \sigma_{\varepsilon s}^2} = \frac{\sigma_s^2}{\left(\sigma_s^2 + 2\theta\Delta\right)^2} > 0.$$

4.5.4 Bonus, effort, and comparative statics (cont.)

- The fixed payment rises with signal correlation only when effort is costly enough:

$$\frac{\partial a^*}{\partial \sigma_{\varepsilon s}^2} > 0 \iff \theta > \frac{3\sigma_s^2}{2\Delta}.$$

- Intuitively, a sharper signal first lets the principal shift pay toward the bonus (lowering a^*); only when effort is expensive does she also raise the fixed payment.

4.5.4 (cont.) – Equilibrium profit

- Equilibrium profit is **always positive**:

$$V^* = \frac{\sigma_s^2}{4\theta\left(\sigma_s^2 + 2\theta\Delta\right)} > 0.$$

- Profit rises when the signal is more informative about effort:

$$\frac{\partial V^*}{\partial \sigma_{\varepsilon s}^2} = \frac{\sigma_s^2}{2\left(\sigma_s^2 + 2\theta\Delta\right)^2} > 0.$$

- **Bottom line.** Adding an informative signal is strictly valuable for the principal, and the value grows with how tightly the signal is correlated with the noise in output.
- That is exactly the content of Holmström's informativeness principle: signals that carry information about the hidden action belong in the contract.

Example 4.4 – Numerical illustration

- Parameters: $\sigma_\varepsilon^2 = \sigma_s^2 = 1$, $\sigma_{\varepsilon s} = \frac{1}{2}$, $c(e) = \theta e^2$.
- Optimal signal weight, from $\alpha^* = -\sigma_{\varepsilon s}/\sigma_s^2$:

$$\alpha^* = -\frac{1/2}{1} = -\frac{1}{2}.$$

- Effective-signal variance at the optimum, $\sigma_\varepsilon^2 - \sigma_{\varepsilon s}^2/\sigma_s^2$:

$$\mathrm{Var}(\varepsilon + \alpha^* s) = 1 - \frac{1}{4} = \frac{3}{4} < 1,$$

so the signal cuts wage risk by a quarter.

Example 4.4 – Numerical illustration (cont.)

- With $\Delta = \sigma_\varepsilon^2 \sigma_s^2 - \sigma_{\varepsilon s}^2 = \frac{3}{4}$, the general formulas give the effort

$$e^* = \frac{1}{\theta(2+3\theta)}.$$

- Bonus:

$$b^* = \frac{2}{2+3\theta}.$$

- Fixed payment:

$$a^* = \frac{3\theta - 2}{2\theta(2+3\theta)^2}.$$

Example 4.4 – Comparing to no-signal case

- Versus Example 4.3 (no signal, same θ, $\sigma^2 = 1$):
  - Effort: $\dfrac{1}{\theta(2+3\theta)} > \dfrac{1}{2\theta(1+2\theta)}$ – **higher**.
  - Bonus: $\dfrac{2}{2+3\theta} > \dfrac{1}{1+2\theta}$ – **higher**.
  - Fixed payment: $\dfrac{3\theta-2}{2\theta(2+3\theta)^2} < \overline{u} - \dfrac{1-2\theta}{4\theta(1+2\theta)^2}$ – **lower**.
- Adding an informative signal:
  - Reduces the agent's wage uncertainty.
  - Allows higher bonus, higher effort, lower fixed payment.
  - Increases efficiency.

Section 4.6
Moral Hazard with Multiple Tasks

Multitask – Setup

- The worker now has **two** effort choices, e_1 and e_2.
- Example: a salesman splits time between customer visits (e_1) and building a website or learning new sales tools (e_2).
- Cost is separable and convex, with no interaction between tasks:

$$c(e_1, e_2) = \tfrac{1}{2}e_1^2 + \tfrac{1}{2}e_2^2.$$

Multitask – Setup (cont.)

- Output depends on both efforts:

$$x = e_1 + \gamma e_2 + \varepsilon, \quad \varepsilon \sim F.$$

- Parameter γ is the efficacy of task 2: $\gamma = 0$ means e_2 is irrelevant, $0 < \gamma < 1$ means it matters less than e_1, $\gamma > 1$ means it matters more, and $\gamma < 0$ means it harms output.
- The principal still offers a single linear contract $w(x) = a + bx$ on total output.

4.6.1 Principal's problem

- The objective nets the wage out of total output, $(1-b)(e_1+\gamma e_2)-a$, so the principal solves

$$\max_{a,b,e_1,e_2} \ (1-b)(e_1+\gamma e_2)-a$$

$$\text{s.t. } a+b(e_1+\gamma e_2)-\tfrac{1}{2}\text{Var}(w)-c(e_1,e_2)\geq 0, \ \text{(PC)}$$

$$\max_{e_1,e_2\geq 0} \ a+b(e_1+\gamma e_2)-\tfrac{1}{2}\text{Var}(w)-c(e_1,e_2). \ \text{(IC)}$$

4.6.1 Principal's problem (cont.)

- Now **four** choice variables, a, b, e_1, e_2, with a single bonus b priced on total output.
- We apply the FOA to **both** efforts: the worker's two first-order conditions replace his argmax.
- The tension is that one bonus must steer two tasks at once, so the mix of effort the worker picks is not something the principal sets directly.

4.6.2 Agent's FOCs

- With $\text{Var}(w) = b^2\sigma^2$, the worker solves

$$\max_{e_1, e_2 \geq 0} \; a + b(e_1 + \gamma e_2) - \tfrac{1}{2}b^2\sigma^2 - \tfrac{1}{2}e_1^2 - \tfrac{1}{2}e_2^2.$$

- **FOC in e_1:** $b - e_1 = 0 \Rightarrow b = e_1$.
- **FOC in e_2:** $b\gamma - e_2 = 0 \Rightarrow e_2 = b\gamma = \gamma e_1$.
- So the worker splits effort **proportionally**, $e_2 = \gamma e_1$: the better task 2 converts into output, the more of it he does.

Solving the principal's problem (multitask)

- **Step 1. Substitute** $b = e_1$ and $e_2 = \gamma e_1$ into the program:

$$\max_{a, e_1} (1 + \gamma^2) e_1 (1 - e_1) - a$$

$$\text{s.t. } a + (1 + \gamma^2) e_1^2 - (1 + \gamma^2 + \sigma^2) \tfrac{1}{2} e_1^2 \geq 0. \text{ (PC)}$$

Solving the principal's problem (multitask, cont.)

- **Step 2. Binding PC**, solve for a:

$$a = \tfrac{1}{2}(\sigma^2 - \gamma^2 - 1)e_1^2.$$

- **Step 3. Substitute into the objective**, leaving one variable:

$$\max_{e_1 \geq 0} (1 + \gamma^2)e_1(1 - e_1) - \tfrac{1}{2}(\sigma^2 - \gamma^2 - 1)e_1^2.$$

4.6.3 Equilibrium efforts and bonus

- **Step 4. FOC in e_1:**

$$(1+\gamma^2)(1-2e_1) - (\sigma^2 - \gamma^2 - 1)e_1 = 0.$$

- Solve for the first effort, which equals the bonus:

$$e_1^* = b^* = \frac{1+\gamma^2}{1+\gamma^2+\sigma^2}.$$

- The bonus rises with task-2 efficacy, $\dfrac{\partial b^*}{\partial \gamma} = \dfrac{2\gamma\sigma^2}{(1+\gamma^2+\sigma^2)^2} > 0$: a more productive second task makes output a better target to reward.

Equilibrium efforts and bonus (cont.)

- Second effort:

$$e_2^* = \gamma e_1^* = \frac{\gamma(1+\gamma^2)}{1+\gamma^2+\sigma^2}.$$

- When $\gamma = 0$: $e_2^* = 0$ and $e_1^* = \dfrac{1}{1+\sigma^2}$, recovering Section 4.3 (with $\theta = 1/2$, $\rho = 1$).

4.6.3 (cont.) – Effect of task-2 efficacy

- Comparative statics with respect to γ (efficacy of task 2):

$$\frac{\partial V^*}{\partial \gamma} = \frac{\gamma\left(1+\gamma^2\right)\left(1+\gamma^2+2\sigma^2\right)}{\left(1+\gamma^2+\sigma^2\right)^2} > 0.$$

- **Reading.** Both worker and principal gain from empowering task 2 – the second effort is a complement, not a distraction.
- **Reading the multitask fix.** When tasks reinforce each other ($\gamma > 0$), the principal safely rewards output; when they compete for the worker's attention, the principal has to be more cautious about high-powered incentives on the measured task.
- **Numerical illustration.** With $\sigma^2 = 1$, doubling γ from 0.5 to 1 raises b^* from 0.56 to 0.67 and profit by a similar percentage.

Equilibrium fixed payment a^*

- Substitute e_1^* into $a = \frac{1}{2}(\sigma^2 - \gamma^2 - 1)e_1^2$:

$$a^* = \frac{(\sigma^2 - \gamma^2 - 1)(1 + \gamma^2)^2}{2(1 + \gamma^2 + \sigma^2)^2}.$$

- Its sign in γ turns on a quartic bracket:

$$\frac{\partial a^*}{\partial \gamma} = \frac{-\gamma(1 + \gamma^2)\left[\gamma^4 + (2 + 3\sigma^2)\gamma^2 + (1 + 3\sigma^2 - 2\sigma^4)\right]}{(1 + \gamma^2 + \sigma^2)^3}.$$

Equilibrium fixed payment – Calibration $\sigma^2 = 2$

- At $\sigma^2 = 2$ (used in Fig. 4.6 so that a^* stays positive):

$$a^* = \frac{(1-\gamma^2)(1+\gamma^2)^2}{2(3+\gamma^2)^2}, \qquad b^* = \frac{1+\gamma^2}{3+\gamma^2}.$$

- The fixed payment is **non-monotonic** in γ (rising for low γ, then falling), while the bonus increases throughout.

Reading a^* and b^*

- The bonus $b^* = \dfrac{1}{1+\rho\sigma^2}$ has a clean reading:
  - If $\rho = 0$ (risk-neutral worker), $b^* = 1$ – a franchise contract that hands the whole marginal profit to the worker.
  - If $\rho\sigma^2$ is large (very risk-averse worker or very noisy output), $b^* \to 0$ – the wage becomes almost flat, recovering the Chapter 2 result.
- So the bonus interpolates between the two Chapter-2 cases we already know, with the intensity of the interpolation set by $\rho\sigma^2$.
- **The fixed payment** a^* shifts to keep PC binding. It absorbs the risk premium the worker demands for facing income uncertainty, so the total expected pay meets $\overline{u}$.

Reading a^* and b^* (cont.)

- **Two moving parts, one job each:** the bonus b^* handles the IC (it sets the worker's effort), while the fixed payment a^* handles the PC (it sets his pay level).
- Economically, this split is why the comparative statics stay clean: a change in ρ, σ^2, or θ moves b^* through incentives and a^* through participation, without the two getting tangled.

Figure 4.6 – Multitask comparative statics

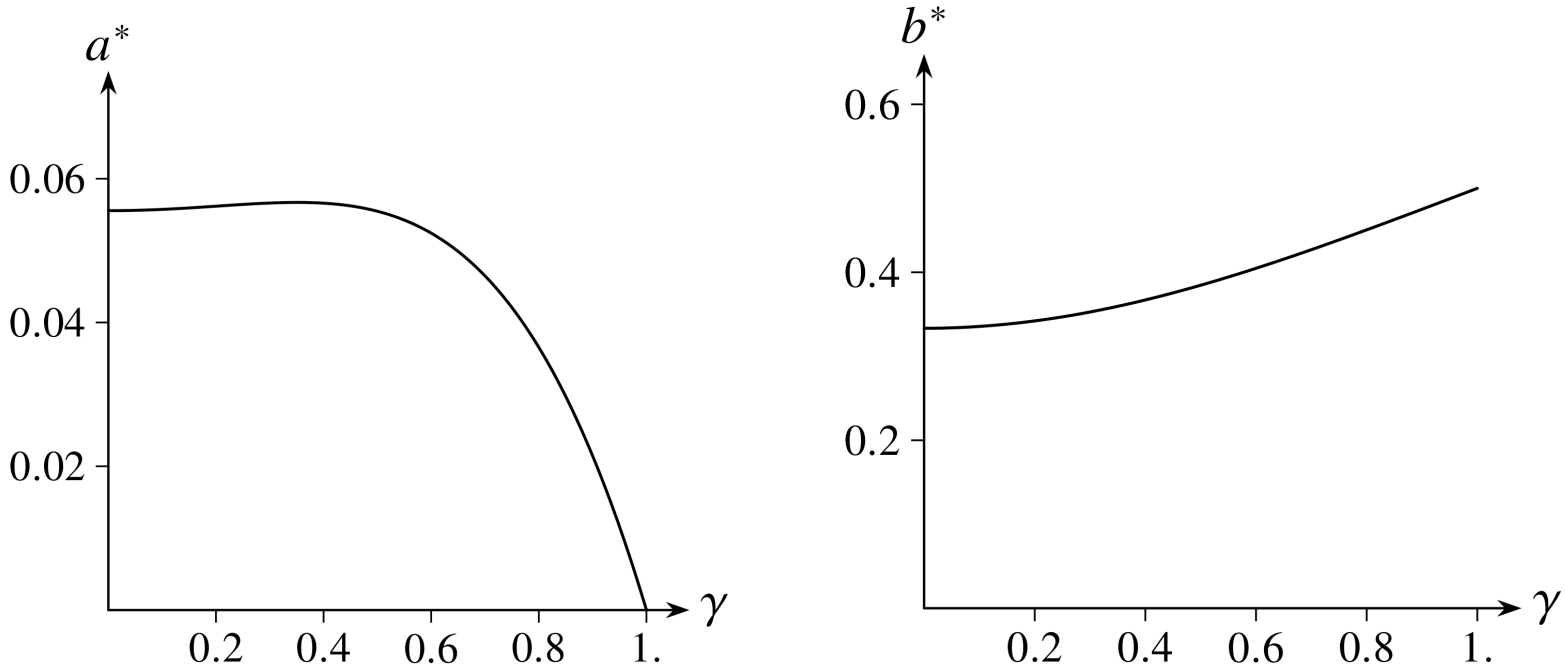


Fig. 4.6a (left). Fixed payment a^* non-monotonic in γ. Fig. 4.6b (right). Bonus b^* increasing in γ.

Figure 4.6c – Effort e_2^*

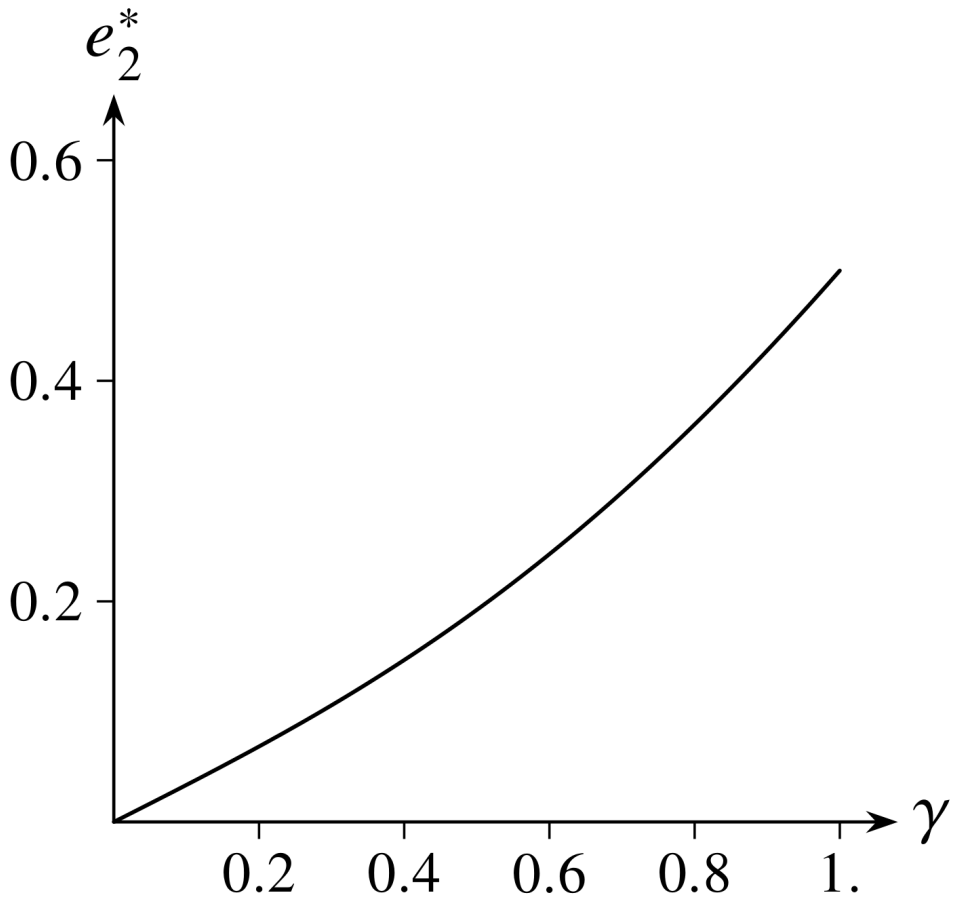


Fig. 4.6c. Effort level e_2^* increasing in γ. When $\gamma = 0$, $e_2^* = 0$ (irrelevant task).

Multitask – Holmström-Milgrom insight

- When the agent has multiple tasks, incentives on one task **affect effort on others**.
- Here: higher γ raises b^* AND raises e_2^*.
- More general principle (Holmström-Milgrom 1991): incentives should be **balanced** across tasks.
- Application: do not pay teachers based only on test scores – this distorts effort toward test prep and away from other valuable activities ("teaching to the test").
- When one task is hard to measure, optimal contract reduces bonus on *all* tasks to prevent distortion.
- Key takeaway: **multi-tasking constrains the use of high-powered incentives**.

Multitask – The teaching lesson

- Real jobs bundle several tasks; not all are equally measurable.
  - A teacher's task is student learning; the measurable proxy is test scores.
  - A salesperson's task is customer relationship; the measurable proxy is quarterly revenue.
- **Result.** High-powered incentives on the measurable task divert effort away from the unmeasurable one – test-score gaming, customer poaching, short-termism.
- **Design responses:**
  - *Low-powered incentives.* Flatter compensation preserves effort on unmeasurable tasks.
  - *Multiple measures.* Add proxies for the second task (student surveys, customer retention scores).
  - *Split the jobs.* Assign the two tasks to different agents when possible.

Section 4.7
Summary

Summary – First-order approach

- Continuous effort makes the IC an argmax over an interval – the **first-order approach** replaces it with the agent's FOC (Eq. 4.2).
- Generalized MRS-type condition (Eq. 4.3):

$$\lambda + \mu \frac{f'(\pi_i|e)}{f(\pi_i|e)} = \frac{1}{u'(w_i)}.$$

Summary – First-order approach (cont.)

- Key term: $\dfrac{f'(\pi_i|e)}{f(\pi_i|e)}$, best read as an **elasticity** of f in effort (high profit more elastic in effort $\Rightarrow$ higher pay).

Summary – First-order approach (cont.)

- Wages under AI exceed wages under SI (as in Chapter 3).
- *Caveat – when is the FOA valid?* Replacing the IC by its FOC is justified only under conditions such as the **monotone likelihood-ratio property** and the **convexity of the distribution function** (Rogerson, 1985). The agent's utility levels are endogenous, so the FOC is necessary but not, in general, sufficient.

Summary – Linear contracts under CARA

- Setup: $x = e + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$, CARA agent, $c(e) = \frac{1}{2}e^2$.
- Optimal linear contract:

$$b^* = \frac{1}{1+\rho\sigma^2}, \quad e^* = b^*, \quad a^* = \overline{u} - \frac{1-\rho\sigma^2}{2\left(1+\rho\sigma^2\right)^2}.$$

- General cost: $b^* = \dfrac{1}{1+\rho\sigma^2 c''\left(e^*\right)}$.
- Comparative statics: more noise or risk aversion $\Rightarrow$ lower bonus, higher fixed payment, lower effort.

Summary – Extensions

- **Performance vs. output** (Section 4.4): wage based on $p = \beta e + \phi$. Higher $\beta \Rightarrow$ higher bonus, more effort.
- **Informative signals** (Section 4.5): optimal weight $\alpha^* = -\sigma_{\varepsilon s}/\sigma_s^2$. Adding informative signals raises bonus, effort, and reduces risk premium – **informativeness principle**.
- **Multitask** (Section 4.6): agent allocates effort $e_2 = \gamma e_1$; bonus and efforts increase in γ. Strong incentives can distort effort across tasks.

Appendix 4.A
Certainty Equivalent as Mean and Variance

Appendix 4.A – Setup

- Bernoulli utility $u(x)$, income $x \sim N(\mu, \sigma^2)$.
- Goal: express the certainty equivalent CE as a function of μ, σ^2.
- Certainty equivalent definition:

$$u(CE) = E[u(x)].$$

- Apply first-order Taylor on LHS, second-order on RHS:

$$\begin{aligned} u(\mu) + u'(\mu)(CE - \mu) &= E[u(\mu) + u'(\mu)(x - \mu) \\ &\quad + \tfrac{1}{2}u''(\mu)(x - \mu)^2]. \end{aligned}$$

Appendix 4.A – Derivation

- **Step 1. Take expectations**, using $E[x] = \mu$ and $E[(x-\mu)^2] = \sigma^2$:

$$u(\mu) + u'(\mu)(CE - \mu) = u(\mu) + \frac{\sigma^2}{2}u''(\mu).$$

- **Step 2. Cancel** $u(\mu)$ and solve for CE:

$$CE = \mu + \frac{\sigma^2}{2}\frac{u''(\mu)}{u'(\mu)}.$$

Appendix 4.A – Derivation (cont.)

- **Step 3. Insert Arrow-Pratt** $\rho = -u''/u'$:

$$CE = \mu - \tfrac{1}{2}\rho\sigma^2.$$

- **Step 4. Apply to the wage** in Section 4.3, with $\mu = E(w)$ and $\sigma^2 = \text{Var}(w)$:

$$CE = E(w) - \tfrac{1}{2}\rho\text{Var}(w).$$

Appendix 4.A – Interpretation

- Certainty equivalent of a risky income $x \sim N(\mu, \sigma^2)$:
  - **Increases** in mean income μ.
  - **Decreases** in variance σ^2 (agent dislikes risk).
  - **Decreases** in risk aversion ρ.
- Risk premium: $RP = \mu - CE = \frac{1}{2}\rho\sigma^2$.
- The mean-variance form is exact for CARA + normal, and a good approximation otherwise (for small variance).

Appendix 4.A – Interpretation (cont.)

- This makes the linear-contract analysis tractable: it is the foundation of Section 4.3.
- One caveat: under a no-bankruptcy (limited-liability) constraint the certainty equivalent cannot be negative, which requires $E(w) \geq \frac{1}{2}\rho \mathrm{Var}(w)$ – expected pay must at least cover the risk premium.

Looking ahead – Chapter 5

- Chapters 2–4 covered **moral hazard** (hidden action).
- Chapter 5 turns to **adverse selection** (hidden type):
  - Agent's type θ is private information *before* contracting.
  - Akerlof's "Market for Lemons" – adverse selection unravels markets.
  - Threshold probability $\bar{q} = \dfrac{s_H - b_L}{b_H - b_L}$ determines whether the market exists.
  - Signaling (Spence): agent sends a costly signal to reveal type.
  - Screening: principal offers a menu of contracts to separate types.
  - PBE refinement via the Intuitive Criterion.
- Tools used: Bayesian games, perfect Bayesian equilibrium.

Common student mistakes to avoid (Chapter 4)

- **Using FOA outside its comfort zone.** The first-order approach needs MLRP + CDFC to be valid. Drop those and the FOC no longer identifies the agent's optimum.
- **Reading b^* as the wage.** b^* is the *slope* of the linear contract, not the wage. The wage is $a^* + b^*\pi$.
- **Forgetting the risk premium.** a^* absorbs both the outside option *and* the risk premium; ignoring either mismeasures the PC.

Common student mistakes (Chapter 4) (cont.)

- **Assuming an informative signal is always a bonus.** A signal that is a sufficient statistic for effort given π adds nothing to the contract.
- **Treating multitask as a curiosity.** In multitask settings, high-powered incentives can backfire – an important practical lesson, not a technical corner case.

Key takeaways from Chapter 4

1. Continuous effort $\Rightarrow$ the IC argmax becomes a single FOC. This is the **first-order approach**.
2. Under CARA + normal output, optimal linear contract has closed form:
$$b^* = \frac{1}{1+\rho\sigma^2}, \quad e^* = b^*.$$
3. Higher risk aversion or noise lowers the bonus, raises the fixed payment.
4. **Informativeness principle:** a signal reduces wage variance and allows higher bonus and effort – $\alpha^* = -\sigma_{\varepsilon s}/\sigma_s^2$.
5. **Multitask**: incentives on one task affect effort on others; multi-tasking constrains high-powered incentives.
6. This chapter completes the moral hazard toolkit. Chapter 5 turns to the opposite information problem – adverse selection.

End of Chapter 4

Chapter 5

Adverse Selection

Instructor Slides

Contract Theory: An Introduction with Applications

Ana Espinola-Arredondo, Felix Muñoz-Garcia, and Pak-Sing Choi

Outline of Chapter 5

- This chapter shifts from **moral hazard** (hidden action) to **adverse selection** (hidden type).
- Sections:
  - 5.1 Introduction – Akerlof's lemons problem.
  - 5.2–5.3 The lemons model – two car types.
  - 5.4 Continuous qualities and unraveling.
  - 5.5 Application to labor markets.
  - 5.6 Solutions: warranties, screening, signaling.
  - 5.7–5.8 Signaling games – separating and pooling PBEs.
  - 5.9 Refinement – ruling out implausible beliefs.
  - 5.10 Summary. 5.11 Exercises.

What you will learn

- How **hidden types** (rather than hidden actions) can cause markets to fail.
- The **threshold** probability $\bar{q} = \dfrac{s_H - b_L}{b_H - b_L}$ for trade to occur in equilibrium.
- The **unraveling result**: under continuous qualities, the market shrinks to the lowest type.
- How **warranties**, **screening**, and **signaling** can mitigate adverse selection.
- How to use the **5-step procedure (Tool 5.1)** to find Perfect Bayesian Equilibria (PBEs) in signaling games.
- Why **equilibrium dominance** refines away implausible PBEs.

Section 5.1
Introduction

What is adverse selection?

- **Adverse selection** arises when one party has private information about a relevant characteristic ("type") before contracting.
- Akerlof (1970), "The Market for Lemons" – Nobel Prize 2001 – showed that adverse selection can:
  - Cause efficient trades to fail to occur.
  - Lead markets to **unravel** – only the worst types trade.
  - Generate persistent inefficiency that price signals alone cannot fix.
- Classic example: used cars.
  - Seller knows the car's quality (peach vs lemon).
  - Buyer cannot distinguish them at the time of sale.

What is adverse selection? (cont.)

- Same logic applies to many markets: insurance, credit, labor, online platforms.

Adverse selection vs. moral hazard – Side by side

- Both involve asymmetric information, but the *timing* is different.
  - **Moral hazard** (Chapters 2–4): information becomes asymmetric *after* the contract, because the agent chooses a hidden action.
  - **Adverse selection** (this chapter and Chapter 6): information is already asymmetric *before* the contract, because the agent knows his own type.
- **Fix that follows.**
  - Moral hazard: design a wage schedule so the agent *wants* to take the right action.
  - Adverse selection: design a menu (or a signal, or a screening device) so each type *reveals* what he knows.

Adverse selection vs. moral hazard – The fix

- Different friction, different fix – but both live under the same principal-agent umbrella.

Roadmap for Chapter 5

- **Sections 5.2–5.5** – demonstrating the problem:
  - Two-type lemons model (5.2).
  - Generalize valuations (5.3).
  - Continuous qualities and unraveling (5.4).
  - Labor markets (5.5).
- **Sections 5.6–5.9** – solutions and signaling games:
  - Warranties, certification, screening.
  - Signaling: Spence's labor-market game.
  - Separating and pooling Perfect Bayesian Equilibria (PBEs).
  - Refinement by equilibrium dominance.

Section 5.2
The Market for Lemons – Two Car Types

Setup – Peaches and lemons

- Two types of cars:
  - **Peach** (high quality): buyer values at \$3,000; seller values at \$2,000.
  - **Lemon** (low quality): buyer values at \$1,000; seller values at \$100.
- Prices given exogenously: p_H for peaches, p_L for lemons.
- Seller observes the car's quality; buyer does *not*.

Setup – Peaches and lemons (cont.)

- Payoffs if sale occurs:
  - Buyer: $\$3{,}000 - p_H$ (peach) or $\$1{,}000 - p_L$ (lemon).
  - Seller: $p_H - \$2{,}000$ (peach) or $p_L - \$100$ (lemon).
- If no sale: buyer gets 0; seller keeps the car (worth \$2,000 or \$100).

5.2.1 Symmetric information benchmark

- **Step 1. State who sees what.** Both buyer and seller observe the car's quality, so the buyer knows whether he faces a peach or a lemon before deciding.
- **Step 2. Buyer's rule for a peach.** He buys iff the surplus is nonnegative, since a negative surplus is worse than walking away (utility 0):

$$\$3{,}000 - p_H \geq 0 \iff p_H \leq \$3{,}000.$$

  The buyer pays at most his own valuation of the peach.
- **Step 3. Buyer's rule for a lemon.** The same logic with the lemon valuation:

$$\$1{,}000 - p_L \geq 0 \iff p_L \leq \$1{,}000.$$

- Each type is priced on its own value because quality is observed – no inference is needed.

Symmetric information benchmark (cont.)

- **Step 4. Seller's rule.** The seller accepts a sale only if the price covers what the car is worth to her, since otherwise keeping the car is better:

$$p_H \geq \$2{,}000 \quad \text{(peach)}, \qquad p_L \geq \$100 \quad \text{(lemon)}.$$

- **Step 5. Overlap the two rules.** A trade occurs at any price agreeable to both sides:
  - Peaches traded at $p_H \in [\$2{,}000, \$3{,}000]$.
  - Lemons traded at $p_L \in [\$100, \$1{,}000]$.
- Both ranges are nonempty because the buyer values each type above the seller – gains from trade exist for *every* car.
- **Both car types are sold** – no market failure.
- Numeric check: a peach at \$2,900 gives the buyer $\$3{,}000 - \$2{,}900 = \$100 > 0$ and the seller $\$2{,}900 - \$2{,}000 = \$900 > 0$, so both agree.

Why the symmetric-information benchmark matters

- Under symmetric information the buyer sees the seller's type directly and pays the type-specific value.
  - Every peach and every lemon trades at its fair price.
  - Total surplus is maximized – the first-best benchmark.
- **Take this as the target.** Every result under asymmetric information (unraveling, warranty menus, signaling) will be measured against this benchmark.
- **Same role as Chapter 2 played in Chapters 3–4.** The benchmark isolates the pure informational friction from other market frictions.

Why the SI benchmark matters (cont.)

- Realistic instances of near-symmetric information: certified pre-owned cars, quality-audited medical devices, credit ratings after formal disclosure – each an example of institutions that push the market closer to the benchmark.

5.2.2 Asymmetric information – Setup

- Now the seller observes quality but the buyer only sees the price p, so the buyer must *infer* quality from the price he is quoted.
- Buyer's prior: the car is a peach with probability $q \in [0, 1]$ and a lemon with probability $1 - q$, where q reflects the market-wide proportion of peaches for that make and model.
- The price itself carries information, because a type is offered at a given price only when that sale is profitable for the seller. We therefore split the price line into two regions:
  1. Low prices $p \leq \$2{,}000$ – a peach cannot be profitably offered (its sale would lose money), so only lemons appear.
  2. High prices $p > \$2{,}000$ – a peach *can* be profitably offered, so both types may appear.
- The cutoff \$2,000 is the seller's peach valuation: below it she would rather keep a peach than sell it.

Low prices region $p \leq \$2{,}000$

- **Step 1. Read the price.** At $p \leq \$2{,}000$ a peach sale would earn the seller $p - \$2{,}000 \leq 0$, so no peach is offered here.
- **Step 2. Infer the type.** Seeing such a price, the buyer correctly concludes *the car must be a lemon* – the price has fully revealed quality.
- **Step 3. Apply the buyer's rule for a lemon.** He buys iff his surplus is nonnegative:

$$\$1{,}000 - p \geq 0 \iff p \leq \$1{,}000.$$

- Result: lemons trade at $p \in [\$100, \$1{,}000]$, exactly the symmetric-information range.
- Intuitively, the lemon market is undamaged: a low price is bad news, but it is *correct* news, so no efficient lemon trade is lost.

High prices region $p > \$2{,}000$

- **Step 1. Both types may appear.** At $p > \$2{,}000$ a peach sale is profitable too, so the price no longer reveals the type.
- **Step 2. Write expected utility.** The buyer weighs each type by its probability and buys iff the expected surplus is nonnegative:

$$q(\$3{,}000 - p) + (1 - q)(\$1{,}000 - p) \geq 0.$$

- **Step 3. Simplify.** Collecting the valuation terms and the price:

$$\$3{,}000q + \$1{,}000(1 - q) \geq p \iff \$1{,}000 + \$2{,}000q \geq p.$$

- The left side is the buyer's *expected value* of the car; he pays at most that amount.

High prices region $p > \$2{,}000$ (cont.)

- **Step 4. Impose both conditions.** A high-price purchase needs the price to clear the seller's peach floor *and* stay within the buyer's expected value:

$$\$2{,}000 < p \leq \$1{,}000 + \$2{,}000q.$$

  - Lower bound $p > \$2{,}000$: the seller offers a peach only above it.
  - Upper bound $p \leq \$1{,}000 + \$2{,}000q$: the buyer pays no more than expected value.
- The pairs (p, q) satisfying both make up **region** B in Figure 5.1.
- Numeric check at $q = \frac{2}{3}$: the band is $\$2{,}000 < p \leq \$1{,}000 + \$2{,}000(\frac{2}{3}) = \$2{,}333.33$, a nonempty interval, so peaches can trade.

Figure 5.1 – Purchase region B

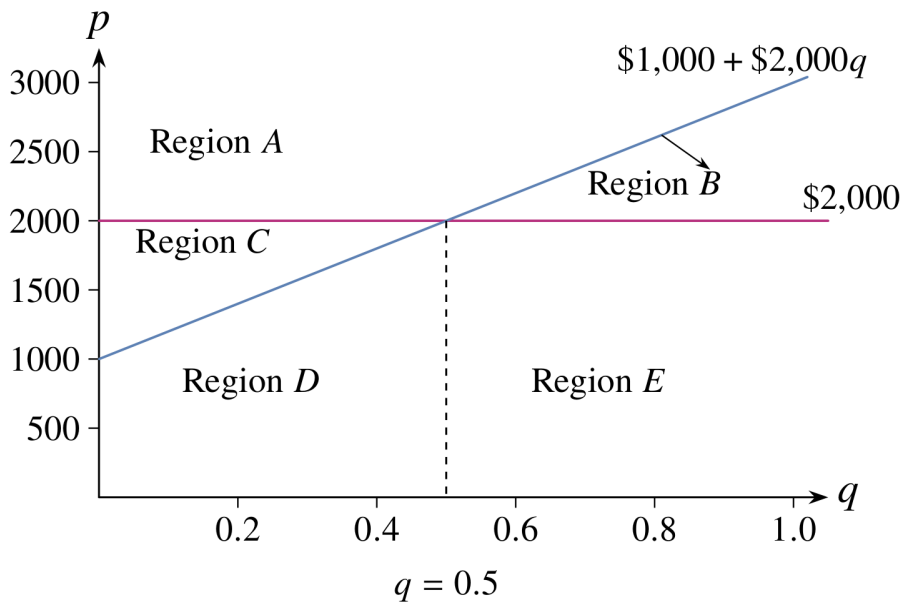


Figure 5.1. Purchases only occur in region B, where $p \leq \$1{,}000 + \$2{,}000q$ and $p > \$2{,}000$ both hold. Region B exists only for $q > 1/2$.

- Region B is the sliver above $p = \$2{,}000$ and below the expected-value line, opening up only once peaches are common enough.

Market failure when $q \leq 1/2$

- **Step 1. Region B must be nonempty.** Both bounds have to hold at once:

$$\$1{,}000 + \$2{,}000q \geq p > \$2{,}000.$$

- **Step 2. Require the band to have room.** The upper bound must exceed the lower bound, so set $\$1{,}000 + \$2{,}000q > \$2{,}000$ and solve:

$$\$2{,}000q > \$1{,}000 \iff q > \tfrac{1}{2}.$$

- **Step 3. Read off the two cases.** If $q \leq \frac{1}{2}$ the band is empty: there is no price both high enough for the seller and low enough for the buyer, so the seller stops offering peaches and only lemons trade.
- **Inefficient equilibrium:** a peach could be traded under symmetric information at any price in $[\$2{,}000, \$3{,}000]$, yet under asymmetric information its market fails to exist.

Numerical illustration – $q = \frac{1}{3}$ vs $q = \frac{2}{3}$

- Suppose the buyer's prior is $q = \frac{1}{3}$:
  - Expected value: $\frac{1}{3}(3000) + \frac{2}{3}(1000) = \$1{,}666.67$.
  - Below seller's peach cutoff ($2,000) – only lemons trade.
- Suppose instead $q = \frac{2}{3}$:
  - Expected value: $\frac{2}{3}(3000) + \frac{1}{3}(1000) = \$2{,}333.33$.
  - Above $2,000 – a market for peaches can exist at price up to $2,333.33.
- The fraction of peaches in the market **determines** whether high-quality trade survives.
- Why the two cases split at $q = \frac{1}{2}$: expected value $\$1{,}000 + \$2{,}000q$ crosses the seller's peach floor $2,000 exactly there, so $q > \frac{1}{2}$ is precisely when the buyer's expected value clears the price the seller needs.

Reading the market failure

- The unraveling is subtle: nobody sets out to destroy the market. Every agent behaves optimally – and the market collapses anyway.
- **Two forces at work:**
  - Buyers cannot tell peaches from lemons, so they price the average quality.
  - At that price, peach owners refuse to sell – their reservation value exceeds the offered price.
- Result: only lemons trade; the peach market disappears.

Reading the market failure (cont.)

- **The efficiency loss.** Every peach-buyer match that fails represents surplus lost to the informational friction – a strictly Pareto-inferior outcome.
- In the next section (Section 5.3), we characterize the range of parameters for which this failure occurs, and in Section 5.6 we study the market's own responses (warranties, certificates, signaling).

5.2.3 Lemons in other markets

- Adverse selection applies in many domains:
  - **Health insurance**: insurers cannot observe individual health risks (Cutler & Reber, 1998).
  - **Credit markets**: credit rationing rather than risk-adjusted rates (Stiglitz & Weiss, 1981).
  - **IPOs**: underpricing as a response to asymmetric info (Rock, 1986).
  - **Online platforms**: reputation systems on eBay/Amazon (Resnick et al., 2006).
  - **Labor markets**: temp agencies signal worker quality (Autor, 2001).
  - **Peer-to-peer lending**: alternative screening models (Iyer et al., 2016).
- Even niche markets – online dating (Lee & Niederle, 2015) and **haunted houses** disclosure law – exhibit the same logic.

Section 5.3
Two Car Types – General Valuations

General valuations

- We now replace the specific dollar figures with general valuations, to see which features of the numbers drove the result.
- Buyer values peach at b_H, lemon at b_L, with $b_H > b_L$; seller values peach at s_H, lemon at s_L, with $s_H > s_L$.
- Maintained ranking:

$$b_H > s_H > b_L > s_L.$$

  - $b_i > s_i$: each type has positive gains from trade, so trade is efficient for both.
  - $s_H > b_L$: the buyer's lemon value sits below the seller's peach value, which is what makes a peach unprofitable once the buyer fears a lemon.
- Under symmetric information: peaches trade at $[s_H, b_H]$, lemons at $[s_L, b_L]$ – every car sells, as before.

The cutoff probability $\bar{q}$

- **Step 1. Buyer's rule at a high price** $p \in [s_H, b_H]$. He buys iff expected value exceeds the price:

$$q(b_H - p) + (1 - q)(b_L - p) > 0.$$

- **Step 2. Rearrange** into expected value versus price:

$$b_L + (b_H - b_L)q > p.$$

- **Step 3. Impose the seller's peach floor** $p > s_H$. For the buyer's rule to hold at some admissible price we need

$$b_L + (b_H - b_L)q > s_H \iff (b_H - b_L)q > s_H - b_L.$$

- The left side rises with q; the right side is fixed, so a large enough q tips the inequality.

The cutoff probability $\bar{q}$ (cont.)

- **Step 4. Solve for** q (dividing by $b_H - b_L > 0$ keeps the inequality direction):

$$q > \bar{q} \equiv \frac{s_H - b_L}{b_H - b_L}.$$

- Peaches trade in equilibrium iff $q > \bar{q}$: the proportion of peaches must clear this cutoff.
- Numeric check with the Section 5.2 numbers $s_H = \$2{,}000$, $b_L = \$1{,}000$, $b_H = \$3{,}000$:

$$\bar{q} = \frac{\$2{,}000 - \$1{,}000}{\$3{,}000 - \$1{,}000} = \frac{1{,}000}{2{,}000} = \tfrac{1}{2},$$

recovering the $q > \frac{1}{2}$ threshold found earlier.

5.3.1 Comparative statics of $\bar{q}$

- Read $\bar{q} = \dfrac{s_H - b_L}{b_H - b_L}$ as a ratio: a smaller $\bar{q}$ means trade survives for a wider range of q, so it is the market-friendliness we want low.
- **Value premium** $b_H - b_L$ (the denominator). A larger premium makes $\bar{q}$ fall, since a peach is worth much more than a lemon, so the buyer risks a purchase even when peaches are less common.
- **Preference divergence** $s_H - b_L$ (the numerator). A larger divergence makes $\bar{q}$ rise: the seller's peach value sits far above the buyer's lemon value, so the buyer needs peaches to be very likely before he will pay the price a peach seller requires.

Comparative statics of $\bar{q}$ (cont.)

- The three partial derivatives make each channel precise:

$$\frac{\partial \bar{q}}{\partial s_H} = \frac{1}{b_H - b_L} > 0, \qquad \frac{\partial \bar{q}}{\partial b_H} = -\frac{s_H - b_L}{(b_H - b_L)^2} < 0,$$

$$\frac{\partial \bar{q}}{\partial b_L} = \frac{s_H - b_H}{(b_H - b_L)^2} < 0 \quad \text{(negative since } s_H < b_H\text{)}.$$

- Reading each:
  - $s_H \uparrow$: a pricier peach for the seller raises $\bar{q}$, so trade needs more peaches.
  - $b_H \uparrow$: a more valuable peach for the buyer lowers $\bar{q}$, so trade is easier.
  - $b_L \uparrow$: valuing lemons more also lowers $\bar{q}$, since a mistaken lemon is less costly, so fewer peaches are needed for the buyer to risk it.

Section 5.4
Lemons with Continuous Qualities

Setup – Continuous qualities

- From here on q denotes the car's *quality*, now a continuous variable, rather than the probability of a peach used in Sections 5.2–5.3. We follow the book's notation; the meaning is set by context.
- Quality is uniformly distributed on $[0, Q]$ with $Q \in (1, 2)$:

$$q \sim U[0, Q].$$

- Buyer's valuation of a quality-q car is q; seller's valuation is $\frac{q}{Q}$.
- Gains from trade exist for every quality, since

$$q > \frac{q}{Q} \quad \text{for all } q > 0 \text{ (because } Q > 1).$$

- Examples: $q \sim U[0, 1.3]$ and $q \sim U[0, 1.5]$.

From two types to a continuum – What changes?

- With just **two types** (peaches and lemons), the market either functions or collapses – a discrete outcome.
- With a **continuum** of types, we can get *partial* unraveling: only the top of the distribution drops out.
- So the fraction of quality that trades varies continuously with the parameters – richer comparative statics.
- **The lesson survives.** Whichever setup we choose, the mechanism is the same: buyers price average quality, and the highest types find that price too low.
- **Empirical relevance.** Real-world quality distributions are continuous, not binary. Used-car quality, applicant ability, and project riskiness all vary smoothly.

5.4.1 Symmetric information

- **Step 1. Buyer's acceptance rule.** Observing quality q, the buyer accepts iff $q - p \geq 0$, i.e. the price does not exceed the value he sees.
- **Step 2. Seller's problem.** She sets the price to maximize her margin, subject to the buyer accepting:

$$\max_{p \geq 0} \; p - \frac{q}{Q} \quad \text{s.t.} \quad q - p \geq 0. \quad \text{(PC)}$$

- **Step 3. PC binds.** The seller raises the price until the buyer is just indifferent, so $p = q$; any lower price would leave money on the table.

5.4.1 Symmetric information (cont.)

- **Step 4. Substitute** $p = q$ into the objective:

$$\max_{p \geq 0} \; p - \frac{p}{Q} = p\left(\frac{Q-1}{Q}\right).$$

- **Step 5. Sign the derivative.** Since $\dfrac{Q-1}{Q} > 0$ when $Q > 1$, the objective rises in p, so the seller pushes p to its PC ceiling:

$$p^{SI} = q.$$

- Every quality trades, from q near 0 (lemons) up to q near Q (peaches).
- No market failure arises: when the buyer observes quality, each car is priced at its own value.

5.4.2 Asymmetric information

- **Step 1. Buyer forms an expectation.** Unable to observe quality, the buyer accepts iff the price does not exceed the *average* quality he expects:

$$p \leq E[q]. \quad \text{(PC)}$$

- **Step 2. Seller's problem.**

$$\max_{p \geq 0} \; p - \frac{q}{Q} \quad \text{s.t.} \quad p \leq E[q].$$

- **Step 3. PC binds.** As before the seller charges the most the buyer will accept, so $p = E[q]$. The key change: the price is tied to average quality, not the car's own quality.

5.4.2 Asymmetric information (cont.)

- **Step 4. When is a sale profitable?** With $p = E[q]$ the margin is $E[q] - \frac{q}{Q}$, which is nonnegative iff

$$q \leq Q \cdot E[q].$$

A car of quality above $Q\,E[q]$ costs the seller more than the average price, so she withholds it.

- **Step 5. Evaluate $E[q]$.** Since $q \sim U[0, Q]$, $E[q] = \frac{Q}{2}$, so the condition becomes

$$q \leq Q \cdot \frac{Q}{2} = \frac{Q^2}{2}.$$

- Only qualities $q \leq \frac{Q^2}{2}$ are offered; the top of the distribution, $\left(\frac{Q^2}{2}, Q\right]$, disappears.

Figure 5.2 – Lemons with continuous quality

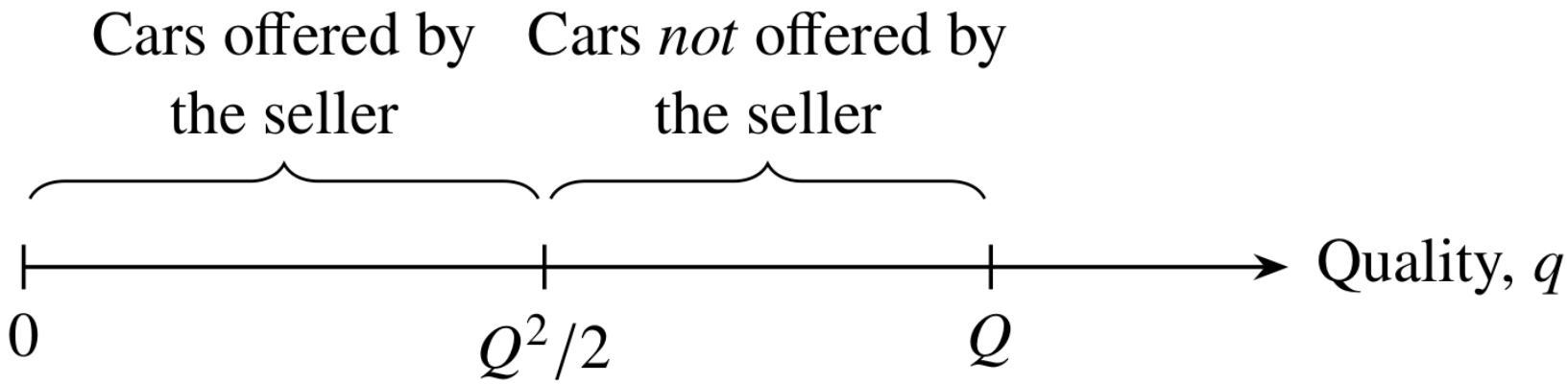


Figure 5.2. The lemons problem with continuous quality levels. Only qualities $q \leq Q^2/2$ are profitably offered; high-quality cars in $(Q^2/2, Q]$ disappear from the market.

- The flat ceiling at $p = E[q] = Q/2$ is all the seller can charge; any car whose own value q/Q sits above it is withheld, which is the shaded high-quality region cut from the market.

Example 5.1 – Highest qualities not sold

- **Step 1. Fix a value of Q.** Take $Q = \frac{3}{2}$, so quality is $q \sim U[0, \frac{3}{2}]$.
- **Step 2. Compute expected quality.**

$$E[q] = \frac{3/2}{2} = \frac{3}{4}, \qquad \text{so the PC ceiling is } p \leq \tfrac{3}{4}.$$

- **Step 3. Find the profitable cutoff** $q \leq Q \cdot E[q]$:

$$q \leq \frac{3}{2} \times \frac{3}{4} = \frac{9}{8} = 1.125.$$

- **Step 4. Read off the excluded cars.** Qualities in $(1.125, 1.5]$ are no longer offered – the top third of the range.
- The market for the highest-quality cars **disappears** under asymmetric information, even though each would trade under symmetric information.

5.4.3 Unraveling result – Round 2

- **Step 1. Anticipate the seller's withholding.** A rational buyer knows only $q \in [0, Q^2/2]$ will be offered, so he conditions his expectation on that truncated range.
- **Step 2. Update expected quality** on $[0, Q^2/2]$:

$$E\left[q \mid q \leq \tfrac{Q^2}{2}\right] = \frac{Q^2/2 - 0}{2} = \frac{Q^2}{4}.$$

- **Step 3. New ceiling.** The PC price falls to $p \leq \dfrac{Q^2}{4}$, lower than the first-round $Q/2$, because the buyer now expects a worse pool.

Unraveling result – Round 2 (cont.)

- **Step 4. New profitable cutoff.** With $p = \dfrac{Q^2}{4}$ the margin $\dfrac{Q^2}{4} - \dfrac{q}{Q}$ is nonnegative iff

$$q \leq \frac{Q^3}{4}.$$

- **Step 5. Confirm the range shrank.** The new bound is below the old one,

$$\frac{Q^3}{4} < \frac{Q^2}{2} \iff Q < 2,$$

  which holds since $Q \in (1, 2)$.

- So another slice of **higher-quality cars exits**: the buyer's rationality feeds back and tightens the market again.

Example 5.2 – Further restriction

- **Step 1. Continue Example 5.1** with $Q = \frac{3}{2}$, now applying the Round-2 cutoff $q \leq Q^3/4$.
- **Step 2. Evaluate the new bound:**

$$q \leq \frac{(3/2)^3}{4} = \frac{27}{32} \simeq 0.84.$$

- **Step 3. Compare to Round 1.** The offered range shrinks from $[0, 1.125]$ to $[0, 0.84]$; cars in $[0.84, 1.5]$ now disappear too.
- Each iteration restricts the quality range further, so the buyer's rationality *amplifies* the lemons problem rather than curing it.

Figure 5.3 – More iterations

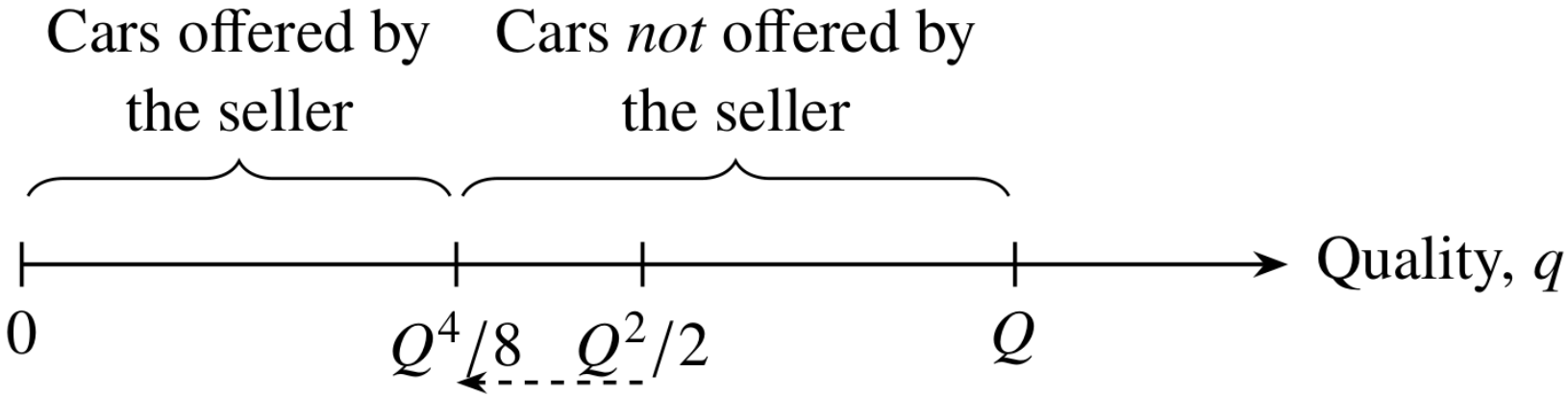


Figure 5.3. After more iterations, the upper bound of qualities offered shrinks to $Q^4/8$, then $Q^n/2^{n-1}$.

- Each round moves the right-hand edge of the traded interval leftward, since every update lowers the expected pool and hence the price the seller can obtain.

Round 3 and beyond

- **Step 1. Round-3 update.** Conditioning on $q \le Q^3/4$ and repeating the profit check gives

$$E\left[q \mid q \le \tfrac{Q^3}{4}\right] = \frac{Q^3}{8}, \qquad q \le \frac{Q^4}{8}.$$

- For $Q = \frac{3}{2}$ the bound is $\dfrac{(3/2)^4}{8} = \dfrac{81}{128} \simeq 0.63$, below the Round-2 value 0.84.
- **Step 2. Spot the pattern.** After n rounds the upper bound is

$$\frac{Q^n}{2^{n-1}} = 2\left(\frac{Q}{2}\right)^n.$$

- Since $Q < 2$ we have $\dfrac{Q}{2} < 1$, so the bound shrinks geometrically, driving more high-quality sellers out each round.

The unraveling result

- After n iterations, qualities offered shrink to $\left[0, \frac{Q^n}{2^{n-1}}\right]$.
- Taking the limit as $n \to \infty$:

$$\lim_{n\to\infty} 2\left(\frac{Q}{2}\right)^n = 0 \quad \text{since } \frac{Q}{2} < 1.$$

- In the limit only the **lowest quality** $q = 0$ is traded; markets for all positive qualities $q \in (0, Q]$ **collapse**.
- **Akerlof's main insight:** adverse selection together with the buyer's iterative inference unravels the market to nothing useful.
- Intuitively, pricing at the average always makes the best remaining car unprofitable, so the "best remaining" keeps falling until nothing is left. Solutions follow in Sections 5.6–5.9.

Section 5.5
Application to Labor Markets

Setup – Labor market

- The lemons logic is not special to cars. Here the **worker** sells labor and the **firm** buys it, with the roles of the car example carried over.
- The worker privately observes his productivity $\theta \sim U[0, 1]$; the firm cannot observe it, mirroring the buyer who could not see car quality.
- Many firms compete for workers (Bertrand-style competition), which will pin the wage to expected productivity.
- Question: what wages emerge in equilibrium under symmetric versus asymmetric information?

5.5.1 Symmetric information

- **Step 1. Firm observes θ.** It offers a wage $w = \theta$, and the two one-sided deviations fail:
  - $w > \theta$: the firm pays more than the worker produces, so it makes a loss.
  - $w < \theta$: a rival firm can offer slightly more and still profit, bidding the wage up.
- **Step 2. Competition pins the wage.** Bertrand-style undercutting drives w exactly to productivity:

$$w^{SI} = \theta.$$

- **Step 3. Who is employed.** Every worker from $\theta = 0$ to $\theta = 1$ works, each paid his own productivity.
- No market failure – analogous to symmetric-info car market in §5.4.1.

5.5.2 Asymmetric information – Round 1

- **Step 1. Firms price the average.** Unable to observe θ, each firm knows only $E[\theta] = \frac{1}{2}$ and offers that wage:
  - A higher wage would mean an expected loss, since pay exceeds expected productivity.
  - A lower wage would be undercut, as a rival can pay a bit more and still profit.
- **Step 2. Workers self-select.** A worker with $\theta > \frac{1}{2}$ faces $w = \frac{1}{2} < \theta$, so the wage is below his own productivity and he declines the job.
- **Step 3. Who stays.** Only workers with $\theta \leq \frac{1}{2}$ accept, so **high-productivity workers exit the market** – the lemons pattern in wages.

Iterations and unraveling

- **Step 1. Firm re-updates.** Knowing only $\theta \in [0, \frac{1}{2}]$ remain, it recomputes expected productivity:

$$E\left[\theta \mid \theta \leq \tfrac{1}{2}\right] = \frac{1/2 - 0}{2} = \frac{1}{4}.$$

- **Step 2. Wage falls, more exit.** The new wage $w = \frac{1}{4}$ drives out workers with $\theta \in (\frac{1}{4}, \frac{1}{2}]$.
- **Step 3. Repeat.** The next update gives

$$E\left[\theta \mid \theta \leq \tfrac{1}{4}\right] = \frac{1}{8}.$$

- Continuing, $w \to 0$ and only the lowest-productivity worker $\theta = 0$ remains: the **labor market unravels**, exactly as the car market did.

Section 5.6
Solutions to Adverse Selection

5.6.1 Warranties

- **Warranty:** the seller promises to repair or replace the product if it malfunctions within a set period.
- Why a warranty can signal quality: the expected repair cost it carries differs across types.
  - A high-quality seller faces low expected repair costs, so the promise is cheap for her to make.
  - A low-quality seller faces high expected repair costs, so the same promise is expensive.
- This cost wedge can sustain a separating equilibrium where only high-quality sellers offer warranties, letting buyers read a warranty as a mark of quality.
- Intuitively, the warranty works for the same reason education will in Spence's model: it is a signal only because it is cheaper for the better type.

Warranties (cont.)

- **Empirical evidence:**
  - Bond (1982): warranties raised average quality in used-pickup-truck market.
  - Tadelis (1999): warranties on electronics lower return rates, raise satisfaction.

Example 5.3 – Warranties in the lemons market

- **Step 1. Values and risk.** Buyer: peach \$4,000, lemon \$2,000; seller: peach \$3,000, lemon \$1,500. Breakdown probability 0.1 (peach), 0.8 (lemon); repair costs \$800.
- **Step 2. High-quality seller** (cost $0.1 \times \$800 = \$80$) beats her \$3,000 value:

$$\pi_H = \$4{,}000 - \$80 = \$3{,}920 > \$3{,}000. \checkmark$$

- **Step 3. Low-quality seller** (cost $0.8 \times \$800 = \$640$) falls short of her \$1,500 value:

$$\pi_L = \$2{,}000 - \$640 = \$1{,}360 < \$1{,}500. \times$$

- **Separating equilibrium**: only peaches carry warranties, so a warranty reveals a peach.

Example 5.4 – When warranties fail

- Same valuations as Example 5.3, but now the breakdown probabilities are generic, $p_H < p_L$, with repair cost c.
- Here p_H, p_L denote *breakdown probabilities*, not the car prices p_H, p_L of Section 5.2; the context sets the meaning.
- **Step 1. High-quality IC.** She offers a warranty iff it beats her \$3,000 reservation value:

$$\$4{,}000 - p_H c > \$3{,}000 \iff c < \frac{\$1{,}000}{p_H}.$$

- **Step 2. Low-quality IC.** She abstains iff the warranty fails to beat her \$1,500 reservation value:

$$\$2{,}000 - p_L c \leq \$1{,}500 \iff c \geq \frac{\$500}{p_L}.$$

Example 5.4 – Three regions of c

- Since $p_L > p_H$, the two cutoffs satisfy $\frac{\$500}{p_L} < \frac{\$1{,}000}{p_H}$, which orders the repair cost into three regions:
  1. $c < \frac{\$500}{p_L}$: repairs are cheap, so **both** types offer – a pooling (warranty) outcome, uninformative to buyers.
  2. $\frac{\$500}{p_L} \leq c < \frac{\$1{,}000}{p_H}$: **only the high-quality** seller offers – the separating outcome, where a warranty reveals a peach.
  3. $c \geq \frac{\$1{,}000}{p_H}$: repairs are dear, so **neither** offers – a pooling (no-warranty) outcome.
- Intuitively, a warranty separates types only when repair cost is neither so low that lemons can mimic nor so high that even peaches drop out.

Figure 5.4 – Three equilibria in the lemons market

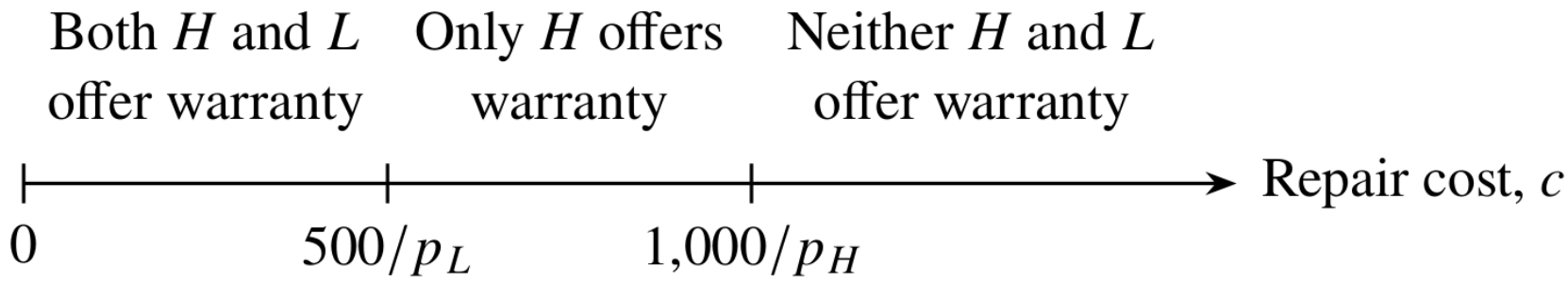


Figure 5.4. Three equilibria depending on the repair cost c: (i) pooling warranty for $c < \$500/p_L$, (ii) separating equilibrium for intermediate c, (iii) pooling no-warranty for high c.

- The two cutoffs mark where each type's warranty decision flips; only the middle band gives separation.

Warranties – General valuations

- Replace the dollar figures by buyer valuations v_H, v_L and seller valuations s_H, s_L, with $v_H > s_H$ and $v_L > s_L$.
- Here v_H, v_L play the role of the buyer values b_H, b_L of Section 5.3; we switch letters to match the book's warranty notation, and p_H, p_L remain breakdown probabilities.
- **High-quality IC:** she offers a warranty iff

$$c < \frac{v_H - s_H}{p_H}.$$

- **Low-quality IC:** she abstains iff

$$c \geq \frac{v_L - s_L}{p_L}.$$

- The same three regions as Example 5.4 arise (Figure 5.5).

Warranties – Comparative statics

- **Low-type surplus or risk.** A higher surplus $v_L - s_L$, or a lower breakdown probability p_L, raises the lower cutoff $\dfrac{v_L - s_L}{p_L}$ and expands the pooling (both-offer) region, since lemons are then more tempted to offer the warranty.
- **High-type surplus or risk.** A higher surplus $v_H - s_H$, or a lower p_H, raises the upper cutoff $\dfrac{v_H - s_H}{p_H}$ and widens the separating region.
- Both forces shrink the no-warranty pooling region from the right.
- Intuitively, the separating region survives only when the warranty is cheap enough for peaches but still too costly for lemons.

Figure 5.5 – General valuations

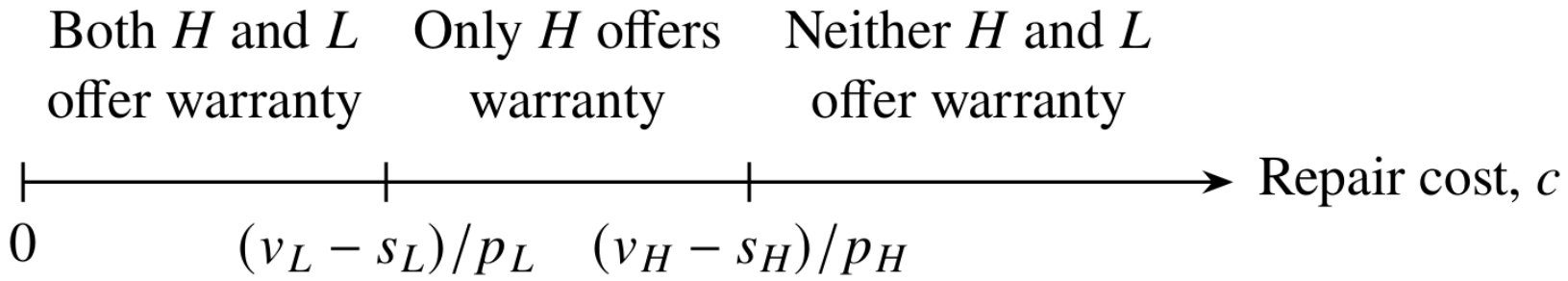


Figure 5.5. Three equilibrium regions in the warranty model with general valuations v_H, v_L, s_H, s_L. Cutoffs $\dfrac{v_L - s_L}{p_L}$ and $\dfrac{v_H - s_H}{p_H}$.

- Shifting either cutoff rightward (more surplus, less risk) trades area away from the no-warranty region on the right.

Warranties – Caveats

- Warranties only mitigate adverse selection if:
  - Warranty terms are **enforceable** – legally or reputationally.
  - Sellers' commitment to honor warranties is **credible**.
  - Cost of providing warranties is not prohibitive.
- Without enforcement a warranty becomes cheap talk: a low-quality seller can promise it at no real cost, collapsing the separation.
- In thin-margin industries even a high-quality seller may find the warranty unprofitable, giving a no-warranty pooling outcome.

Warranties – Caveats (cont.)

- **Policy implications:** consumer-protection laws, lemon laws, warranty standards.

5.6.2 Screening and quality certificates

- **Screening**: the uninformed party induces the informed party to reveal information, rather than waiting for a signal.
- **Quality certificates**: third-party verification (e.g., Carfax, AAA inspection) that attests to the car's condition.
- The certification cost is typically higher for a low-quality car, which needs more work to pass inspection – the same type-dependent cost wedge that drives warranties.
- Separating PBE: high-quality sellers certify, low-quality sellers do not, so a certificate reveals a peach.
- **Empirical evidence:** Jin & Kato (2006) – online field experiment showed third-party certification raised average quality and prices of cars sold.

Example 5.5 – Quality certificates

- **Step 1. Costs.** Same valuations as Example 5.3; certification costs \$500 for a peach and \$1,000 for a lemon.
- **Step 2. Peach seller.** Selling at \$4,000 net of the \$500 cost beats her \$3,000 reservation value:

$$\$4{,}000 - \$500 = \$3{,}500 > \$3{,}000. \checkmark$$

- **Step 3. Lemon seller.** Selling at \$2,000 net of the \$1,000 cost falls short of her \$1,500 reservation value:

$$\$2{,}000 - \$1{,}000 = \$1{,}000 < \$1{,}500. \times$$

- **Step 4. Conclusion.** Only peaches certify, so the certificate separates the types and buyers read it as a mark of quality.

5.6.3 Signaling – Spence's labor model

- **Signaling**: *informed* party takes a costly action to credibly reveal type.
- Many examples:
  - Limit pricing to deter entry (Bhattacharya, Kose-Williams).
  - Dividends signal future profitability (Milgrom-Roberts).
  - Bank capitalization signals financial health (Diamond-Dybvig).
  - Carbon disclosure signals sustainability (Bénabou-Tirole).
  - Education signals worker productivity (Spence, 1973).
- Spence's labor model is the canonical example – we develop it in Sections 5.7–5.9.

Reading the signaling logic

- **Who moves first?** Under screening, the uninformed party moves first with a menu. Under signaling, the informed party moves first with a costly action.
- **What makes education a credible signal?** Cost, not productivity. Spence's key insight: education is a signal precisely because it is cheaper for high-ability workers.
  - If education were free, everyone would get it – no separation.
  - If education were equally costly for both types, no separation either.
  - The wedge in the cost of the signal is what does the separating work.

Reading the signaling logic (cont.)

- **The single-crossing property.** Marginal willingness to invest in the signal decreases with type – exactly the condition that generates separating equilibria.

Screening vs. signaling – A comparison

- **Screening:** *uninformed* party moves first by offering a menu of options that induces self-selection.
  - Example: insurance company offers high-coverage + high-premium and low-coverage + low-premium plans.
- **Signaling:** *informed* party moves first by taking a costly action that credibly reveals his type.
  - Example: worker chooses education before applying.
- Both can produce **separating** equilibria where types self-identify.

Screening vs. signaling (cont.)

- Chapter 6 develops screening models in depth in the principal-agent context.

Spence's signaling game – Setup

- **Nature moves first:** the worker is high-productivity with probability $\frac{1}{3}$ and low-productivity with probability $\frac{2}{3}$, and privately observes his own type.
- **Worker moves (informed party):** he chooses education (E) or no education (NE) – the costly, observable signal.
- **Firm moves (uninformed party):** it sees only the education decision, not the type, and responds by hiring the worker as Manager (M) or Cashier (C).
- Strategy-profile notation $(\text{decision}^H, \text{decision}^L)$: e.g. (E^H, NE^L) means the high type educates and the low type does not.
- Because the informed party moves first, this is a *signaling* game, solved with Perfect Bayesian Equilibrium.

Figure 5.6 – Labor-market signaling game tree

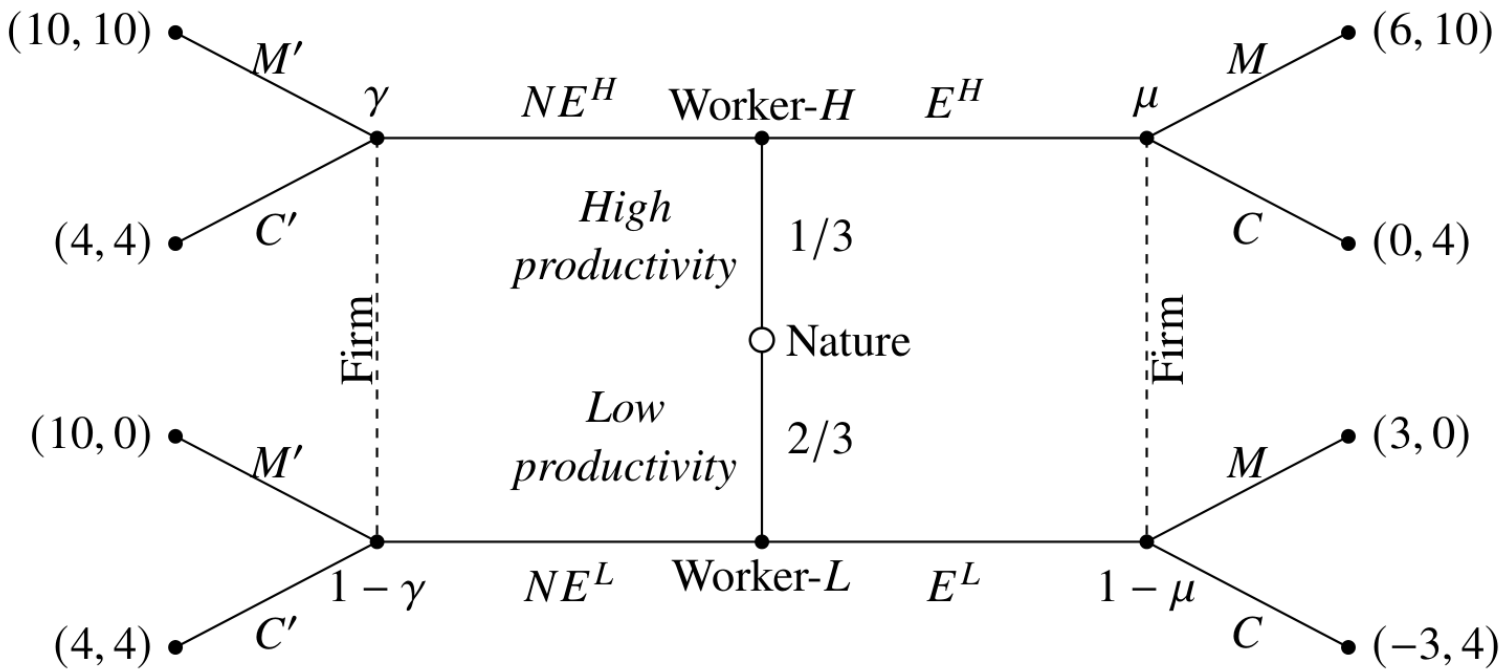


Figure 5.6. Labor-market signaling game. Nature picks type (top: high prob 1/3, bottom: low prob 2/3). Worker chooses E or NE; firm responds with M or C. First payoff: worker; second: firm. Dashed lines denote firm's information sets (it cannot tell type).

Payoffs in Figure 5.6

- **Cost of education = the payoff drop from educating.**
  - High type: as a manager his payoff falls $10 \to 6$, so the cost is $10 - 6 = \$4$.
  - Low type: as a manager his payoff falls $10 \to 3$, so the cost is $10 - 3 = \$7$.
- **Single-crossing:** education is cheaper for the high type ($\$4 < \$7$), the condition that lets education separate the types.
- **Firm's payoffs** depend on the match, not on education:
  - Manager: \$10 if the worker is high, \$0 if low – a mismatch is costly.
  - Cashier: \$4 regardless of type – a safe outside option.

Payoffs in Figure 5.6 (cont.)

- Education is *not* productivity-enhancing – isolates the signaling role.

Tool 5.1 – Finding PBEs in signaling games

1. **Specify** a strategy profile to test (e.g., (E^H, NE^L)) – one message for each type.
2. **Update beliefs** by Bayes' rule wherever the message occurs on the equilibrium path; leave off-path beliefs free.
3. **Find the firm's best response** given those beliefs, comparing its expected payoff from each hiring decision.

- We apply this in §5.7 (separating) and §5.8 (pooling).

Tool 5.1 – Finding PBEs (cont.)

4. **Find each worker type's optimal decision** given the firm's responses, checking whether either type gains by deviating.
5. **Compare** with the profile postulated in Step 1:
   1. If they coincide, the profile is sustained as a PBE (with the Step-2 beliefs and Step-3 responses).
   2. If they differ, the profile cannot be sustained.

- The deviation check in Step 4 is the heart of the method: a profile is an equilibrium only if *no* type wants to send a different signal.

Section 5.7
Signaling – Separating Equilibrium

Step 1 – Strategy profile (E^H, NE^L)

- Test the separating strategy where:
  - High-productivity worker acquires education (E^H).
  - Low-productivity worker does not (NE^L).
- Probabilities of acquiring education:

$$\alpha^H = 1, \quad \alpha^L = 0.$$

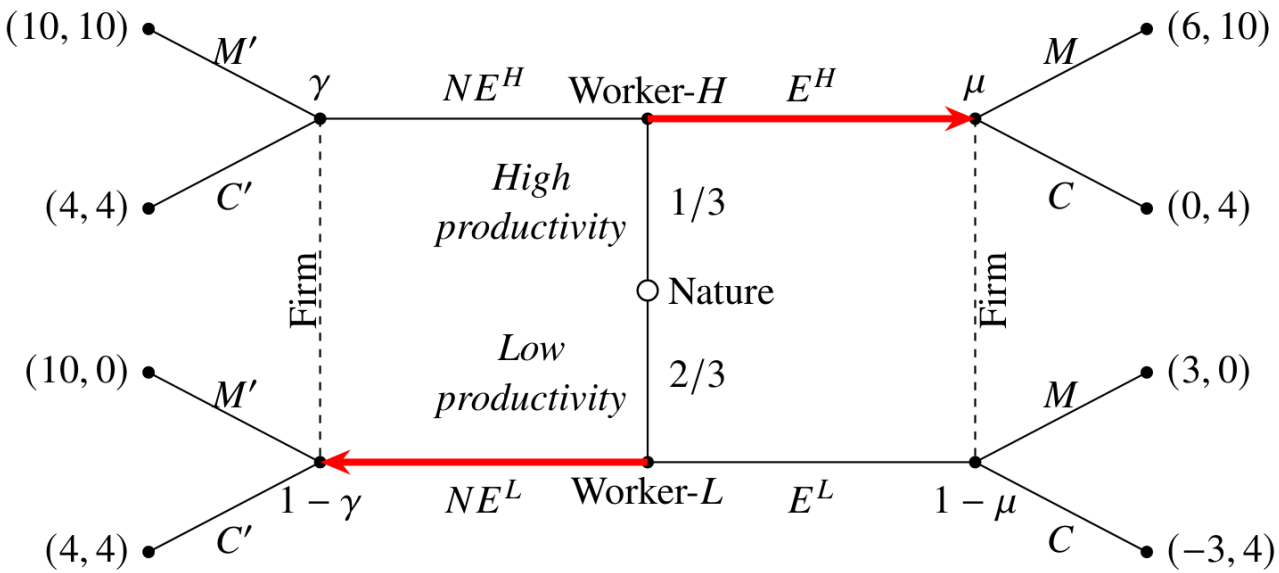


Figure 5.7. Separating strategy profile (E^H, NE^L).

Step 2 – Bayes' rule

- Update beliefs by Bayes' rule, each numerator being prior $\times$ strategy probability, where α^H, α^L are the chances each type educates.
- Belief after observing education (μ):

$$\mu = \frac{\frac{1}{3}\alpha^H}{\frac{1}{3}\alpha^H + \frac{2}{3}\alpha^L} = \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot 1 + \frac{2}{3}\cdot 0} = 1,$$

  since only the high type educates ($\alpha^H = 1, \alpha^L = 0$).
- Belief after observing no education (γ):

$$\gamma = \frac{\frac{1}{3}(1-\alpha^H)}{\frac{1}{3}(1-\alpha^H) + \frac{2}{3}(1-\alpha^L)} = \frac{0}{0 + \frac{2}{3}} = 0.$$

- So education signals a high type with certainty, and no education a low type with certainty – the profile fully separates.

Step 3 – Firm's optimal responses

- **After observing education** ($\mu = 1$): the firm is sure the worker is high-productivity.
  - Manager M: payoff 10; Cashier C: payoff 4.
  - Optimal: **hire as Manager** M, since $10 > 4$.
- **After observing no education** ($\gamma = 0$): the firm is sure the worker is low-productivity.
  - Manager M': payoff 0; Cashier C': payoff 4.
  - Optimal: **hire as Cashier** C', since $4 > 0$ (a low-type manager is a costly mismatch).

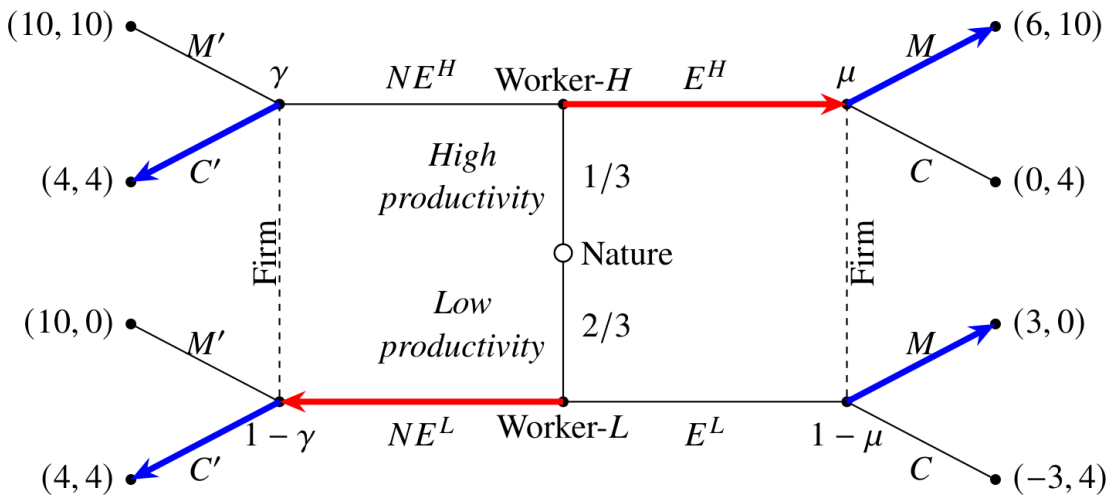


Figure 5.8. Firm's responses (M, C') in the separating PBE.

Step 4 – Worker's optimal decisions

- **High-productivity worker** (does educating beat deviating?):
  - E^H: hired as M, payoff 6. NE^H: hired as C', payoff 4.
  - Check: $6 > 4$, so he stays with E^H. ✓
- **Low-productivity worker** (does deviating to education pay?):
  - NE^L: hired as C', payoff 4. E^L: hired as M but pays the high cost, payoff $10 - 7 = 3$.
  - Check: $4 > 3$, so he stays with NE^L. ✓
- Neither type gains by deviating, so the **separating PBE is confirmed**: the low type's higher education cost is exactly what stops him mimicking.

Step 5 – Summary of separating PBE

- **Separating PBE:**
  - Worker strategy: (E^H, NE^L).
  - Firm response: (M, C').
  - Beliefs: $\mu = 1$ (after E), $\gamma = 0$ (after NE).
- Intuition: low-type's higher cost prevents him from mimicking the high-type. Education is informative.
- **Alternative profile** (NE^H, E^L) where only low-type educates: cannot be sustained – low-type wastes education cost while still being recognized as low.

Step 5 – Summary of separating PBE (cont.)

- Single-crossing condition (cost differential) is the key driver.

Example 5.6 – Separating PBE with general costs

- **Step 1. Generalize the low-type cost** to $c_L = 4 + c$ with $c \geq 0$, the low type's cost disadvantage. At $c = 3$ we recover Figure 5.6.
- **Step 2. Beliefs and firm responses unchanged**: $\mu = 1$, $\gamma = 0$, firm plays (M, C') as before.
- **Step 3. Low-type deviation check.** He stays at NE^L (payoff 4) rather than educating (manager payoff $6 - c$) iff

$$4 \geq 6 - c \iff c \geq 2.$$

- If $c \geq 2$ the separating PBE (E^H, NE^L) holds; if $c < 2$ the low type mimics and it **fails**.
- Numeric check: at $c = 3$, $6 - c = 3 < 4$, so he abstains – matching the baseline game.

Section 5.8
Signaling – Pooling Equilibrium

Step 1 – Strategy profile (NE^H, NE^L)

- Test the pooling strategy: **no** worker type acquires education.
- Probabilities: $\alpha^H = \alpha^L = 0$.

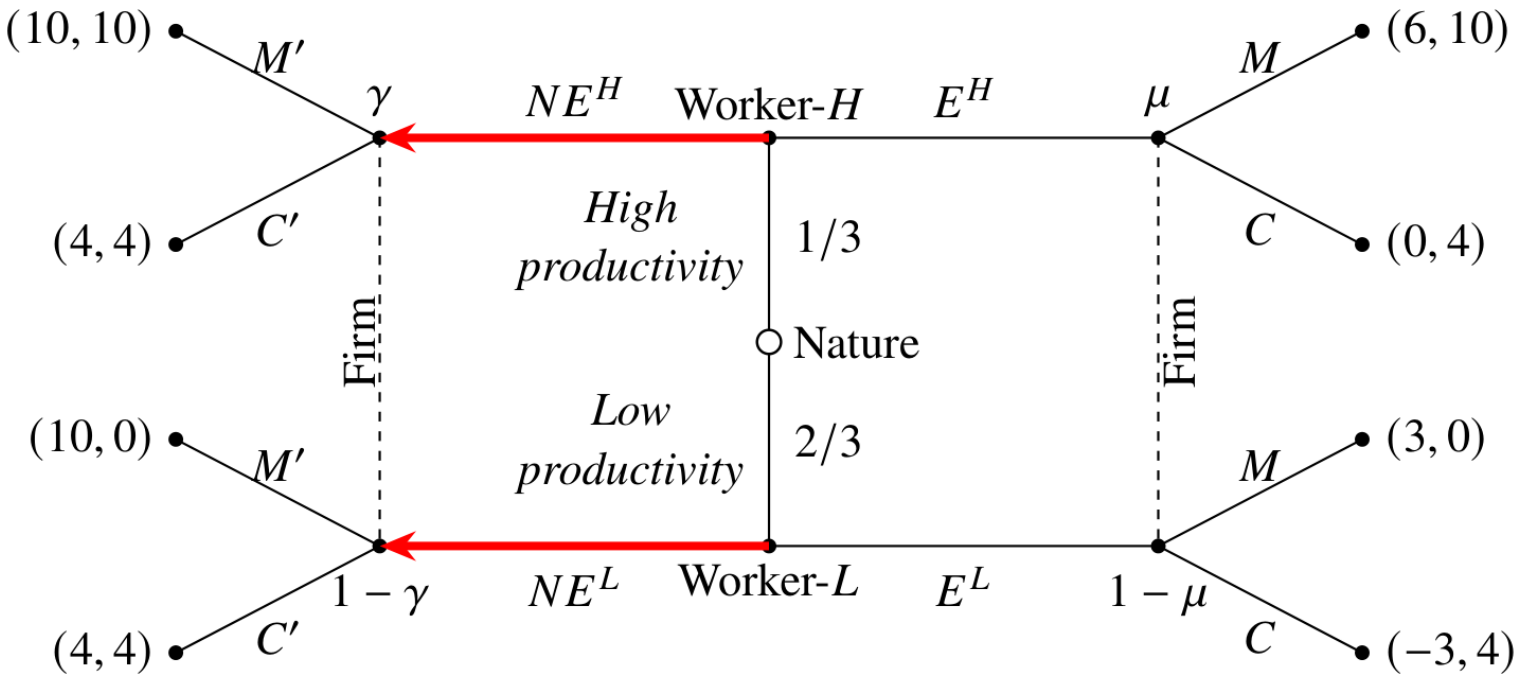


Figure 5.10. Pooling strategy profile (NE^H, NE^L).

Step 2 – Bayes' rule

- After no education (γ), with $\alpha^H = \alpha^L = 0$:

$$\gamma = \frac{\frac{1}{3}(1 - \alpha^H)}{\frac{1}{3} + \frac{2}{3}} = \frac{1}{3}.$$

  The posterior equals the prior, because both types pool on *NE*, so no education is **uninformative**.

- After observing education (μ): this never happens on the equilibrium path, so Bayes' rule is indeterminate:

$$\mu = \frac{\frac{1}{3} \cdot 0}{\frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 0} = \frac{0}{0} \quad \text{(undefined)}.$$

- These are **off-path beliefs**; Bayes' rule cannot pin them, so we leave $\mu \in [0, 1]$ free – the source of the multiplicity we refine later.

Step 3 – Firm's responses

- **After no education ($\gamma = 1/3$):**

$$E\pi(M') = 10 \cdot \tfrac{1}{3} + 0 \cdot \tfrac{2}{3} = \tfrac{10}{3} \simeq 3.33,$$

$$E\pi(C') = 4.$$

  Firm prefers C' (cashier).

- **After observing education (off-path):**

$$E\pi(M) = 10\mu, \quad E\pi(C) = 4.$$

  The firm hires as M iff $10\mu > 4$, i.e. $\mu > \frac{2}{5}$ – it needs a high-type likely enough to risk the manager slot.

- So the response after education splits into two cases:
  1. $\mu > \frac{2}{5}$: firm chooses M after education.
  2. $\mu \leq \frac{2}{5}$: firm chooses C after education.

Case 1: $\mu > 2/5$ – PBE fails

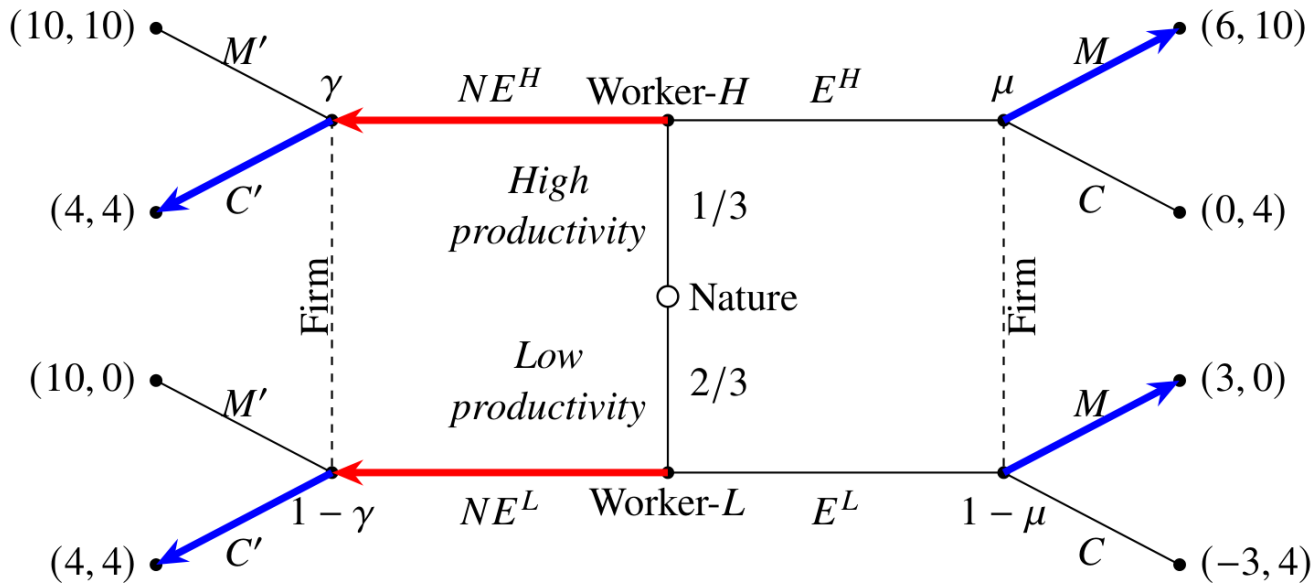


Figure 5.11. Pooling with off-path response (C', M).

- With $\mu > \frac{2}{5}$ an educated worker is hired as manager, so check the high type's deviation to E^H: NE^H gives cashier payoff 4, but E^H gives manager payoff $10 - 4 = 6$.
- Check: $6 > 4$ is a **profitable deviation**, so the pooling profile is **not sustained** when $\mu > \frac{2}{5}$.

Case 2: $\mu \leq 2/5$ – PBE sustained

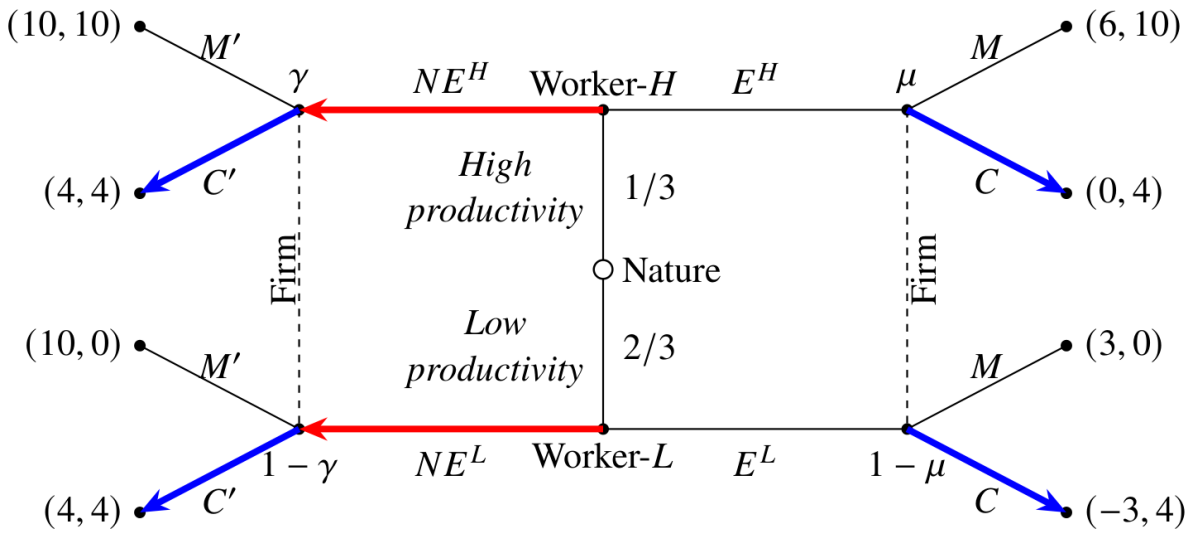


Figure 5.12. Pooling with off-path response (C', C).

- With $\mu \leq \frac{2}{5}$ an educated worker is still a cashier, so deviating to E never earns the manager wage:
  - **High type:** NE^H gives 4 vs. E^H's $4 - 4 = 0$; no deviation.
  - **Low type:** NE^L gives 4 vs. E^L's $4 - 7 = -3$; no deviation.
- Neither gains, so the **pooling PBE is sustained**.

Summary – Pooling (NE^H, NE^L)

- **PBE specification:**
  - Worker strategy: (NE^H, NE^L) – no one educates.
  - Firm response: (C', C) – cashier after both no education and education.
  - On-path beliefs: $\gamma = \frac{1}{3}$.
  - Off-path beliefs: $\mu \leq \frac{2}{5}$.
- Only sustained for sufficiently pessimistic off-path beliefs.
- These beliefs are *permissive* – the model alone doesn't restrict them. We refine them in Section 5.9.

5.8.1 Another pooling – (E^H, E^L)?

- **Step 1–2. Profile and beliefs.** Now both types educate, so education is on-path with $\mu = \frac{1}{3}$, while no education is off-path with $\gamma \in [0, 1]$ unrestricted.
- **Step 3. Firm's responses.**
  - After E: hire as cashier C, since $\mu = \frac{1}{3} < \frac{2}{5}$ makes $10\mu < 4$.
  - After NE (off-path): hire as manager M' iff $\gamma > \frac{2}{5}$, else cashier C'.
- In the postulated profile each type is educated and hired as a cashier: the high type earns $4 - 4 = 0$, the low type $4 - 7 = -3$.

Another pooling – (E^H, E^L)? (cont.)

- **Step 4. Deviation to *NE*, by belief region.**
  - If $\gamma > \frac{2}{5}$: a deviator is hired as manager, so the high type gets $10 > 0$ and the low type $3 > -3$; both deviate.
  - If $\gamma \leq \frac{2}{5}$: a deviator is hired as cashier, so the high type gets $4 > 0$ and the low type $4 > -3$; both still deviate.
- **Step 5. Conclusion.** In *every* belief region at least one type – in fact both – strictly gains by dropping education, so (E^H, E^L) **cannot be a PBE** for any γ.
- Intuitively, education here buys nothing (an educated worker is still a cashier), so paying for it is pure waste.

Section 5.9
Ruling Out PBEs with Insensible Beliefs

The puzzle of $\mu \leq 2/5$

- The pooling PBE (NE^H, NE^L) survives only under the off-path belief $\mu \leq \frac{2}{5}$.
- Read literally, this says that an unexpectedly educated worker is *more likely to be low-productivity* – a strange inference, since education is costlier for the low type.
- Ask who would ever deviate to education. Only the low type could benefit, yet even in the best case he loses:
  - Fooling the firm into M still yields $10 - 7 = 3 < 4$, his pooling payoff.
  - Being seen as a cashier yields $4 - 7 = -3$, far worse.
- So no low type would ever deviate to education – which makes a belief that blames the low type hard to defend.

The puzzle (cont.)

- For low-type, E^L is **equilibrium dominated**: his equilibrium payoff (4) exceeds any deviation payoff.
- Therefore the firm should believe an educated worker is the *high* type.

Equilibrium dominance refinement

- A deviation is **equilibrium dominated** for a type when its equilibrium payoff exceeds every payoff that deviation could bring. For the low type, E^L is dominated: 4 beats both 3 and -3.
- **Equilibrium dominance criterion:** off-path beliefs must put zero probability on a type for whom the observed deviation is equilibrium dominated.
- Applied here, the firm must not blame the low type for education, so $\mu = 1$ – which rules out the beliefs $\mu \leq \frac{2}{5}$ that the pooling PBE needed.
- Absent the refinement (NE^H, NE^L) is still a PBE; with it, only separation survives.

Equilibrium dominance refinement (cont.)

- Refined unique PBE: **separating** (E^H, NE^L) with firm response (M, C').
- **Stronger refinements** (for richer games):
  - Cho & Kreps (1987): **Intuitive Criterion**.
  - Banks & Sobel (1987): **Divinity Criterion**.
- These tools eliminate further implausible PBEs in games with more than two types or messages.

The intuitive criterion – One-line summary

- Cho and Kreps (1987): a belief-consistent PBE should also be robust to *sensible* out-of-equilibrium beliefs.
- **The rule.** Suppose only one type ever finds a particular off-path signal profitable. Then the uninformed party should assign probability one to that type when it sees the signal.
- **What this rules out.** Pooling PBEs sustained by pessimistic off-path beliefs that punish a deviator of *any* type equally – when in fact only one type could gain from deviating.

The intuitive criterion (cont.)

- **What survives.** Separating PBEs, and pooling PBEs whose off-path beliefs are consistent with the players' incentives to send that signal.
- **Why we care.** Without a refinement, adverse-selection models produce a large multiplicity of equilibria. The intuitive criterion selects the ones that survive a mild rationality test on beliefs.

Section 5.10
Summary

Summary – The lemons problem

- **Two car types**: under SI both peaches and lemons trade; under AI, peaches disappear unless $q > \frac{1}{2}$.
- **General valuations**: trade requires $q > \bar{q} = \dfrac{s_H - b_L}{b_H - b_L}$.
- **Continuous qualities**: iterative buyer rationality leads to **full unraveling**; only $q = 0$ trades in the limit.
- **Labor markets**: high-productivity workers exit $\Rightarrow$ wages converge to 0; same unraveling logic.
- Asymmetric information $\Rightarrow$ market failure.

Summary – Solutions

- **Warranties**: high-quality sellers offer; low-quality don't (separating).
  - Range of c determines pooling-warranty, separating, or pooling-no-warranty equilibria.
- **Quality certificates**: third-party verification with type-dependent costs $\Rightarrow$ separating.
- **Signaling** (Spence): informed party takes a costly action.
  - Separating PBE (E^H, NE^L) sustained by single-crossing in costs.
  - Pooling PBE (NE^H, NE^L) requires off-path beliefs $\mu \leq \frac{2}{5}$.
  - Pooling (E^H, E^L) not sustainable.

Summary – Refinement and equilibrium concepts

- **Perfect Bayesian Equilibrium** (PBE):
  - Strategies are sequentially rational given beliefs.
  - Beliefs follow Bayes' rule on the equilibrium path.
  - Beliefs off path are unrestricted – the source of multiplicity.
- **Refinements** eliminate implausible PBEs:
  - Equilibrium dominance: rule out beliefs assigning probability to types whose deviation is equilibrium-dominated.
  - Cho-Kreps Intuitive Criterion, Banks-Sobel Divinity Criterion for richer games.
- In our example: refinement leaves separating PBE (E^H, NE^L) as unique.

The single-crossing property – A general lesson

- Spence's signaling works because of **single-crossing**: high-types' marginal cost of education is **lower** than low-types'.
- Without single-crossing, separating becomes impossible – low-types can always mimic high-types.
- Single-crossing appears throughout contract theory:
  - Insurance: low-risk types more willing to accept high deductibles.
  - Labor: high-ability workers find effort cheaper.
  - Auctions: high-value bidders' willingness to pay rises faster.

Single-crossing (cont.)

- Chapter 6 makes single-crossing the technical backbone of the asymmetric-information principal-agent problem.

Empirical evidence on signaling

- **Education premium:** Lange (2007) – employer learning about productivity over time, consistent with signaling.
- **Veterans' education:** Tyler, Murnane & Willett (2000) – GED passage as a signal of productivity.
- **Firm signaling:** Spiegel & Spulber (1997) – dividends as signals of future earnings.

Empirical evidence (cont.)

- **Online platforms:** Cabral & Hortaçsu (2010) – eBay reputation scores function as signals of seller quality.
- **Environmental signaling:** firms voluntarily disclosing emissions to signal sustainability (Lyon & Maxwell, 2008).
- Across domains: costly, type-dependent signals do shift equilibrium outcomes.

Looking ahead – Chapter 6

- Chapter 5 introduced adverse selection in markets.
- Chapter 6 studies the **general principal-agent problem** combining moral hazard and adverse selection:
  - Profit function $\pi(e) = \log(e) - w$; utility $u(w, e) = w - \theta e^2$.
  - Types $\theta_L < \theta_H$.
  - Symmetric info: $e_i^{SI} = (1/2\theta_i)^{1/2}$.
  - Asymmetric info: PC_H and IC_L bind.
  - **No distortion at the top**: $e_H^{AI} = e_H^{SI}$.
  - **Downward distortion at the bottom**: $e_L^{AI} < e_L^{SI}$.
  - **Information rent** for high type: $u_L^{AI} = (\theta_H - \theta_L)c(e_H) > 0$.
- Plus extensions: single-crossing, screening with product qualities, non-linear pricing.

Adverse selection vs. moral hazard

- **Moral hazard** (Chs. 2–4):
  - Hidden *action after* contracting.
  - Optimal contract: wage spread provides effort incentives.
  - Trade-off: risk-sharing vs. incentives.
- **Adverse selection** (Ch. 5):
  - Hidden *type before* contracting.
  - Optimal market institutions: separate types via signaling / screening.
  - Trade-off: information rents vs. efficient allocation.
- Chapter 6 **integrates** both: agent's type affects effort cost; principal designs contracts that screen *and* provide incentives.

Welfare implications

- Adverse selection generates **deadweight loss**: efficient trades fail to occur.
- In the lemons model:
  - Total welfare loss = sum of surplus from non-traded high-quality units.
  - Buyer also loses access to peaches; seller loses sale of peaches.
- In labor markets: high-productivity workers exit, output is lower than under SI.

Welfare implications (cont.)

- **Solutions reduce DWL but rarely eliminate it:**
  - Signaling: costly action consumes resources.
  - Certification: third-party fees.
  - Warranties: insurance premium absorbed in price.
- **Second-best vs first-best**: signaling/screening get us closer to first-best, but never exactly there.

Policy implications

- **Disclosure regulations:** mandatory reporting (food labels, mortgage disclosure, used-car odometer disclosure) reduce hidden information.
- **Mandatory insurance:** pooling everyone into the market (e.g., universal health insurance) prevents adverse-selection unraveling.
- **Lemon laws:** legal warranties on used cars.

Policy implications (cont.)

- **Credit-rating agencies:** standardized credit scores reduce information gap.
- **Reputation systems:** platform-level mechanisms (eBay, Amazon, Uber) substitute for formal certification.
- Each policy reduces information asymmetry but introduces its own costs (compliance, mandates, false reputation).

Common student mistakes to avoid (Chapter 5)

- **Confusing lemons with moral hazard.** The seller of a lemon is not *doing* anything hidden – he already *is* a lemon. The friction is on the type, not the action.
- **Forgetting who moves first.** Under signaling the *informed* party moves first; under screening the *uninformed* party moves first. The order matters for the equilibrium concept.
- **Reading "separating" as "good."** A separating PBE can still be inefficient (education has no productive value in the Spence model), but it is the least-bad among achievable equilibria.

Common student mistakes (Chapter 5) (cont.)

- **Confusing the intuitive criterion with a solution concept.** It is a refinement of PBE, not a new equilibrium concept.
- **Assuming single-crossing.** It is a property, not a default. Without it, screening menus and separating signals need not exist.

Key takeaways from Chapter 5

1. **Adverse selection** arises with hidden types; the lemons problem is its canonical illustration.
2. The cutoff $\bar{q} = \dfrac{s_H - b_L}{b_H - b_L}$ determines whether a market for high-quality goods exists.
3. With **continuous quality**, iterative rationality drives the market to **unravel** to the lowest type.
4. **Warranties**, **certification**, and **signaling** mitigate adverse selection by allowing types to separate.
5. Spence's signaling model: separating PBE requires **single-crossing** in signal costs; pooling PBE depends on off-path beliefs.
6. **Equilibrium dominance** refines away implausible PBEs – leaving the separating equilibrium as the natural prediction.

End of Chapter 5

Chapter 6

Screening

Instructor Slides

Contract Theory: An Introduction with Applications

Ana Espinola-Arredondo, Felix Muñoz-Garcia, and Pak-Sing Choi

Outline of Chapter 6

- This chapter develops the **general principal-agent problem** combining *hidden type* and the design of *menus*.
- Sections:
  - 6.1 Introduction.
  - 6.2 Model: log-linear, quadratic cost.
  - 6.3 General utility and cost functions; single-crossing.
  - 6.4 Applications: product quality screening, non-linear pricing.
  - 6.5 Summary. 6.6 Exercises. Appendix 6.A.

What you will learn

- The **5-step roadmap** for the principal's problem under symmetric and asymmetric information.
- Two key adverse-selection results:
  - **No distortion at the top**: efficient type's effort is undistorted.
  - **Downward distortion at the bottom**: inefficient type exerts less effort than first-best.
- The **information rent** formula: $u_L^{AI} = (\theta_H - \theta_L)c(e_H) > 0$.

What you will learn (cont.)

- The **single-crossing property** – the key technical condition.
- Applications: product-quality screening and non-linear pricing (uniform vs single tariff vs menu).

Section 6.1
Introduction

The principal-agent problem – A unified view

- Imagine a firm (**principal**) and a worker (**agent**).
- Firm's profit rises with effort but at a diminishing rate (concave), and falls with the wage paid; a concrete form is $\pi(e) = \log(e) - w$.
- Agent's utility rises with the wage but falls with effort, and effort is more painful the larger his cost parameter θ; a concrete form is $u(w, e) = w - \theta e^2$.
- **Intuitively,** the worker and the firm value effort in opposite ways – effort is costly to the worker but valuable to the firm – so the contract must reconcile the two sides.

The unified view (cont.)

- **Hidden information**: principal does not observe θ.
- **Hidden action** is set aside here – effort is observable once contracted. Chapter 6 focuses purely on adverse selection within a principal-agent setting.
- Solution: the principal offers a **menu of contracts** to induce self-selection.
- **Intuitively,** since the principal cannot see the type, she lets the worker reveal it through the contract he picks, rather than asking him to report it.

Where Chapter 6 sits among contract problems

- Chapters 2–4 gave us **moral hazard** – the friction is on the *action*.
- Chapter 5 gave us **adverse selection with signaling and lemons** – the friction is on the *type*, and the market partly unravels.
- Chapter 6 is the third branch: adverse selection with the *principal* designing a menu of contracts to **screen** the agent's type.
  - The principal moves first, offering a menu.
  - Each agent type selects the option intended for him.
  - Truth-telling is enforced by the incentive compatibility constraint on *types*, not on actions.

Where Chapter 6 sits (cont.)

- **Two big lessons ahead.** No distortion at the top (the efficient type gets the first-best allocation) and downward distortion at the bottom (the inefficient type's allocation is bent to hold rents in check).

Roadmap for Chapter 6

- **Sections 6.2–6.3** – core theory:
  - Log-linear/quadratic model (6.2): closed-form solutions.
  - General utility/cost (6.3): single-crossing property.
  - Both sections solve the SI then AI cases, using the 5-step roadmap.
- **Section 6.4** – applications to monopoly:
  - Product-quality screening.
  - Non-linear pricing (two-part tariffs).
- Throughout: same key results – **no distortion at top**, **downward distortion at bottom**, **information rents**.

Section 6.2
Model

Principal and agent

- **Principal's profit:**

$$\pi(e) = \log(e) - w.$$

  - Concave in e since $\partial \log(e)/\partial e = 1/e > 0$ and $\partial^2 \log(e)/\partial e^2 = -1/e^2 < 0$.

- **Agent's utility:**

$$u(w, e) = w - \theta e^2.$$

  - Marginal cost of effort: $2\theta e$, increasing in e.
  - Parameter θ scales the cost: $\theta_H > \theta_L > 0$.
  - High-cost type (θ_H) finds effort more painful.

6.2.1 Symmetric information – Step 1

- Firm observes θ; solves:

$$\max_{w,e} \ \log(e) - w$$

$$\text{s.t.} \quad w - \theta e^2 \geq 0. \quad \text{(PC)}$$

- PC: the worker signs only if his utility weakly exceeds his reservation level (normalized to 0).
- Two choice variables (w and e); one constraint.
- Five-step approach:
  1. Formulate problem.
  2. Show PC binds.
  3. Substitute binding PC.
  4. Solve for effort.
  5. Compute wage.

Steps 2–3 – PC binds, simplify

- **Step 2 (PC binds):** otherwise the firm could lower wage and still keep worker. So:
$$w = \theta e^2.$$
- **Step 3 (substitute):** insert $w = \theta e^2$ into objective:
$$\max_{e \geq 0} \ \log(e) - \theta e^2.$$
- One choice variable left: effort e.

Step 4 – Effort under SI

- **Step 4 (FOC).** Differentiate $\log(e) - \theta e^2$ and set it to zero,

$$\frac{1}{e} - 2\theta e = 0,$$

  so the marginal benefit of effort $1/e$ equals its marginal cost $2\theta e$ (the objective is globally concave, so this point is the maximum).
- Solving gives $e^2 = \frac{1}{2\theta}$, hence

$$e^{SI} = \left(\frac{1}{2\theta}\right)^{1/2}.$$

- Decreasing in θ, since $\frac{\partial e^{SI}}{\partial \theta} = -(2\theta)^{-3/2} < 0$: the high-cost type exerts less effort.

Steps 4–5 – Effort by type and wages

- Evaluating by type (recall $\theta_H > \theta_L$):

$$e_H^{SI} = \left(\frac{1}{2\theta_H}\right)^{1/2} < \left(\frac{1}{2\theta_L}\right)^{1/2} = e_L^{SI}.$$

- **Step 5 (wages).** Substitute each e_i^{SI} into the binding PC, $w = \theta e^2$ (the wage just covers the effort cost):

$$w_H^{SI} = w_L^{SI} = \theta_i \cdot \frac{1}{2\theta_i} = \$\frac{1}{2}.$$

- Both types earn the same wage: the higher cost per unit of effort for θ_H is offset by the lower effort the firm asks of him.

Intuition for the SI equilibrium

- Why are wages equal across types?
  - Both PCs bind, so each type gets exactly her reservation utility ($\bar{u} = 0$).
  - Wage = compensation for cost of effort: $w_i = \theta_i e_i^2$.
  - Since high-cost type exerts **less** effort, the products $\theta_H \cdot e_H^2$ and $\theta_L \cdot e_L^2$ coincide.
- Effort levels differ because the principal balances the marginal benefit (concave $\log e$) against the marginal cost ($2\theta_i e$).
- Higher θ_i pushes the optimum to a lower effort.

Intuition for the SI equilibrium (cont.)

- This is the *first-best* benchmark: the firm extracts all surplus, so each type is held to his reservation utility and earns no information rent.
- **Economically,** with the type observed there is nothing to hide, so the contract is efficient and the only question is how much effort each type should supply.
- Next we contrast this with asymmetric information, where hiding the type forces the firm to concede a rent.

Example 6.1 – Numerical SI

- Let $\theta_H = 0.15$, $\theta_L = 0.1$.
- Effort levels (plug into $e_i^{SI} = (1/2\theta_i)^{1/2}$):

$$\begin{aligned} e_H^{SI} &= \left(\frac{1}{2 \times 0.15}\right)^{1/2} = \sqrt{\tfrac{10}{3}} \simeq 1.83, \\ e_L^{SI} &= \left(\frac{1}{2 \times 0.1}\right)^{1/2} = \sqrt{5} \simeq 2.24. \end{aligned}$$

- The low-cost worker is asked for more effort, $e_L^{SI} > e_H^{SI}$.

Example 6.1 – Numerical SI (cont.)

- Wages are equal across types, $w_H^{SI} = w_L^{SI} = \$\frac{1}{2}$.
- **Sanity check (PC binds).** The high-cost worker's utility is exactly zero:

$$u_H^{SI} = \frac{1}{2} - 0.15 \cdot \frac{10}{3} = \frac{1}{2} - \frac{1}{2} = 0, \ \checkmark$$

  and the same holds for the low-cost worker.
- Firm profit per type: $\log(1.83) - 0.5 \simeq 0.10$ and $\log(2.24) - 0.5 \simeq 0.30$, so the low-cost worker is the more worthwhile hire.

6.2.2 Asymmetric Information

The principal's problem under AI

- Firm cannot observe θ; knows $\gamma = \Pr(\theta_H)$.
- Designs a **menu** of contracts (w_H, e_H) and (w_L, e_L).
- Problem (expected profit, weighting each type's contract by its probability):

$$\max \ \gamma[\log(e_H) - w_H] + (1-\gamma)[\log(e_L) - w_L].$$

- Because the type is hidden, the same menu must work for both, which adds the incentive constraints below.

The principal's problem under AI (cont.)

- Subject to **four constraints**:
  - PC_H: $w_H - \theta_H e_H^2 \geq 0$.
  - PC_L: $w_L - \theta_L e_L^2 \geq 0$.
  - IC_H: $w_H - \theta_H e_H^2 \geq w_L - \theta_H e_L^2$.
  - IC_L: $w_L - \theta_L e_L^2 \geq w_H - \theta_L e_H^2$.

Why menus instead of single contracts?

- Under AI, principal cannot just ask the worker his type – worker would always claim to be the high-cost type to demand more wage for less effort.
- Solution: design **two contracts** (w_H, e_H) and (w_L, e_L) such that each type **voluntarily** picks the one designed for her.
- This requires:
  - PCs so each type participates.
  - ICs so each type self-selects truthfully.

Why menus instead of single contracts? (cont.)

- ICs are the genuinely new constraints relative to SI.
- Designing a menu means letting the worker reveal his type through the contract he chooses.
- **Intuitively,** the firm cannot ask "what is your type?" and trust the answer, so it lets the choice of contract do the talking – each type picks the option built for him.

The 4 constraints explained

- PC_H: high-type accepts the high-type contract.
- PC_L: low-type accepts the low-type contract.
- IC_H: high-type prefers her own contract to the low-type's.

The 4 constraints explained (cont.)

- IC_L: low-type prefers her own contract to the high-type's.
- PCs ensure voluntary participation; ICs ensure self-selection.
- **Key insight (Step 2):** two of the four constraints bind in equilibrium – specifically, PC_H and IC_L.
- The other two (PC_L and IC_H) are slack.

Step 2a – PC_L is slack

- Start from IC_L: $w_L - \theta_L e_L^2 \geq w_H - \theta_L e_H^2$.
- Since $\theta_L < \theta_H$:

$$w_H - \theta_L e_H^2 > w_H - \theta_H e_H^2 \geq 0$$

  (last inequality is PC_H).
- Combining: $w_L - \theta_L e_L^2 > 0$.
- This is PC_L with strict inequality $\Rightarrow$ PC_L **is slack**.
- **Interpretation:** if the low-cost agent prefers her own contract over the high-cost contract (IC_L), and the high-cost agent participates (PC_H), then the low-cost agent participates as well.

Step 2b – IC_L binds

- If IC_L were slack, the principal could lower w_L and still keep the low type choosing his own contract, raising profit. So it holds with equality:

$$w_L - \theta_L e_L^2 = w_H - \theta_L e_H^2,$$

  which rearranges to

$$w_L = w_H + \theta_L (e_L^2 - e_H^2).$$

- This ties the low type's wage to the high type's, plus the effort-cost difference he saves by taking his own contract.

Steps 2c–2d – IC_H slack, PC_H binds

- **Step 2c (IC_H slack).** Substituting the binding IC_L into IC_H and cancelling w_H leaves

$$(\theta_H - \theta_L)(e_L^2 - e_H^2) \geq 0,$$

  which holds strictly once $e_L > e_H$ (verified below), so the high type does not want the low type's contract.
- **Step 2d (PC_H binds).** Otherwise the principal could lower w_H and still keep the high type, so

$$w_H = \theta_H e_H^2.$$

Step 3 – Wages from the binding constraints

- From binding PC_H, $w_H = \theta_H e_H^2$. Substituting this into binding IC_L gives the low type's wage:

$$\begin{aligned} w_H &= \theta_H e_H^2, \\ w_L &= \theta_H e_H^2 + \theta_L (e_L^2 - e_H^2). \end{aligned}$$

- The first term, $\theta_H e_H^2$, is what the high type is paid; the extra $\theta_L(e_L^2 - e_H^2)$ is the low type's information rent expressed in wage terms.

Step 3 – Substitute into the objective

- Insert w_H and w_L into expected profit:

$$\max_{e_L, e_H} \ \gamma[\log(e_H) - \theta_H e_H^2] + (1-\gamma)[\log(e_L) - \theta_H e_H^2 - \theta_L(e_L^2 - e_H^2)].$$

- Only two choice variables remain, e_H and e_L (the wages are already pinned down).
- We ignore PC_L and IC_H for now, then verify ex post that they hold.

Step 4 – FOC for e_L: No distortion at the top

- Differentiate w.r.t. e_L:

$$(1-\gamma)\left(\frac{1}{e_L}-2\theta_L e_L\right)=0.$$

- Since $1-\gamma>0$, the bracket must vanish, giving exactly the SI condition:

$$e_L^{AI}=\left(\frac{1}{2\theta_L}\right)^{1/2}=e_L^{SI}.$$

- **No distortion at the top:** the efficient type's effort under AI coincides with SI.
- Intuition: e_L enters no incentive constraint, so distorting it would just lower profit.

Step 4 – FOC for e_H

- Differentiate w.r.t. e_H; the γ-term is the direct benefit, the second term the extra rent cost of raising e_H:

$$\gamma\left(\frac{1}{e_H}-2\theta_H e_H\right)-2e_H(1-\gamma)(\theta_H-\theta_L)=0.$$

- Solving for e_H:

$$e_H^{AI}=\left(\frac{\gamma}{2[\theta_H-(1-\gamma)\theta_L]}\right)^{1/2}.$$

Step 4 – Downward distortion at the bottom

- **Downward distortion:** $e_H^{AI} < e_H^{SI}$ because

$$\frac{\gamma}{\theta_H - (1-\gamma)\theta_L} < \frac{1}{\theta_H} \iff \theta_H > \theta_L, \ \checkmark$$

  which holds by assumption.

- **Intuitively,** the firm asks the high-cost type for less effort than first-best to make his contract less tempting to the low-cost type; this relaxes the binding IC_L and shrinks the rent the firm must pay.

Step 5 – Equilibrium wages

- Insert the equilibrium efforts into the binding constraints. From $w_H = \theta_H e_H^2$:

$$w_H^{AI} = \frac{\gamma\theta_H}{2[\theta_H - (1-\gamma)\theta_L]} < \frac{1}{2} = w_H^{SI},$$

lower than under SI, since $\theta_H > \theta_L$.

- From $w_L = \theta_H e_H^2 + \theta_L(e_L^2 - e_H^2)$:

$$w_L^{AI} = \frac{(1+\gamma)\theta_H - \theta_L}{2[\theta_H - (1-\gamma)\theta_L]} > \frac{1}{2} = w_L^{SI},$$

higher than under SI – this gap is the low type's information rent.

Step 5 – Reading the equilibrium wages

- **Key results under AI:**
  - High-cost type: lower wage, lower effort (utility still 0).
  - Low-cost type: **higher** wage, same effort (a positive rent).
- The wage spread widens from 0 under SI to $w_L^{AI} > w_H^{AI}$ under AI.
- **Economically,** hidden information flips the wage ranking: the efficient worker, paid the same as the inefficient one under SI, now earns strictly more.

Intuition – No distortion at the top

- The low-cost (efficient) type's effort is chosen to maximize joint surplus, exactly as under SI.
  - Distorting e_L would shrink the pie without relaxing any IC.
  - All the inefficiency is absorbed in the high-type contract instead.
- **Intuitively,** no rent is attached to the top type's effort, so the firm has no reason to bend it away from first-best.

Intuition – Downward distortion at the bottom

- The high-cost type's effort is reduced below first-best.
  - Lowering e_H makes the high-cost contract less attractive to the low type (less effort, but also less wage).
  - This relaxes IC_L, so the principal can lower w_L and pay a smaller information rent.
- **Trade-off:** efficiency loss (lower e_H) versus rent saved on the low type; the firm sets e_H where the two balance at the margin.

The two-distortion rule – Concrete numbers

- Take $\theta_H = 2$, $\theta_L = 1$, $\gamma = \frac{1}{2}$ (equal type probabilities).
- **Top type** (efficient, low cost):
  $e_L^{AI} = e_L^{SI} = \left(\frac{1}{2\theta_L}\right)^{1/2} = \frac{1}{\sqrt{2}} \simeq 0.71$ – **no distortion at the top**.
- **Bottom type** (inefficient, high cost): first-best is
  $e_H^{SI} = \left(\frac{1}{2\theta_H}\right)^{1/2} = \frac{1}{2}$, while

$$e_H^{AI} = \left(\frac{\gamma}{2[\theta_H - (1-\gamma)\theta_L]}\right)^{1/2} = \left(\frac{1}{6}\right)^{1/2} \simeq 0.41 < \frac{1}{2} = e_H^{SI},$$

  a **downward distortion at the bottom**.

The two-distortion rule – Why (cont.)

- **Why?** Distorting e_L (the top type) would cost surplus and buy nothing in rents, since no binding constraint rewards it.
- Distorting e_H *downward*, in contrast, makes the high-cost contract less attractive to the efficient low-cost type, which relaxes the binding IC_L and lets the principal pay him a smaller information rent.
- **The bottom-line trade-off.** Each unit of downward distortion at the bottom (e_H) saves rent paid to the top (the low-cost type) but loses surplus at the bottom; the principal balances the two at the margin.

Table 6.1 – Comparative statics under AI

	γ	θ_H	θ_L	Relative to *SI*
e_L^{AI}	N/A	N/A	$-$	Same effort
e_H^{AI}	$+$	$-$	$+$	Lower effort
w_L^{AI}	$+$	$+$	$-$	Higher wage
w_H^{AI}	$+$	$-$	$+$	Lower wage

Table 6.1. Equilibrium effort and wages under asymmetric information.

- Note the **cross-type effects**: w_L^{AI} rises with θ_H, e_H^{AI} rises with θ_L.
- Each type's contract depends on *both* types' parameters.

Example 6.2 – Numerical AI

- Continuing Example 6.1: $\theta_H = 0.15$, $\theta_L = 0.1$, $\gamma = \frac{1}{3}$.
- Efforts (low type undistorted, high type bent down):

$$e_L^{AI} = \left(\frac{1}{2 \times 0.1}\right)^{1/2} = \sqrt{5} \simeq 2.24,$$

$$e_H^{AI} = \left(\frac{1/3}{2[0.15 - (2/3) \cdot 0.1]}\right)^{1/2} = \sqrt{2} \simeq 1.41.$$

- Indeed $e_H^{AI} = 1.41 < 1.83 = e_H^{SI}$, while $e_L^{AI} = 2.24 = e_L^{SI}$.

Example 6.2 – Numerical AI (wages)

- Wages:

$$\begin{aligned} w_L^{AI} &= \$\tfrac{3}{5} = \$0.60, \\ w_H^{AI} &= \$\tfrac{3}{10} = \$0.30. \end{aligned}$$

- **Sanity check.** The high type's utility is $0.30 - 0.15 \cdot 2 = 0$, while the low type's is $0.60 - 0.1 \cdot 5 = \$0.10 > 0$ – his information rent.

Example 6.2 – Numerical AI (cont.)

- Compare with SI: both wages were \$0.5. Under AI, $w_L \uparrow$ to 0.60 and $w_H \downarrow$ to 0.30; the firm earns more from the low-cost type ($0.20 > 0.047$).
- Ex-ante expected profit (weights $\gamma = \frac{1}{3}$ on the high type and $1 - \gamma = \frac{2}{3}$ on the low type):

$$E[V^*] = \tfrac{1}{3}\left(\tfrac{1}{2}\log 2 - \tfrac{3}{10}\right) + \tfrac{2}{3}\left(\tfrac{1}{2}\log 5 - \tfrac{3}{5}\right) = \frac{\log 50 - 3}{6} \simeq 0.15,$$

positive, so the principal contracts even before the worker's type is realized.

6.2.3 Information Rents

No rents for the high-cost type

- Under SI: $u_H^{SI} = \frac{1}{2} - \theta_H \cdot \frac{1}{2\theta_H} = 0$.
- Under AI: $w_H \downarrow$ but $e_H \downarrow$ too.
- Compute:

$$u_H^{AI} = \frac{\gamma\theta_H}{2[\theta_H - (1-\gamma)\theta_L]} - \theta_H \cdot \frac{\gamma}{2[\theta_H - (1-\gamma)\theta_L]} = 0.$$

- **Both effects cancel**: high-cost worker gets zero utility (reservation) in both info regimes.
- Inefficient type extracts **no rent**.

Information rents for the low-cost type

- Under SI: $u_L^{SI} = \frac{1}{2} - \theta_L \cdot \frac{1}{2\theta_L} = 0$.
- Under AI: $w_L \uparrow$, e_L unchanged.
- Compute:

$$u_L^{AI} = \frac{(1+\gamma)\theta_H - \theta_L}{2[\theta_H - (1-\gamma)\theta_L]} - \frac{1}{2} = \frac{\gamma(\theta_H - \theta_L)}{2[\theta_H - (1-\gamma)\theta_L]} > 0.$$

- This is the **information rent** – payment for not mimicking the high-cost contract.
- Comparative statics: rent grows with $(\theta_H - \theta_L)$ (more incentive to mimic) and with γ (higher chance of facing the inefficient type).

Information rent decomposition

- Mechanically: rent $u_L^{AI} - u_L^{SI} = u_L^{AI} - 0 = u_L^{AI} > 0$.
- Decomposition:

$$u_L^{AI} = \underbrace{(w_L^{AI} - w_L^{SI})}_{\text{wage gain}} - \underbrace{\theta_L (e_L^{AI})^2 - \theta_L (e_L^{SI})^2}_{\text{effort cost change}}.$$

- Since $e_L^{AI} = e_L^{SI}$, the effort cost difference is zero.

Information rent decomposition (cont.)

- All the rent comes from the wage increase:

$$u_L^{AI} = w_L^{AI} - w_L^{SI} > 0.$$

- **The cost of information asymmetry**: paid as extra wages to the efficient type.
- For the high type: $w_H \downarrow$ but $e_H \downarrow$ too, keeping utility at 0.

Example 6.3 – Measuring information rent

- With $\theta_H = 0.15$, $\theta_L = 0.1$, $\gamma = \frac{1}{3}$:

$$u_L^{AI} = \frac{\gamma(\theta_H - \theta_L)}{2[\theta_H - (1-\gamma)\theta_L]} = \frac{\frac{1}{3}(0.15 - 0.1)}{2[0.15 - \frac{2}{3} \cdot 0.1]} = \frac{1}{10} = 0.10.$$

- Positive information rent (\$0.10) for the low-cost worker.
- **Cross-check:** this equals the wage gain $w_L^{AI} - w_L^{SI} = 0.60 - 0.50 = \$0.10$, since the effort is unchanged, so the whole rent is paid as extra wage.
- In contrast $u_H^{AI} = 0$: the inefficient type earns no rent.
- **General lesson:** information rents accrue to the *efficient* type, to keep him from masquerading as inefficient.

Section 6.3
General Approach

General setup – The principal

- **Principal's profit:**

$$\pi(e, w) = x(e) - w,$$

  with $x'(e) > 0$ (effort raises revenue) and $x''(e) \leq 0$ (at a weakly decreasing rate).
- Examples: $x(e) = a + be$ (linear), $x(e) = \log e$, $x(e) = e^{1/2}$.
- The log-quadratic model of Section 6.2 is the special case $x(e) = \log e$.

General setup – The agent

- **Agent's utility:**

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

with $u' > 0$, $u'' \leq 0$ (risk-averse over wages), and $c_e > 0$, $c_{ee} \geq 0$ (effort is costly and convex).

- **Note.** Here $u(w)$ is the utility from the wage alone, so total utility is $v(w, e \mid \theta)$; the log-quadratic case is $u(w) = w$, $c(e, \theta) = \theta e^2$.
- Cross-derivative $c_{e\theta}(e, \theta) > 0$ – the **single-crossing** condition (high types have a steeper marginal cost of effort).

Single-crossing property

- Cross-derivative condition: $c_{e\theta}(e, \theta) > 0$.
- For two types $\theta_H > \theta_L$:

$$c_e(e, \theta_H) > c_e(e, \theta_L).$$

- Indifference curves in (e, w) space:

$$\begin{aligned} \text{Low type:} \qquad & w = u^{-1}\left[v + c(e, \theta_L)\right], \\ \text{High type:} \qquad & w = u^{-1}\left[v + c(e, \theta_H)\right]. \end{aligned}$$

- High-type curve *steeper* in e.
- Indifference curves cross **once**: hence "single-crossing."

Figure 6.1 – Single-crossing property

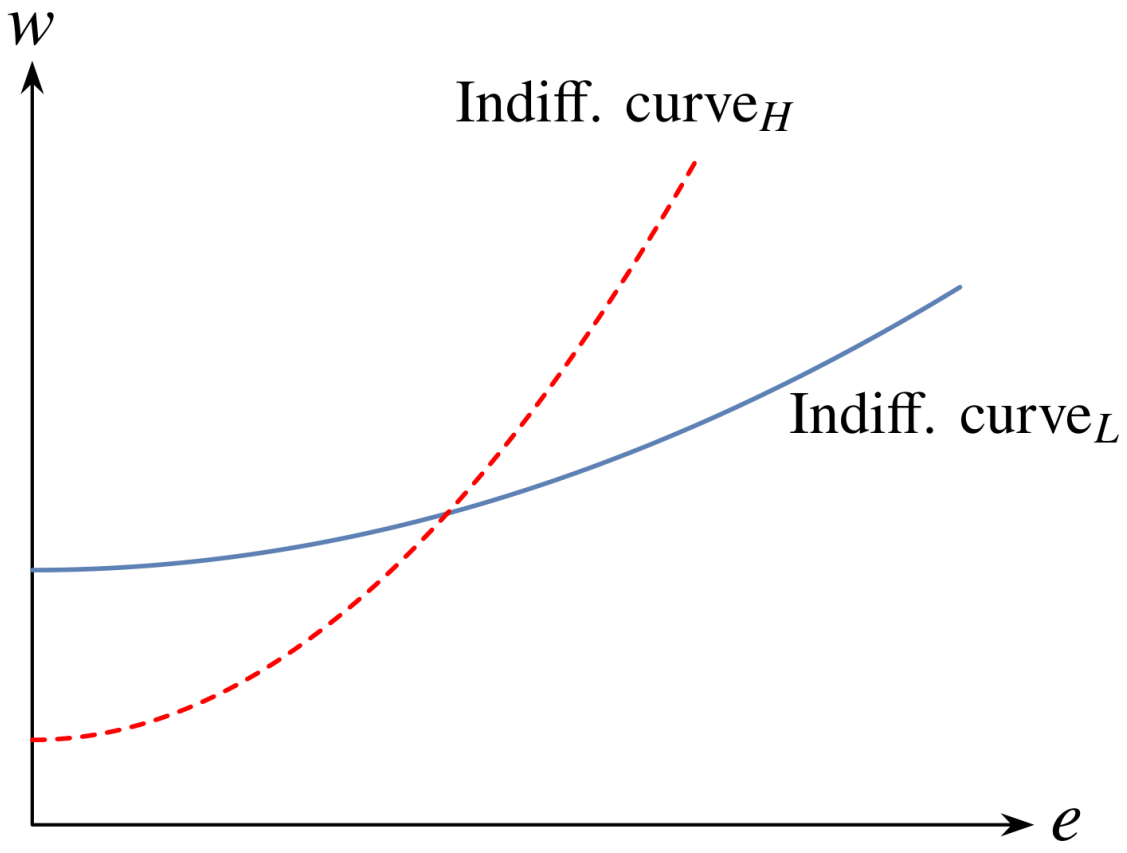


Figure 6.1. Single-crossing property: indifference curve of high-type worker is steeper than low-type's, crossing only once.

Example 6.4 – Numerical single-crossing

- Let $u(w) = \sqrt{w}$, $c(e, \theta) = \theta e^2$, and fix utility at $v = 10$.
- Low type ($\theta_L = 0.1$): from $\sqrt{w} = 10 + 0.1e^2$, squaring gives

$$w = (10 + 0.1e^2)^2.$$

- High type ($\theta_H = 0.15$): from $\sqrt{w} = 10 + 0.15e^2$, squaring gives

$$w = (10 + 0.15e^2)^2.$$

- Both curves start at $w = 100$ when $e = 0$ (same intercept).

Example 6.4 – Single-crossing (cont.)

- Slopes in e: $0.4e(10 + 0.1e^2)$ for the low type and $0.6e(10 + 0.15e^2)$ for the high type.
- Their difference is

$$0.6e(10 + 0.15e^2) - 0.4e(10 + 0.1e^2) = e(2 + 0.05e^2) > 0,$$

  so the high-type curve is always steeper.
- **Intuitively,** starting from a common intercept the high type's indifference curve rises faster, so the two curves cross only once.

Figure 6.2 – Indifference curves of the two types

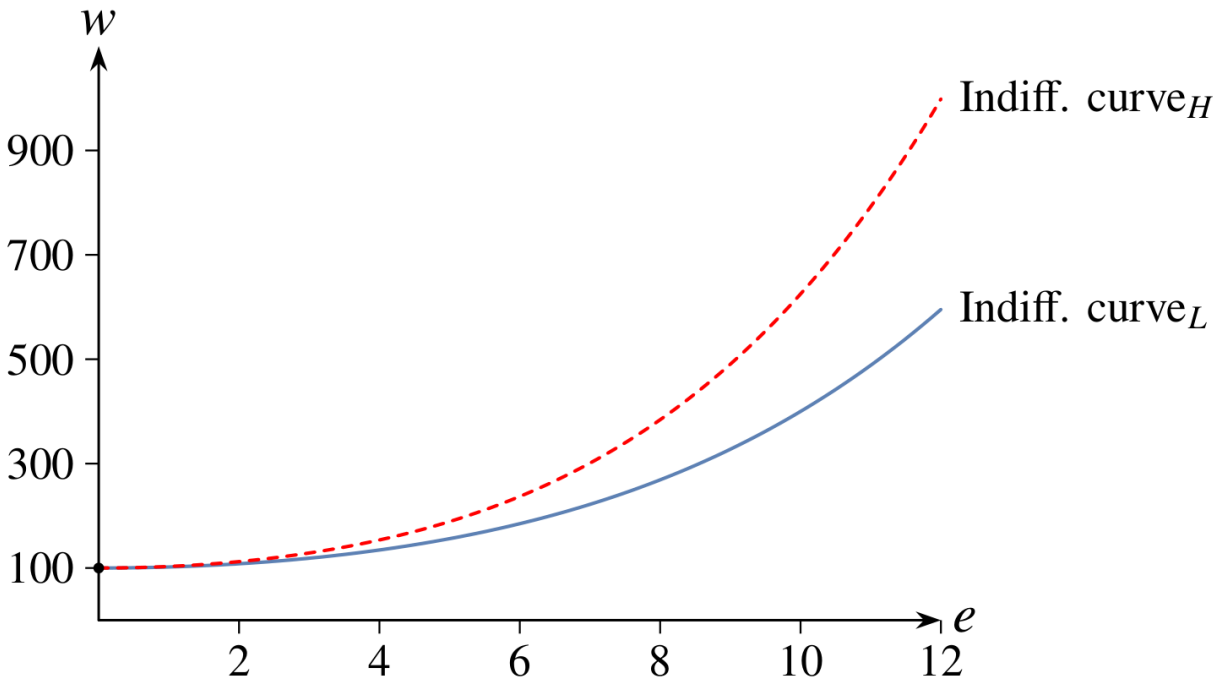


Figure 6.2. Indifference curves: low-type $w = (10 + 0.1e^2)^2$ and high-type $w = (10 + 0.15e^2)^2$. They start at the same vertical intercept but the high-type's grows faster.

Examples of single-crossing

- Single-crossing appears in many models:
  - **Labor / contracts**: high-ability workers find effort cheaper (as in this chapter).
  - **Insurance**: low-risk individuals find high deductibles less costly to accept.
  - **Pricing**: high-valuation consumers tolerate higher prices for higher quality.
  - **Auctions**: bidders with higher private values are willing to bid more.
  - **Signaling** (Ch. 5): high-ability workers find education less costly.
- Mathematical condition: $\frac{\partial}{\partial \theta}(MRS_A) > 0$ – types are ordered by their marginal trade-off.
- Without single-crossing, separation can fail and pooling becomes inevitable.

6.3.1 General SI – The MRS condition

- Principal's problem (observing θ_i):

$$\max_{w_i, e_i} \; x(e_i) - w_i \quad \text{s.t.} \quad u(w_i) - c(e_i, \theta_i) \geq 0.$$

- Steps 2–3: PC binds, $w_i = u^{-1}[c(e_i, \theta_i)]$.
- Step 4: FOC w.r.t. e_i yields the **MRS condition**:

$$x'(e_i) = \frac{c_e(e_i, \theta_i)}{u'\{w_i\}} \Leftrightarrow MRS_P = MRS_A.$$

- Principal increases effort until her marginal benefit equals agent's marginal cost.

Figure 6.3 – Effort under SI

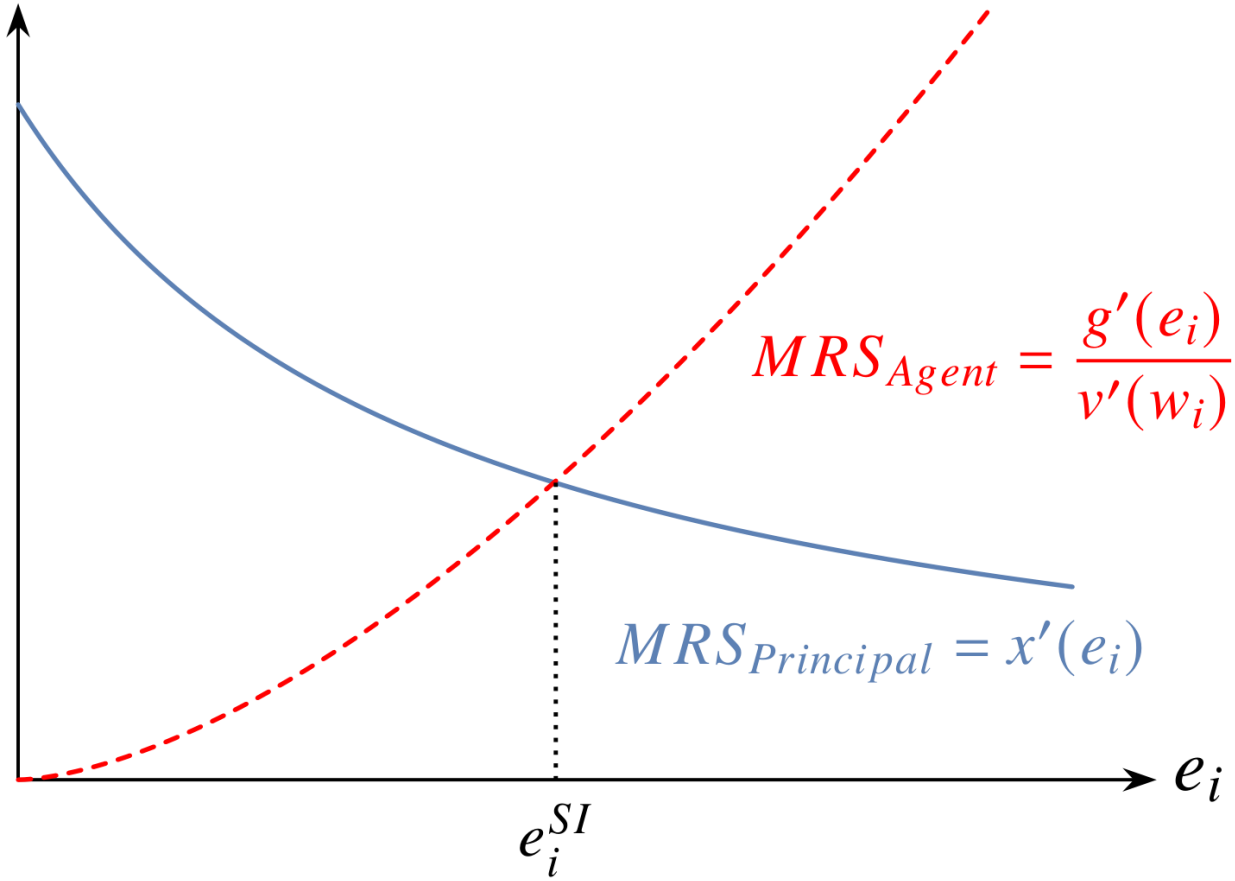


Figure 6.3. $MRS_P = x'(e)$ weakly decreasing; $MRS_A = c_e(e, \theta)/u'(w)$ weakly increasing. They cross at e_i^{SI}.

Example 6.5 – Log-quadratic SI

- $x(e) = \log e$, $u(w) = w$, $c(e, \theta) = \theta e^2$.
- Steps 1–4 yield FOC: $\frac{1}{e_i} = 2\theta_i e_i$.
- Solving: $e_i^{SI} = (1/2\theta_i)^{1/2}$, same as Section 6.2.
- Wages: $w_i^{SI} = \theta_i \cdot 1/(2\theta_i) = \frac{1}{2}$.
- Wage-effort pairs (with $\theta_H = 0.15$, $\theta_L = 0.1$):

$$\begin{aligned} (w_H^{SI}, e_H^{SI}) &= (0.5,\ 1.83), \\ (w_L^{SI}, e_L^{SI}) &= (0.5,\ 2.24). \end{aligned}$$

6.3.2 General AI – The same recipe

- Same 4 constraints (PC_H, PC_L, IC_H, IC_L) but with general $u(\cdot)$ and $c(\cdot, \cdot)$.
- **Step 2:** PC_H and IC_L bind; PC_L and IC_H are slack (by the same proof as in §6.2).
- Using $c(e, \theta) = \theta c(e)$ for tractability:

$$\begin{aligned} w_H &= u^{-1}[\theta_H c(e_H)], \\ w_L &= u^{-1}[(\theta_H - \theta_L)c(e_H) + \theta_L c(e_L)]. \end{aligned}$$

- Substitute into objective; differentiate.

General AI – FOCs

- **FOC for e_L:**

$$x'(e_L) = \frac{\theta_L c'(e_L)}{u'(w_L)} \equiv MRS_A^{SI}(e_L).$$

  This is the same condition as under SI, so **no distortion at the top.**

- **FOC for e_H:**

$$x'(e_H) = \frac{\theta_H c'(e_H)}{u'(w_H)} + \underbrace{\frac{(1-\gamma)(\theta_H - \theta_L)}{\gamma} \frac{c'(e_H)}{u'(w_L)}}_{\text{New}}.$$

- The extra term is positive, so $MRS_A^{AI}(e_H) > MRS_A^{SI}(e_H)$: raising e_H now carries an added rent cost.
- **Downward distortion at the bottom**: $e_H^{AI} < e_H^{SI}$.

Figure 6.4 – Effort under AI for the high type

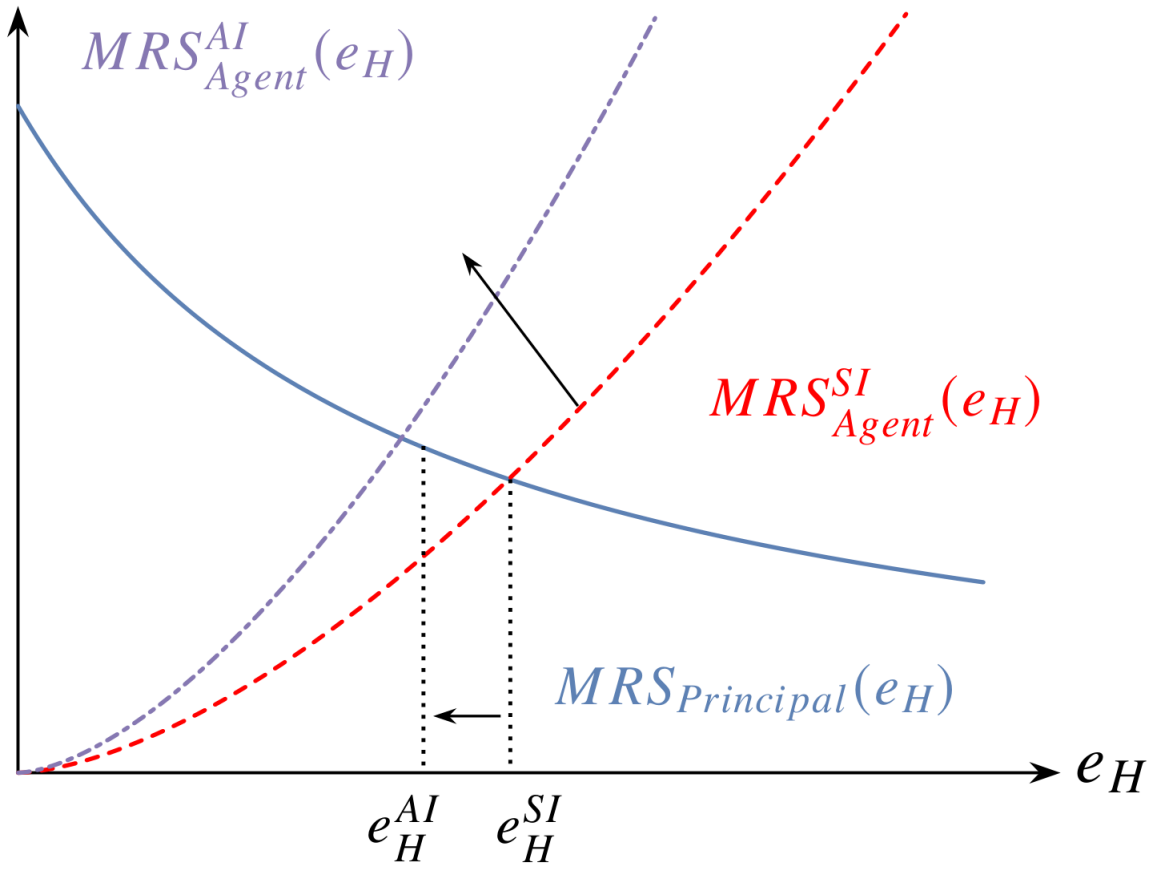


Figure 6.4. $MRS^{AI}_A(e_H)$ lies above $MRS^{SI}_A(e_H)$ for all $e_H > 0$. Hence the crossing with MRS_P happens at a lower $e^{AI}_H < e^{SI}_H$.

Example 6.6 – General AI numerical

- Same parameters as Example 6.5 plus $\gamma = \frac{1}{2}$.
- Effort levels:

$$\begin{aligned} e_L^{AI} &= \sqrt{5} \simeq 2.24 \text{ (no distortion)}, \\ e_H^{AI} &= \left[\frac{\gamma}{2(\theta_H - (1-\gamma)\theta_L)}\right]^{1/2} = \frac{\sqrt{10}}{2} \simeq 1.58. \end{aligned}$$

- Wages:

$$\begin{aligned} w_H^{AI} &= 0.15 \cdot \tfrac{5}{2} = \$\frac{3}{8}, \\ w_L^{AI} &= 0.05 \cdot \tfrac{5}{2} + 0.1 \cdot 5 = \$\frac{5}{8}. \end{aligned}$$

Example 6.6 (cont.) – Rent and profits

- Information rent: $u_L^{AI} = \frac{5}{8} - \frac{1}{2} = \frac{1}{8} = 0.125$.
- Principal's ex-post profit:

$$\begin{aligned} \log(e_L^{AI}) - w_L^{AI} &= \tfrac{1}{2}\log 5 - \tfrac{5}{8} \simeq 0.18 \text{ (low type)}, \\ \log(e_H^{AI}) - w_H^{AI} &= \tfrac{1}{2}\log \tfrac{5}{2} - \tfrac{3}{8} \simeq 0.08 \text{ (high type)}. \end{aligned}$$

- Ex-ante profit (with $\gamma = \frac{1}{2}$): $\frac{1}{2}(0.18) + \frac{1}{2}(0.08) \simeq 0.13$.
- **Reading.** The low type is easier to serve profitably; the high type earns a positive rent that the principal cannot squeeze.

6.3.3 The information rent formula

- Start with binding IC_L: $u(w_L) - \theta_L c(e_L) = u(w_H) - \theta_L c(e_H)$.
- Add/subtract $\theta_H c(e_H)$:

$$u(w_L) - \theta_L c(e_L) = \underbrace{u(w_H) - \theta_H c(e_H)}_{0 \text{ by binding } PC_H} + (\theta_H - \theta_L) c(e_H).$$

- **Information rent formula:**

$$u_L^{AI} = u(w_L) - \theta_L c(e_L) = (\theta_H - \theta_L) c(e_H) > 0.$$

- Grows with cost differential $(\theta_H - \theta_L)$ – the gap that makes mimicking attractive.
- High-cost type's rent is 0 (PC_H binds).

Reading the information rent formula

- The rent formula $u_L^{AI} = (\theta_H - \theta_L)c(e_H)$ tells us three things at once:
  - Rent is *proportional to the type gap* $\theta_H - \theta_L$: bigger heterogeneity, bigger rent.
  - Rent is *proportional to the high type's distorted effort* $c(e_H)$: the more the bottom type is asked to do, the more tempting mimicry becomes.
  - Rent *vanishes* when types coincide ($\theta_H = \theta_L$) or when the high type's effort is driven to zero.
- **Design lever.** The principal trades off the rent against surplus by choosing how far down to distort e_H.
- **In practice,** the rent shows up in phone plans, insurance menus, and airline fare classes as the extra surplus captured by heavier users, riskier customers, or premium travelers.

Section 6.4.1
Screening Using Product Qualities

Setup – Product-quality screening

- Monopolist offers a good with quality q and price p.
- Two consumer types A (high quality preference) and B (low):

$$u_A = \theta_A q - p, \quad u_B = \theta_B q - p,$$

  with $\theta_A > \theta_B > 0$.
- Production cost: $c(q) = \frac{q^2}{2}$ (increasing, convex).
- Each consumer buys at most one unit.
- Monopolist chooses (p_i, q_i) for each type.

Symmetric information

- Monopolist solves:

$$\max_{p_i, q_i} \; p_i - \frac{q_i^2}{2} \quad \text{s.t.} \quad \theta_i q_i - p_i \geq 0. \quad (PC_i)$$

- PC_i binds: $p_i = \theta_i q_i$.
- Substitute: $\max_{q_i} \; \theta_i q_i - \frac{q_i^2}{2}$.

Symmetric information (cont.)

- FOC $\theta_i = q_i$ gives $q_i^{SI} = \theta_i$ (quality matches the taste for it). Hence:
$$q_A^{SI} = \theta_A > \theta_B = q_B^{SI}, \quad p_i^{SI} = \theta_i^2.$$
- Profit: $\pi_i^{SI} = \theta_i^2 - \frac{\theta_i^2}{2} = \frac{\theta_i^2}{2}$.
- The binding PC_i means all consumer surplus is extracted: each type is left exactly indifferent to not buying.

Asymmetric information

- Now monopolist can't observe θ_i; $\Pr(A) = \gamma$.
- Four constraints: PC_A, PC_B, IC_A, IC_B.
- **Step 2:** PC_B and IC_A bind (analogous to PC_H and IC_L in §6.2 – the "least efficient" here is B; the "most efficient" is A).
- From binding constraints:

$$p_B = \theta_B q_B, \quad p_A = \theta_A(q_A - q_B) + \theta_B q_B.$$

- Substitute into profit; solve for q_A, q_B.

Equilibrium qualities

- The FOCs yield:

$$\begin{aligned} q_A^{AI} &= \theta_A \quad \text{(no distortion at the top)}, \\ q_B^{AI} &= \frac{\theta_B - \gamma\theta_A}{1-\gamma} < \theta_B = q_B^{SI} \quad \text{(distorted down)}. \end{aligned}$$

- Here the high-valuation type A is the "top" (undistorted) and B is the "bottom" (distorted).
- $q_B^{AI} > 0$ only if $\gamma < \theta_B/\theta_A$; otherwise the monopolist serves the high type alone.

Equilibrium prices

- Insert the qualities into the binding constraints:

$$\begin{aligned} p_B^{AI} &= \theta_B \frac{\theta_B - \gamma\theta_A}{1-\gamma}, \\ p_A^{AI} &= \frac{\theta_A^2 - (1+\gamma)\theta_A\theta_B + \theta_B^2}{1-\gamma}. \end{aligned}$$

- Both fall relative to SI, $p_A^{AI} < p_A^{SI}$ and $p_B^{AI} < p_B^{SI}$: the monopolist must leave the high type an information rent, so it cannot charge the full first-best price.

Real-world product-line examples

- **Airlines**: economy, business, first class – self-selection by willingness to pay for quality.
- **Software**: free / standard / premium versions; tiered feature sets.
- **Streaming services**: basic / standard / premium plans based on resolution and number of screens.

Real-world product-line examples (cont.)

- **Restaurants**: prix fixe menus offering different quality tiers.
- **Universities**: undergraduate / graduate / executive education programs at very different prices.
- In each case, the high-quality option is undistorted, while the low-quality option is deliberately degraded – to make the high-quality one attractive to high-WTP customers without losing low-WTP customers.

Sales vs. margin trade-off

- Monopolist's trade-off:
  - **Margin**: higher price extracts more from high-type but loses low-type.
  - **Sales**: lower price attracts both types but smaller margin per unit.
- Condition $\gamma < \frac{\theta_B}{\theta_A} \Leftrightarrow$ high-types relatively rare $\Rightarrow$ **serve both types**.
- Otherwise: $\gamma \geq \frac{\theta_B}{\theta_A} \Rightarrow$ **serve only high-types** (riskier strategy but worth it when they are common).

Sales vs. margin trade-off (cont.)

- Quality differential:

$$\Delta q^{AI} = q_A^{AI} - q_B^{AI} = \frac{\theta_A - \theta_B}{1 - \gamma},$$

  increasing in the taste gap $\theta_A - \theta_B$ and in γ.
- **Intuitively,** the more the monopolist fears the high type mimicking the low one, the more it spreads the two qualities apart to keep each type self-selecting.
- As $\gamma \to 1$ the low type is rarely present, so the firm bends q_B down sharply and the gap widens.

Example 6.7 – Quality screening numerical

- $\theta_A = \frac{2}{3}$, $\theta_B = \frac{1}{2}$, $\gamma = \frac{1}{3}$.
- Since $\gamma = \frac{1}{3} < \frac{1/2}{2/3} = \frac{3}{4}$, both types are served.
- Qualities:

$$\begin{aligned} q_A^{SI} &= q_A^{AI} = \frac{2}{3}, \\ q_B^{SI} &= \frac{1}{2}, \quad q_B^{AI} = \frac{1/2 - 1/3 \cdot 2/3}{1 - 1/3} = \frac{5}{12} \simeq 0.417. \end{aligned}$$

- The top type's quality is undistorted; the bottom type's drops from 0.5 to 0.417.

Example 6.7 – Prices and profit (cont.)

- Prices: $p_A^{SI} = 4/9 \simeq 0.444$, $p_A^{AI} = 3/8 \simeq 0.375$; $p_B^{SI} = 0.25$, $p_B^{AI} = 5/24 \simeq 0.208$ – both fall under AI.
- Profit: $\pi^{AI} = \frac{19}{144} \simeq 0.132$.
- **Sanity check.** This beats serving only A ($\frac{2}{27} \simeq 0.074$) but falls short of the first best ($\pi^{SI} = \frac{17}{108} \simeq 0.157$) – the cost of hidden information.

Section 6.4.2
Non-Linear Pricing in Monopoly

Setup – Non-linear pricing

- Consumer utility: $u_i(x, y) = \theta_i \sqrt{x} + y$, where x is units of the good and y the numeraire (money for everything else).
- Types $\theta_H > \theta_L \geq 0$, with $\Pr(L) = \gamma$; the high type values the good more at every quantity.
- Monopolist's unit cost $c > 0$ with $\theta_L > c$, so even the low type is worth serving.

Setup – Non-linear pricing (cont.)

- Three pricing regimes:
  1. **Uniform pricing**: single per-unit price p.
  2. **Single two-part tariff**: (T, p) for both types.
  3. **Menu of quantity–charge bundles**: (T_H, x_H) and (T_L, x_L).
- Goal: rank profits across the three regimes.

Why two-part tariffs help

- Uniform pricing leaves consumer surplus on the table – consumers keep the surplus from each marginal unit beyond the price.
- A **fixed fee** T on top of p can extract this surplus without distorting marginal consumption.
- Classic real-world examples:
  - Cell-phone plans: monthly fee + per-minute or per-GB charge.
  - Costco: annual membership + low per-unit prices.
  - Theme parks: entry ticket + per-ride or per-meal pricing.
  - Cable TV: monthly subscription + pay-per-view.

Why two-part tariffs help (cont.)

- Two-part tariffs are a classic "screening through pricing structure" device.
- The per-unit price p controls how much each consumer buys; the fixed fee T then skims the surplus that consumption generates.
- **Intuitively,** splitting the charge into two parts lets the firm set the quantity with p and capture rents with T, instead of using one blunt price for both jobs.

Regime 1 – Uniform pricing

- Each consumer chooses quantity from $\frac{\theta_i}{2\sqrt{x}} = p$, giving demand $x_i(p) = \frac{\theta_i^2}{4p^2}$.
- Aggregate (expected) demand:

$$x(p) = \frac{(1-\gamma)\theta_H^2 + \gamma\theta_L^2}{4p^2}.$$

- The monopolist solves $\max_p (p-c)x(p)$; the FOC yields

$$p^U = 2c.$$

Regime 1 – Uniform pricing (profit)

- Equilibrium profit (insert $p^U = 2c$):

$$\pi^U = \frac{(1-\gamma)\theta_H^2 + \gamma\theta_L^2}{16c}.$$

- Decreasing in the share of low-type consumers:

$$\frac{\partial \pi^U}{\partial \gamma} = -\frac{\theta_H^2 - \theta_L^2}{16c} < 0.$$

- **Intuitively,** more low-value customers pull down the expected margin, since a single price cannot separate the two types.

Regime 2 – Single two-part tariff (both types)

- Add a fixed fee T on top of price p.
- The low type has the smaller surplus, so his PC is the binding one: PC_L binds while PC_H is slack ($T \leq \theta_L^2/4p < \theta_H^2/4p$), with $S_L(p) = \theta_L^2/4p$.
- Set $T = S_L(p)$, substitute, and differentiate in p:

$$p^{ST} = \frac{2c\left[\gamma\theta_L^2 + (1-\gamma)\theta_H^2\right]}{(1+\gamma)\theta_L^2 + (1-\gamma)\theta_H^2}.$$

Regime 2 – Single two-part tariff (cont.)

- Fee:

$$T^{ST} = \frac{\theta_L^2 \left[(1+\gamma)\theta_L^2 + (1-\gamma)\theta_H^2\right]}{8c\left[\gamma\theta_L^2 + (1-\gamma)\theta_H^2\right]}.$$

- Profit:

$$\pi^{ST} = \frac{\left[(1+\gamma)\theta_L^2 + (1-\gamma)\theta_H^2\right]^2}{16c\left[\gamma\theta_L^2 + (1-\gamma)\theta_H^2\right]}.$$

Regime 2 – Serving only the high type

- Alternative: ignore low type, set T, p to maximize profit only from high type.
- PC_H binds: $T = \theta_H^2/4p$.
- FOC yields $p^H = c$. Fee: $T^H = \theta_H^2/4c$.

Regime 2 – Serving only the high type (cont.)

- Profit:

$$\pi^H = \frac{(1-\gamma)\theta_H^2}{4c}.$$

- **Cutoff** for choosing between the two strategies:

$$\frac{\theta_H}{\theta_L} > \bar{\theta} \equiv \left[\frac{(1-\gamma)+2\sqrt{1+\gamma+\gamma^2}}{3(1-\gamma)}\right]^{1/2}.$$

- If above the cutoff the monopolist prefers π^H (high types alone); otherwise π^{ST} (both types).
- **Intuitively,** a large taste gap makes the high type worth targeting alone, since serving the low type would force a big price cut on everyone.

Regime 3 – Menu of quantity–charge bundles

- Offer (T_L, x_L) and (T_H, x_H).
- Same 4 constraints as in §6.2 – PC_L and IC_H bind.
- From binding constraints:

$$\begin{aligned} T_L &= \theta_L \sqrt{x_L}, \\ T_H &= \theta_H \sqrt{x_H} - (\theta_H - \theta_L)\sqrt{x_L}. \end{aligned}$$

- Substitute these tariffs into expected profit, leaving x_H and x_L as the only choice variables.

Regime 3 – Optimal quantities

- Differentiating expected profit gives

$$\begin{aligned} x_H &= \frac{\theta_H^2}{4c^2} \quad \text{(undistorted, top type)}, \\ x_L &= \left(\frac{\theta_L - (1-\gamma)\theta_H}{2\gamma c} \right)^2 \quad \text{(distorted down)}. \end{aligned}$$

- The high type's quantity sets his marginal value equal to cost, $\frac{\theta_H}{2\sqrt{x_H}} = c$; the low type's is bent below first-best to limit the high type's rent.

Menu – Tariffs

- Substitute the quantities into the binding constraints to get the lump-sum charges:

$$\begin{aligned} T_L &= \frac{\theta_L[\theta_L - (1-\gamma)\theta_H]}{2\gamma c}, \\ T_H &= \frac{\theta_H^2 - (2-\gamma)\theta_H\theta_L + \theta_L^2}{2\gamma c}. \end{aligned}$$

- As γ rises, T_L rises and T_H falls – the high type is charged less to keep his bundle incentive compatible.

Menu – Profit and dominance

- Profit:

$$\pi^{MT} = \frac{(1-\gamma)\theta_H^2 - 2(1-\gamma)\theta_L\theta_H + \theta_L^2}{4\gamma c},$$

  decreasing in γ.
- **Key result:** π^{MT} **strictly exceeds** $\max\{\pi^H, \pi^{ST}\}$ and π^U, so the menu dominates every simpler scheme.

Why the menu dominates a single tariff

- Single tariff treats both types the same: forces a single (T, p) or just serves the high type.
- Menu of tariffs:
  - Offers different bundles (T_L, x_L) and (T_H, x_H).
  - Each type self-selects into the bundle designed for her.
- As in screening models: low-type gets distorted quantity $x_L < x_L^{SI}$; high-type gets undistorted $x_H = x_H^{SI}$.

Why the menu dominates a single tariff (cont.)

- Information rent flows to the high type via lower T_H (relative to what would fully extract her surplus).
- Menu pricing always dominates because it can replicate any single tariff, plus optimize the second contract.

Menu vs. single tariff – The teaching lesson

- A single two-part tariff either exploits the high type (leaving the low type out) or serves both at the cost of leaving rents on the table.
- A *menu* of tariffs unbundles those goals: each type gets a bundle designed for him – more surplus captured without excluding anyone.
- **This is the essence of screening.** The principal is *sorting* types through their own choices, using their willingness to pay for extra units.
- **Same idea in many markets:** airline fare buckets, mobile data plans, streaming-service tiers, and quantity discounts on wholesale.

Profit ranking

- Assume $\gamma = \frac{1}{2}$. Two cases on θ_H/θ_L:
- **Case 1:** $\theta_H/\theta_L > \sqrt{(1+2\sqrt{7})/3} \simeq 1.45$. Then:

$$\pi^U < \pi^H < \pi^{MT}.$$

- **Case 2:** $\theta_H/\theta_L \leq \sqrt{(1+2\sqrt{7})/3}$. Then:

$$\pi^U < \pi^{ST} < \pi^{MT}.$$

Profit ranking (cont.)

- **Universal ranking:**

$$\pi^{U} < \pi^{\text{Single}} < \pi^{MT}.$$

- Holds for any $\gamma \in (0, 1)$, not only $\gamma = \frac{1}{2}$.
- Sophistication of pricing increases extractable surplus.

Example 6.8 – Two numerical illustrations

- $\gamma = \frac{1}{2}$, $\theta_L = \frac{1}{2}$, $c = 1$.
- **First example** ($\theta_H = 3/4$, ratio $1.5 > 1.45$): serve only high.
  - $\pi^U \simeq 0.025$, $\pi^H \simeq 0.07$, $\pi^{MT} \simeq 0.078$.
- **Second example** ($\theta_H = 2/3$, ratio $1.33 < 1.45$): serve both with single tariff.
  - $\pi^U \simeq 0.022$, $\pi^{ST} \simeq 0.064$, $\pi^{MT} \simeq 0.069$.

Example 6.8 (cont.)

- In both examples π^{MT} is highest, confirming $\pi^{MT} > \pi^{\text{Single}} > \pi^{U}$.
- The only thing that differs across the two is whether the single tariff serves everyone or the high type alone – the menu wins either way.
- **Economically,** each step toward a richer pricing scheme relaxes an information constraint and lets the firm keep more surplus.

Table 6.2 – Profits across regimes

		Menu pricing	*Single tariff*	*Uniform pricing*
θ_H	Served	π^{MT}	$\pi^{H}\,/\,\pi^{ST}$	π^{U}
3/4	Only high	0.078	0.07	0.025
2/3	Everyone	0.069	0.064	0.022

Table 6.2. Profits across different pricing strategies.

- **Always optimal** to use the menu of quantity–charge bundles.
- Each step "up" relaxes the information constraint and extracts more surplus.

Remark – Quantity bundles vs. genuine two-part tariffs

- The menu above specifies a **quantity and a total charge**, (x_i, T_i): the monopolist forces the quantity each type consumes. This is the standard **nonlinear-pricing** (second-degree price-discrimination) menu.
- It is *not* a menu of genuine two-part tariffs (T_i, p_i). Facing a per-unit price p_i, a consumer chooses her own quantity along demand $x_i(p_i) = \dfrac{\theta_i^2}{4p_i^2}$, which need not equal the second-best x_i.

Section 6.5
Summary

Remark – Quantity bundles (cont.)

- Hence a menu of *genuine* two-part tariffs cannot, in general, implement the second-best outcome, and earns weakly lower profit.
- This previews the **Revelation Principle** (Ch. 7): the optimum is reached by a *direct* mechanism (quantity bundles), whereas a familiar *indirect* instrument (two-part tariffs) need not be optimal.

Summary – Symmetric information

- Principal observes type θ; chooses (w, e) for each type.
- PC binds; principal extracts **all** surplus.
- In the log-quadratic model:

$$e_i^{SI} = \left(\frac{1}{2\theta_i}\right)^{1/2}, \quad w_i^{SI} = \frac{1}{2}.$$

- Same wage; lower effort from high-cost type.
- General MRS condition: $x'(e_i) = \dfrac{c_e(e_i, \theta_i)}{u'(w_i)}$.

Summary – Asymmetric information

- Principal designs a **menu of contracts**; four constraints.
- **Step 2:** PC_H and IC_L bind; PC_L and IC_H slack.
- **Two universal results:**
  - *No distortion at the top*: $e_L^{AI} = e_L^{SI}$.
  - *Downward distortion at the bottom*: $e_H^{AI} < e_H^{SI}$.

Summary – Asymmetric information (cont.)

- **Information rent:**

$$u_L^{AI} = (\theta_H - \theta_L)c(e_H) > 0.$$

- Inefficient type gets 0; efficient type gets positive rent.

Summary – Applications and single-crossing

- **Single-crossing property** ($c_{e\theta} > 0$): indifference curves of types cross only once – the technical backbone of separation.
- **Product-quality screening:** $q_A^{AI} = \theta_A$ (no distortion); $q_B^{AI} = \frac{\theta_B - \gamma\theta_A}{1-\gamma}$ (downward distortion).
- **Non-linear pricing:** menu $>$ single tariff $>$ uniform.

Summary – Applications (cont.)

- Same logic governs every adverse-selection model in the book:
  - Top type undistorted.
  - Bottom type distorted.
  - Top type receives rents.

Appendix 6.A
Lagrangian Proof that PC_H and IC_L Bind

Appendix 6.A – Lagrangian setup

- Lagrangian:

$$\begin{aligned}\mathcal{L} \;=\; & \gamma[x(e_L) - w_L] + (1-\gamma)[x(e_H) - w_H] \\ & +\lambda_H[u(w_H) - c(e_H, \theta_H)] \\ & +\lambda_L[u(w_L) - c(e_L, \theta_L)] \\ & +\mu_H[u(w_H) - c(e_H, \theta_H) - u(w_L) + c(e_L, \theta_H)] \\ & +\mu_L[u(w_L) - c(e_L, \theta_L) - u(w_H) + c(e_H, \theta_L)].\end{aligned}$$

- Multipliers $\lambda_H, \lambda_L, \mu_H, \mu_L \geq 0$.
- Since PC_L is slack, $\lambda_L = 0$; remove that term.
- Take FOCs in w_L, w_H, e_L, e_H to confirm $\lambda_H > 0$ and $\mu_L > 0$ – so PC_H and IC_L bind in equilibrium.

Connecting to earlier chapters

- **Chapters 2–4** (moral hazard): hidden *action* after contracting.
- **Chapter 5** (adverse selection in markets): hidden *type* in market settings (lemons, signaling, screening).
- **Chapter 6** integrates: hidden type inside a principal-agent contract.

Connecting to earlier chapters (cont.)

- Common structure across all chapters:
  - Some constraint binds; others are slack.
  - Optimal contract balances participation, incentives, and rents.
  - Information frictions $\Rightarrow$ second-best outcomes.
- Chapter 7 takes a final step: **mechanism design** generalizes to many agents and arbitrary social objectives.

Empirical evidence on screening

- Screening contracts have been documented in many markets:
  - **Auto insurance:** Cohen & Einav (2007) – menu of deductibles, low-risk drivers choose high deductibles.
  - **Health insurance:** Cardon & Hendel (2001) – limited evidence of adverse selection in employer-sponsored plans.
  - **Annuities:** Finkelstein & Poterba (2004) – long-lived buyers self-select into longer-payout annuities.
  - **Credit cards:** Ausubel (1991) – card issuers offer multiple price-feature combinations.
- Findings are mixed but consistent with the theoretical prediction: menus enable separation, with information rents flowing to private types with valuable information.

Looking ahead – Chapter 7

- Chapter 6 dealt with a single principal designing contracts for agents of unknown type.
- Chapter 7 generalizes to **mechanism design**:
  - Designer (often a social planner) chooses the rules of the game.
  - **Direct** vs. **indirect** revelation mechanisms.
  - **Bayesian** (BIC) and **dominant-strategy** (DSIC) incentive compatibility.
  - **Revelation principle**: WLOG focus on direct, truthful mechanisms.
  - **Groves** and **VCG** mechanisms achieve allocative efficiency.
  - **Myerson's theorem** and **revenue equivalence**.
- Throughout: same logic of constraints, rents, and screening – now generalized to multi-agent settings.

Policy implications

- **Regulation of monopolies**: when monopolists screen using menus, they extract more surplus – but consumers also benefit from having access to differentiated products.
- **Antitrust:** forced uniform pricing may reduce profits but also reduce welfare (fewer types served).
- **Health-insurance design:** mandatory pooling can prevent adverse-selection unraveling but reduces choice for low-risk types.

Policy implications (cont.)

- **Tax policy:** optimal income tax = principal-agent problem where government screens workers by ability.
- **Regulation of utilities:** two-part tariffs (connection fee + usage charge) allowed under cost-recovery rules.
- Understanding screening is crucial for any policy that affects markets with hidden information.

Extensions and open questions

- **More than two types**: same logic, but ICs across adjacent types matter most ("local" ICs).
- **Continuous types**: leads to Myerson's optimal mechanism – Chapter 7.
- **Random allocation**: with continuous types, optimal mechanism may involve stochastic outcomes (lotteries).

Extensions and open questions (cont.)

- **Risk-averse principal**: optimal contract may share more risk; but the no-distortion-at-top result is robust.
- **Multidimensional types**: when types differ along multiple dimensions, screening becomes much harder.
- **Dynamic contracts**: with repeated interactions, principal can update beliefs and adjust contracts – but ratchet effects emerge.

From mechanism design to optimal taxation

- Mirrlees (1971): optimal income tax problem.
  - Government (principal) cannot observe worker ability (type).
  - Designs a tax schedule = menu of (income, tax) pairs.
  - Workers self-select labor supply.
- Same structural results apply:
  - No distortion at the top (highest-ability worker faces zero marginal tax in the limit).
  - Downward distortion at the bottom (low-ability faces high effective marginal tax via phase-outs).
  - Information rents to higher-ability workers.
- Modern public economics is built on these principles – Diamond (1998), Saez (2001).

Common pitfalls in solving these problems

- **Wrong binding constraints**: blindly assume all PCs bind – no, only PC_H does.
- **Forgetting to verify slack constraints**: must check PC_L and IC_H hold ex post.
- **Confusing types**: "H" here = high cost = inefficient; elsewhere "H" may mean high productivity. Read context carefully.

Common pitfalls (cont.)

- **Sign errors in information rents**: rent formula has $(\theta_H - \theta_L) > 0$ – always positive.
- **Skipping single-crossing check**: many results break down without it.
- **Wage**: under SI, the same flat wage; under AI, the *efficient* type gets the higher wage.

Mathematical toolbox used in this chapter

- **Constrained optimization**: Lagrangian with multiple constraints; binding vs. slack.
- **First-order conditions**: in w_i, e_i for SI; in e_H, e_L for AI after substituting binding constraints.
- **Inverse functions**: $u^{-1}[\cdot]$ to solve for w from PC.

Mathematical toolbox (cont.)

- **Single-crossing**: comparing slopes of indifference curves.
- **Comparative statics**: partial derivatives w.r.t. $\gamma, \theta_H, \theta_L$.
- **Substitution**: reducing four-variable problem to two choice variables.
- All of these techniques will reappear (and be extended) in Chapter 7.

Common student mistakes to avoid (Chapter 6)

- **Writing IC on the wrong type.** Under adverse selection, IC restricts each type from claiming to be the *other* type. The binding IC is usually the high type's temptation to claim to be the low type, not the reverse.
- **Assuming PC binds for every type.** Only the low-cost type's PC binds; the high-cost type earns an information rent.
- **Distorting the top.** No distortion at the top is a result, not an assumption; distorting e_H leaves surplus on the table without shrinking any rent.

Common student mistakes (Chapter 6) (cont.)

- **Ignoring single-crossing.** The whole separating menu relies on single-crossing; without it the incentive constraints do not sort types by a scalar.
- **Confusing screening with signaling.** Under screening, the *uninformed* party moves first; under signaling (Chapter 5), the *informed* party moves first.

Key takeaways from Chapter 6

1. Under symmetric info, principal extracts all surplus; both types get reservation utility.
2. Under asymmetric info, two constraints bind (PC_H, IC_L); two are slack (PC_L, IC_H).
3. **No distortion at the top**: $e_L^{AI} = e_L^{SI}$ (or $q_A^{AI} = q_A^{SI}$).

Key takeaways from Chapter 6 (cont.)

1. **Downward distortion at the bottom**: $e_H^{AI} < e_H^{SI}$ (or $q_B^{AI} < q_B^{SI}$).
2. **Information rent** for the efficient type:

$$u_L^{AI} = (\theta_H - \theta_L)c(e_H) > 0.$$

3. **Sophisticated pricing pays**: $\pi^{MT} > \pi^{ST} > \pi^{U}$.
4. Single-crossing is the technical backbone enabling separation.

End of Chapter 6

Chapter 7

Mechanism Design

Instructor Slides

Contract Theory: An Introduction with Applications

Ana Espinola-Arredondo, Felix Muñoz-Garcia, and Pak-Sing Choi

Outline of Chapter 7

- This chapter switches from analyzing given games to **designing the rules** – mechanism design.
- Sections:
  - 7.1 Introduction. 7.2 Model.
  - 7.3–7.4 Implementation, BIC, DSIC.
  - 7.5 Revelation principle.
  - 7.6 Allocative efficiency.
  - 7.7 Groves and VCG mechanisms.
  - 7.8 Participation constraints (ex-ante / interim / ex-post).
  - 7.9 Linear utility, Myerson's theorem, revenue equivalence.
  - 7.10 Optimal Bayesian mechanism.
  - 7.11 Summary. 7.12 Exercises.

What you will learn

- How to model a **mechanism** as a Bayesian game.
- The distinction between **indirect** (IRM) and **direct** revelation mechanisms (DRM).
- Two incentive-compatibility concepts: **BIC** (Bayesian) and **DSIC** (dominant-strategy).

What you will learn (cont.)

- The **revelation principle**: any outcome an IRM can implement can be implemented by a DRM.
- **Groves** and **VCG** mechanisms: how to achieve allocative efficiency in dominant strategies.
- **Myerson's theorem**, the **revenue equivalence theorem**, and how to design the **optimal Bayesian mechanism**.
- Virtual valuations and the information-rent intuition.

Section 7.1
Introduction

From given games to designed games

- In Chapters 2–6 we took the rules of the game as given.
- In mechanism design, the planner **chooses** the rules to achieve her objective:
  - Auctioneer designs the auction format to maximize revenue.
  - Government designs taxes / subsidies / procurement rules.
  - Regulator designs price schedules under hidden costs.
- **Central question:** which game design ("mechanism") can induce agents to choose strategies that maximize the designer's objective?

From given games to designed games (cont.)

- Often the designer cannot observe agents' preferences directly: a bidder's valuation, a citizen's taste for a bridge, a firm's cost.
- So she must **elicit** that information through the rules themselves, designing the game so that agents reveal what she needs.
- This is the sense in which mechanism design is "reverse game theory": we fix the equilibrium behavior we want and search for a game that delivers it.

Roadmap for Chapter 7

- Definitions (§7.2): players, types, mechanisms, indirect vs. direct.
- Incentive properties (§7.3–7.4):
  - BIC: truth-telling is a Bayesian Nash equilibrium.
  - DSIC: truth-telling is a dominant strategy.
- Revelation principle (§7.5): focus WLOG on DRMs.

Roadmap for Chapter 7 (cont.)

- Specific mechanisms (§7.6–7.7): Groves, VCG.
- Refinements (§7.8): three flavors of participation constraint.
- Linear environment (§7.9): Myerson + revenue equivalence.
- Optimal mechanisms (§7.10): virtual valuations.
- Throughout: same logic as Chapters 5–6, now generalized.

Section 7.2
Model

Players and types

- Players $i = 1, 2, \ldots, n$. We set the mechanism up as a **Bayesian game**: each player holds private information.
- Each player privately observes his **type** $\theta_i \in \Theta_i$ – it encodes his preferences (willingness to pay for an object, valuation of a public project, production cost).
- **State**: the full type profile $\theta = (\theta_1, \ldots, \theta_n) \in \Theta$, where $\Theta \equiv \Theta_1 \times \cdots \times \Theta_n$.
- Why "state"? Once θ is fixed, everything the designer cares about is pinned down; the designer's problem is that she never sees it.

Players and types (cont.)

- The draw of θ follows a **common prior** $\phi(\cdot)$ over Θ: player i's own draw θ_i is private, but the distribution ϕ is common knowledge.
- **Quasilinear** preferences:

$$v_i(k, t_i, \theta_i) = \underbrace{u_i(k, \theta_i)}_{\text{value of allocation}} + \underbrace{t_i}_{\text{money}},$$

  where $k \in K$ is the allocation and t_i is a monetary transfer.
- The reading: utility is the value of the allocation plus money, so transfers enter one-for-one – no wealth effects, which lets us add utilities across agents.
- Transfer $t_i > 0$: agent *receives* money. $t_i < 0$: agent *pays* (e.g., a tax funding the project).

Mechanism designer

- Designer (seller, government) cannot observe types θ_i, yet wants an **outcome** $(k, t_1, \ldots, t_n)$ that depends on them – an allocation $k \in K$ plus a transfer to each player.
- **Self-financing** (no outside funds):

$$\sum_{i=1}^{n} t_i \leq 0.$$

  Strict inequality: designer keeps a surplus; equality: budget balanced.
- Set of feasible outcomes:

$$X = \left\{ (k, t_1, \ldots, t_n) : k \in K, \ t_i \in \mathbb{R}, \ \sum_i t_i \leq 0 \right\}.$$

- **Choice rule**: $f(\theta) = (k(\theta), t_1(\theta), \ldots, t_n(\theta))$ maps each type profile into an outcome.

7.2.2 Indirect revelation mechanism (IRM)

- **IRM**: collection $\Gamma = \{S_1, \ldots, S_n, g(\cdot)\}$ with action sets S_i and outcome function

$$g : S_1 \times \cdots \times S_n \to X.$$

- Player i's pure strategy $s_i : \Theta_i \to S_i$ maps his private type into an action (e.g., a bid).
- Payoff $v_i(g(s), \theta_i)$: the action profile s produces an outcome $g(s)$, valued through i's type.

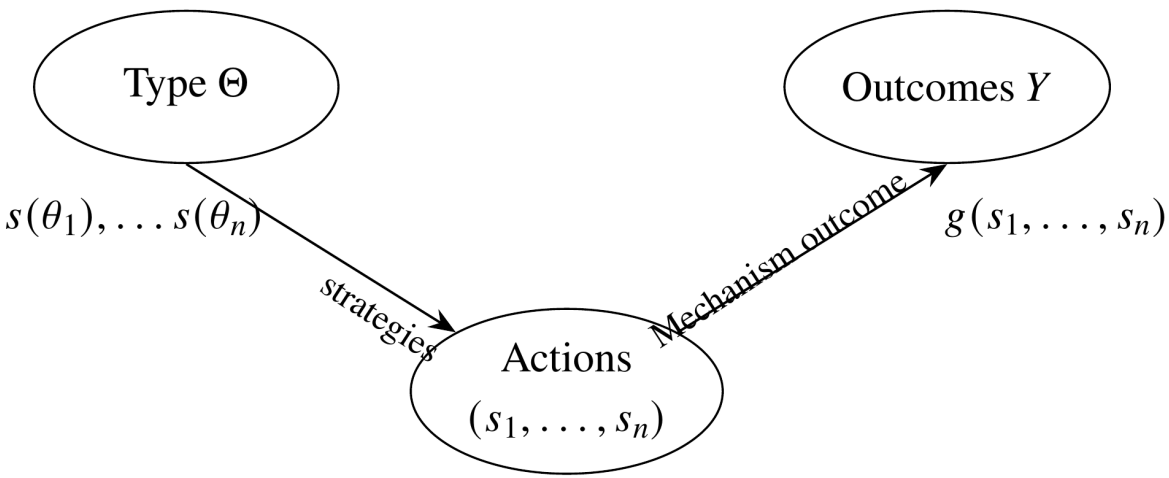


Figure 7.1. Indirect revelation mechanism.

Indirect revelation mechanism (cont.)

- The mechanism works in **two steps**:
  1. Each type picks an action, $s_i(\theta_i)$ (the strategy stage).
  2. The action profile is mapped into an outcome, $g(s)$ (the outcome stage).
- Intuitively, the designer never sees types directly; she sees only the actions they induce, and reads information off the equilibrium that results.
- A first-price auction is an IRM: the action is a bid, and g awards the object to the high bidder at the price he named.

Direct revelation mechanism (DRM)

- **DRM**: the special IRM where $S_i = \Theta_i$ for every i – the only "action" is to report a type.
- Players directly report $\hat{\theta}_i$, truthfully or not. They can report any type, but nothing richer than a type.
- A **social choice function** (SCF) $f : \Theta \to X$ maps the reported profile into an outcome.
- The reading: a DRM collapses the two IRM steps into one, asking agents to hand the designer exactly the information she wants.

Direct revelation mechanism (cont.)

- **Shortcut**: skip the action stage.

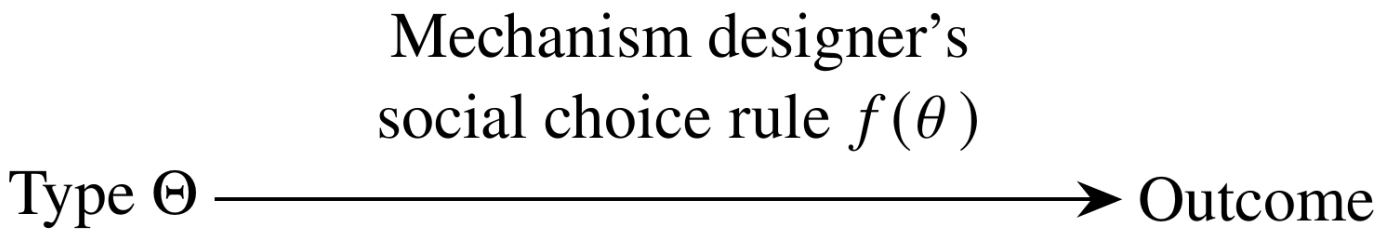


Figure 7.2. Direct revelation mechanism.

- Question: when do DRMs and IRMs yield the same equilibrium outcomes? Answer: see §7.5 (revelation principle).

Example 7.1 – First-price-style DRM

- Seller (agent 0) and two buyers. An outcome lists who gets the object plus the transfers.
- Buyer i's utility: $u_i(k, \theta_i) = \theta_i k_i + t_i$, where $k_i = 1$ if i gets the object – so he enjoys θ_i only when he wins.
- DRM: each buyer reports $\hat{\theta}_i$, then
  - **Allocation**: $k_i(\hat{\theta}) = 1$ if $\hat{\theta}_i \geq \hat{\theta}_j$ (ties to buyer 1) – the high reporter wins.
  - **Transfer**: $t_i(\hat{\theta}) = -\hat{\theta}_i k_i(\hat{\theta})$ – the winner pays *her own report*.
  - Seller collects $t_0 = \hat{\theta}_1 k_1 + \hat{\theta}_2 k_2$, the winner's payment.

Example 7.1 – First-price-style DRM (cont.)

- This DRM **mimics a sealed-bid first-price auction**: the report plays the role of the bid, and the winner pays what she reported.
- Intuitively, a buyer who reports her true value and wins pays her full value, leaving her nothing – so we should expect her to report *below* her value. We confirm this in §7.3.

Example 7.2 – Second-price-style DRM

- Same setup as Example 7.1, but transfer rule changes.
- Allocation: same – the winner is the highest reporter, $k_i(\hat{\theta}) = 1$ if $\hat{\theta}_i \geq \hat{\theta}_j$.
- **Transfer:** $t_i(\hat{\theta}) = -\hat{\theta}_j k_i(\hat{\theta})$ – the winner now pays her *rival's* report.
- So the winner pays the **second-highest reported valuation**, not her own.
- This mimics a Vickrey (second-price) auction.
- Spoiler: this DRM is DSIC – truthful reporting is optimal. The first-price-style DRM (Example 7.1) is not.

Why first-price differs from second-price as a DRM

- In Example 7.1, the winner pays *her own report*. In Example 7.2, the winner pays *the second-highest report*.
- Implication for incentives:
  - Example 7.1: reporting high $\Rightarrow$ win more often, but pay more if you win. **Trade-off**: agents shade their reports.
  - Example 7.2: reporting high $\Rightarrow$ win more often, but payment is determined by another bidder's report. No reason to shade!
- In §7.3 we'll see this means Example 7.2 is DSIC (truthful reporting is dominant) while Example 7.1 is not.

First vs second-price (cont.)

- Same allocation rule, different transfer rules, hence different incentives to report honestly.
- The lesson for design: the **payment rule**, not who wins, is what makes a mechanism truthful. Both examples give the object to the high reporter, yet only the second-price rule removes the incentive to shade.
- We will see this again with Groves and VCG: fix the efficient allocation, then engineer the transfers to deliver truth-telling.

Example 7.3 – Procurement DRM

- Same players and outcome set as Example 7.1, but the seller now buys a service and wants the **cheapest** provider.
- Allocation reverses: $k_i(\hat{\theta}) = 1$ if $\hat{\theta}_i \leq \hat{\theta}_j$ – the firm reporting the *lowest* cost wins the contract.
- Transfer mirrors Example 7.1 (pay the winner her own reported cost) or Example 7.2 (pay her the rival's cost), but with the *opposite sign*: the winning firm is paid, not charged.
- The reading: procurement is an auction run "in reverse," so every rule flips from sell-side to buy-side.

Example 7.4 – Funding a public project

- n citizens decide whether to build a bridge of cost $C > 0$: $k = 1$ build, $k = 0$ do not.
- Each $-t_i$ is the tax paid by citizen i. The project is built only if taxes cover the cost:

$$kC \leq -\sum_{i=1}^{n} t_i.$$

- Utility $u_i(k, t_i, \theta_i) = k\theta_i + t_i$: citizen i enjoys his valuation θ_i only if the bridge is built, and may pay a tax ($t_i < 0$).
- So this is the public-good analog of the auction: the "allocation" is a yes/no build decision shared by everyone.

Example 7.5 – Public-project DRM

- Now ask citizens to report their valuations $\hat{\theta}_i$. The build rule is

$$k(\hat{\theta}) = 1 \text{ iff } \sum_i \hat{\theta}_i \geq C,$$

  build exactly when reported total value covers the cost.
- Equal cost-sharing transfer:

$$t_i(\hat{\theta}) = -\frac{C}{n} k(\hat{\theta}),$$

  so each citizen pays C/n when built, and nothing otherwise.
- The reading: a simple, symmetric rule – but, as with the first-price DRM, we must still ask whether citizens report honestly (§7.3).

Section 7.3
Implementation and Incentive Compatibility

Testing Example 7.1 – First-price DRM is not BIC

- Assume $\theta_1, \theta_2 \sim U[0,1]$, independently.
- **Step 1. Fix the opponent at truth-telling** and write player 1's payoff from report $\hat{\theta}_1$:

$$\max_{\hat{\theta}_1} \underbrace{(\theta_1 - \hat{\theta}_1)}_{\text{margin if win}} \cdot \underbrace{\Pr(\theta_2 \leq \hat{\theta}_1)}_{\text{prob. of winning}}.$$

  If she wins she keeps $\theta_1 - \hat{\theta}_1$ (value minus own payment); she wins when her report tops player 2's.
- **Step 2. Use the uniform prior** so $\Pr(\theta_2 \leq \hat{\theta}_1) = \hat{\theta}_1$, reducing the objective to

$$\max_{\hat{\theta}_1} \theta_1 \hat{\theta}_1 - \hat{\theta}_1^2.$$

Testing Example 7.1 (cont.)

- **Step 3. Take the FOC** in $\hat{\theta}_1$:

$$\theta_1 - 2\hat{\theta}_1 = 0 \quad \Longrightarrow \quad \hat{\theta}_1 = \frac{\theta_1}{2}.$$

- **Step 4. Read the result.** Player 1 reports only *half* her true value: she **shades** her report to lower the price she pays if she wins.
- Since the optimal report is $\hat{\theta}_1 \neq \theta_1$, truth-telling is not a best response, so the SCF of Example 7.1 is **not BIC**.
- Intuitively, paying her own report makes honesty costly – every extra unit reported is an extra unit paid when she wins.

Player 2's analogous problem

- By symmetry, if player 1 reports truthfully, player 2's optimal report is also $\hat{\theta}_2 = \theta_2/2$.
- So **both** players shade their reports by half.
- In equilibrium of this DRM, the bidder with the highest *true* valuation still wins (since $\hat{\theta}_1 = \theta_1/2$ and $\hat{\theta}_2 = \theta_2/2$ preserve ranking).

Player 2's analogous problem (cont.)

- But each player pays only half her valuation – not the bidder's true valuation.
- Hence the SCF of Example 7.1 (as a DRM) fails BIC – truthful reporting is not the BNE.
- Implication: cannot use Example 7.1 as a truthful direct mechanism. Must redesign or use Example 7.2's transfer rule.

Bayesian Incentive Compatibility (BIC)

- **Definition:** SCF $f(\cdot)$ is BIC if truth-telling is a BNE of the DRM:

$$s_i^*(\theta_i) = \theta_i \quad \forall \theta_i \in \Theta_i.$$

- Formally: for every player i and every type θ_i:

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i) \mid \theta_i] \geq E_{\theta_{-i}}\left[u_i(f(\theta_i', \theta_{-i}), \theta_i) \mid \theta_i\right].$$

- Truth-telling beats any other report **in expectation**, given that others tell the truth too.
- Note: BIC allows misreporting to be best for some realizations of θ_{-i}, as long as truth-telling wins on average.

BIC vs. DSIC – One-line comparison

- **Dominant-strategy IC (DSIC).** Truth-telling is optimal for each type *whatever* the other players report – a very strong requirement.
- **Bayesian IC (BIC).** Truth-telling is optimal *on average*, given each type's beliefs about the other players' types – a weaker requirement.
- Every DSIC mechanism is BIC. The reverse is not true: BIC opens up mechanisms that DSIC forbids.

BIC vs. DSIC (cont.)

- **Design implication.** If the setting supports Bayesian beliefs (well-defined prior distributions), the designer can implement more outcomes under BIC than under DSIC – more slack, more revenue.
- **Robustness caveat.** DSIC mechanisms work even when players do not share the designer's prior; BIC mechanisms may fail if beliefs are mis-specified.

BIC – Example 7.2 satisfies it

- In Example 7.2 (second-price-style DRM), suppose player 2 reports truthfully.
- Player 1's payoff from reporting $\hat{\theta}_1$ when she wins ($\hat{\theta}_1 \geq \theta_2$): $\theta_1 - \theta_2$.
- Truthful reporting $\hat{\theta}_1 = \theta_1$:
  - If $\theta_1 > \theta_2$: she wins and gets $\theta_1 - \theta_2 > 0$.
  - If $\theta_1 < \theta_2$: she loses and gets 0.

BIC – Example 7.2 satisfies it (cont.)

- Over-reporting ($\hat{\theta}_1 > \theta_1$): could win when $\theta_2 \in (\theta_1, \hat{\theta}_1)$ – but then pays $\theta_2 > \theta_1$ for negative net payoff.
- Under-reporting ($\hat{\theta}_1 < \theta_1$): could lose when $\theta_2 \in (\hat{\theta}_1, \theta_1)$ – foregoes positive margin.
- Hence truth-telling is optimal, so Example 7.2 is BIC.
- **Numeric check.** Let $\theta_1 = 0.6$. Truthful: wins whenever $\theta_2 < 0.6$ and pays θ_2. Over-reporting to 0.8 changes the price in no state (she still pays θ_2), but adds wins for $\theta_2 \in (0.6, 0.8)$ where she pays $\theta_2 > 0.6$ – a loss. Honesty wins.

Dominant-Strategy Incentive Compatibility (DSIC)

- **Definition:** SCF $f(\cdot)$ is DSIC if truth-telling is a *dominant strategy* in the DRM.
- Formally: for every player i, every type θ_i, every report $\theta_i' \neq \theta_i$, and every action profile s_{-i} of opponents:

$$u_i(f(\theta_i, s_{-i}), \theta_i) \geq u_i(f(\theta_i', s_{-i}), \theta_i).$$

- Truth-telling optimal **regardless** of others' strategies.

DSIC (cont.)

- Stronger than BIC: DSIC implies BIC, but not conversely – averaging a condition that holds state-by-state preserves it, not the reverse.
- Robustness: a DSIC mechanism does not require agents to know the prior or trust others to play truthfully, so it keeps working if beliefs are mis-specified.
- Example 7.2 is DSIC: the winner's price is set by a rival's report, so her own report never affects what she pays – honesty is best whatever others do.

Section 7.4
Implementation in IRMs

7.4.1 Implementation in dominant strategies

- Mechanism M implements SCF $f(\cdot)$ **in dominant strategies** if there is a weakly dominant strategy profile s^* such that

$$g(s^*(\theta)) = f(\theta) \quad \forall \theta \in \Theta.$$

- **Example 7.6 (SPA implements Example 7.2):**
  - In the second-price auction, bidding $s_i(\theta_i) = \theta_i$ is weakly dominant.
  - Object goes to highest bidder; winner pays second-highest bid.
  - Same outcome as Example 7.2's SCF.
- SPA = IRM that implements Example 7.2's SCF in dominant strategies.

7.4.2 Implementation in BNE

- Mechanism M implements SCF $f(\cdot)$ **in BNE** if there is a BNE strategy profile s^* such that

$$g(s^*(\theta)) = f(\theta) \quad \forall \theta \in \Theta.$$

- **Example 7.7 (SPA implements Example 7.2 in BNE):**
  - Recall: SCF of Example 7.2 is BIC, so $s_i^*(\theta_i) = \theta_i$ (truthful) is a BNE of the DRM.
  - Use a **second-price auction** as the IRM: bidder with the highest valuation wins and pays the second-highest submitted bid.
  - Bidding one's own type is a BNE (in fact, weakly dominant).
  - Hence the SPA "implements" the SCF of Example 7.2 in BNE.
- Implementation is weaker than DSIC implementation – depends on common beliefs.

Section 7.5
The Revelation Principle

Figure 7.3 – The revelation principle

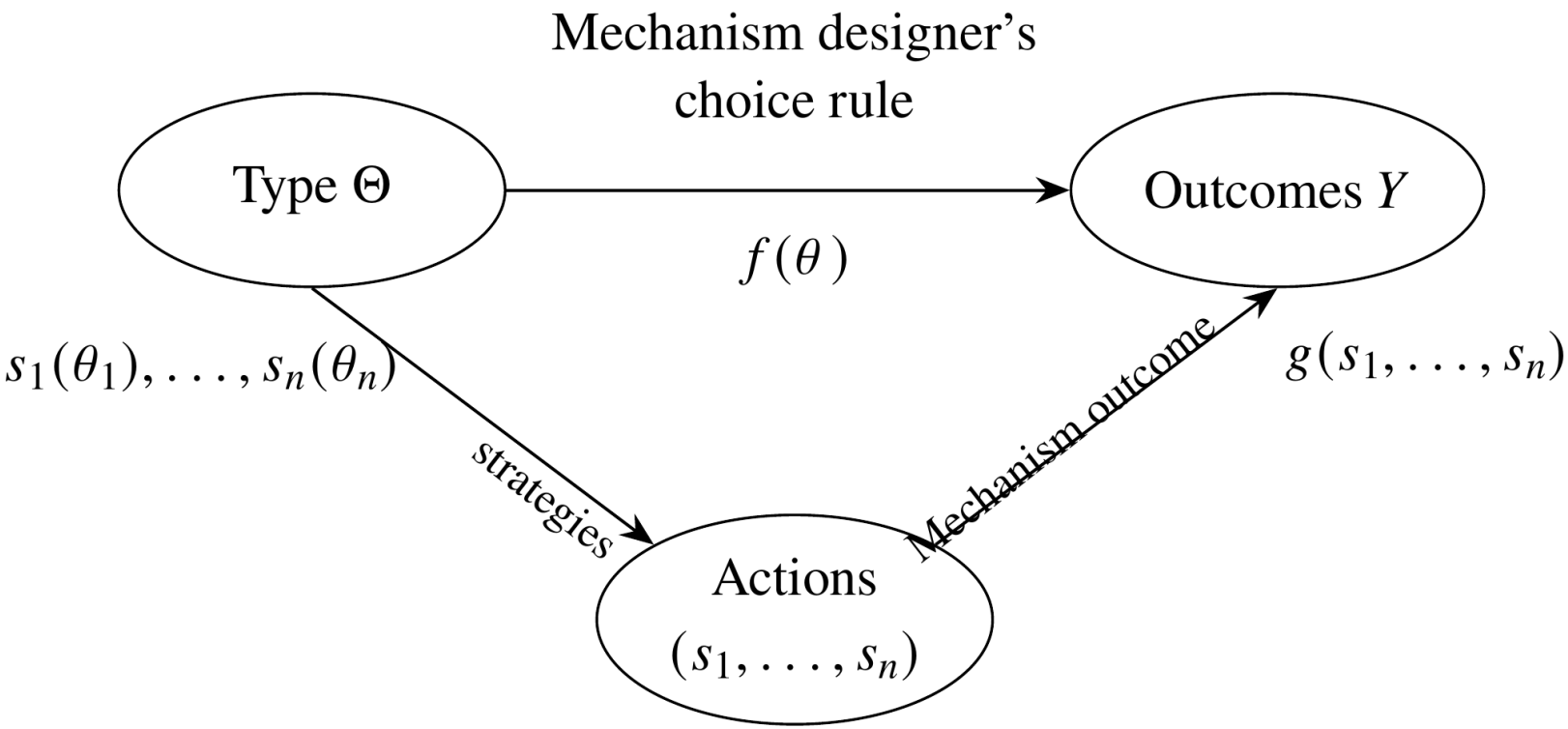


Figure 7.3. The revelation principle. Top route: direct revelation mechanism. Bottom route: indirect revelation through actions.

7.5.1 Revelation principle – BNE version

- **Theorem:** Mechanism M implements SCF $f(\cdot)$ in BNE $\Longleftrightarrow$ $f(\cdot)$ is BIC.
- **Proof sketch** ($\Leftarrow$): If f is BIC, then the DRM (restricting $S_i = \Theta_i$) implements f in BNE.
- ($\Rightarrow$): If M implements f in BNE with strategies $s_i^*(\theta_i)$, then truth-telling is best response in the DRM (substitute $s_i^*(\theta_i)$ for the report).
- **Combined inequality:**

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i) \mid \theta_i] \geq E_{\theta_{-i}}\left[u_i(f(\theta_i', \theta_{-i}), \theta_i) \mid \theta_i\right],$$

which is exactly BIC.

The revelation principle – What it buys us

- Any outcome that can be implemented by *some* mechanism in Bayes-Nash equilibrium can also be implemented by a *direct* mechanism in which truth-telling is a BNE.
- **Practical payoff.** Instead of searching over the space of all mechanisms (auctions, menus, dynamic games), the designer can restrict attention to *direct truthful mechanisms*.
  Massively simplifies the problem.
- **Two ingredients need to hold.**
  - Types are private but distributed according to a common prior.
  - Players are Bayesian rational.

The revelation principle (cont.)

- **What it does not guarantee.** That the direct mechanism is *simple* to describe, or that it survives implementation frictions (side deals, coalitions, mistakes).
- Applications: auction design, matching, voting rules, network formation – all rely on this simplification.

7.5.2 Revelation principle – DSIC version

- **Theorem:** Mechanism M implements SCF $f(\cdot)$ in dominant-strategy equilibrium $\Longleftrightarrow$ $f(\cdot)$ is DSIC.
- Same idea, but now without expectations: condition holds for all θ_{-i} and all s_{-i}.
- **Combined inequality:**

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq u_i(f(\theta_i', \theta_{-i}), \theta_i),$$

which is DSIC.

Revelation principle – Practical implication

- **Summary equivalence:**

$$M \text{ implements } f \text{ in BNE} \iff f \text{ is BIC},$$
$$M \text{ implements } f \text{ in DSE} \iff f \text{ is DSIC}.$$

- **WLOG focus on DRMs**: any outcome achievable by some IRM can be achieved by a truthful DRM.
- This is a **huge** simplification:
  - No need to enumerate all possible game forms.
  - Just optimize over the (much smaller) class of BIC or DSIC SCFs.

Revelation principle – Practical implication (cont.)

- **Caveats.** The principle fixes a *particular* equilibrium: the truthful DRM has truth-telling as *an* equilibrium, but may have others too.
- It is also silent on robustness – the equivalent direct mechanism can be fragile to collusion, side payments, or agents who are not fully rational.
- So "restrict attention to truthful DRMs" is a statement about *what is achievable*, not a promise that the direct mechanism is the one we would run in practice.

Section 7.6
Efficiency

7.6.1 Allocative efficiency

- Quasilinear preferences:

$$v_i(k, \theta_i) = u_i(k, \theta_i) + w_i + t_i.$$

- **Allocative efficiency:** SCF $f(\theta) = (k(\theta), t_1, \ldots, t_n)$ is allocatively efficient if for every θ:

$$k(\theta) \in \arg\max_{k \in K} \sum_{i \in N} u_i(k, \theta_i).$$

- Allocation maximizes **aggregate** payoff for every type profile.

Allocative efficiency (cont.)

- Analogous to Pareto efficiency, applied to allocation rules.
- Note: this is about *allocation* $k(\theta)$ – transfers are separate.

Example 7.8 – An efficient SCF

- Two agents, $\Theta_i = \{20, 60\}$, project cost 50, tax 25 per agent if built.
- Allocation: $k = 1$ iff at least one type is high.
- Transfer: -25 each if $k = 1$.
- Total welfare under $k = 1$:
  - $(20, 20)$: -10 – worse than $k = 0$.
  - $(20, 60)$: 30.
  - $(60, 20)$: 30.
  - $(60, 60)$: 70.
- Rule correctly chooses $k = 0$ at $(20, 20)$, $k = 1$ elsewhere – **allocatively efficient.**

Example 7.9 – An inefficient SCF

- Same setup as 7.8 but cost is zero ($t_i = 0$ for all θ).
- Same allocation rule (build iff at least one type is high).
- But now at $(20, 20)$: not building gives 0; building gives $20 + 20 = 40$.
- Hence the rule *should* build at $(20, 20)$ – but it doesn't.
- Allocation rule is **not** allocatively efficient.
- **Lesson:** efficiency depends on cost structure as well as the allocation rule.

Section 7.7
Common Mechanisms – Groves and VCG

Why we care about specific mechanisms

- The revelation principle tells us **WHAT** a mechanism can achieve; but it doesn't tell us **HOW** to design a practical mechanism.
- In practice, designers want:
  - **Simplicity**: easy to explain to participants.
  - **Robustness**: works regardless of agents' beliefs.
  - **Efficiency**: maximizes total surplus.
  - **Voluntary participation**: agents willing to join.
  - **Budget balance**: mechanism self-financing.
- Groves and VCG mechanisms are famous because they achieve efficiency + DSIC in dominant strategies (robust to beliefs).

Why we care (cont.)

- Next: develop these mechanisms in detail.

7.7.1 The Groves mechanism

- **Theorem (Groves):** If SCF $f(\theta) = (k(\theta), t_1, \ldots, t_n)$ is allocatively efficient, then f is DSIC if transfers have the form:

$$t_i(\theta_i, \theta_{-i}) = \sum_{j \neq i} u_j(k(\theta), \theta_j) + h_i(\theta_{-i}),$$

  where $h_i(\cdot)$ is any function independent of θ_i.
- Intuition: i's transfer reflects the externality her report imposes on others.
- Truth-telling becomes a dominant strategy because i effectively internalizes the social surplus.

Groves – Proof sketch

- **Step 1. Suppose not.** By contradiction, let f be allocatively efficient and Groves but *not* DSIC, so some agent i can profitably misreport $\theta_i' \neq \theta_i$:

$$u_i(f(\theta_i', \theta_{-i}), \theta_i) > u_i(f(\theta_i, \theta_{-i}), \theta_i).$$

- **Step 2. Insert the Groves transfers** and cancel the term $h_i(\theta_{-i})$, which is the same on both sides because it does not depend on i's report:

$$\sum_j u_j(k(\theta_i', \theta_{-i}), \theta_j) > \sum_j u_j(k(\theta_i, \theta_{-i}), \theta_j).$$

- The left side adds i's own utility to the others', so both sides are the *full* social surplus under each allocation.

Groves – Proof sketch (cont.)

- **Step 3. Reach the contradiction.** The inequality says the allocation at the misreport yields *more* aggregate surplus than the one at the truth – impossible, since $k(\theta)$ maximizes that sum by allocative efficiency.
- Hence no profitable misreport exists: Groves transfers plus allocative efficiency imply DSIC. $\square$
- Intuitively, the Groves transfer hands each agent the others' total utility, so his private objective becomes the social objective – telling the truth maximizes both at once.

7.7.2 The VCG mechanism

- **VCG** (Vickrey-Clarke-Groves) is a special Groves mechanism with:

$$h_i(\theta_{-i}) = -\sum_{j \neq i} u_j(k_{-i}(\theta_{-i}), \theta_j),$$

  where k_{-i} is the efficient allocation *ignoring* player i.
- Transfer becomes:

$$t_i(\theta) = \sum_{j \neq i} u_j(k(\theta), \theta_j) - \sum_{j \neq i} u_j(k_{-i}(\theta_{-i}), \theta_j).$$

- Often called the **Clarke pivotal mechanism**.
- Each player pays exactly the externality her presence imposes on others.

VCG – Properties and weaknesses

- **Properties:**
  - Allocatively efficient.
  - DSIC: truth-telling is a dominant strategy.
  - Each agent pays exactly her externality on others.
- **Weaknesses:**
  - **Not budget-balanced** – typically runs a surplus.
  - Can fail **ex-post PC** in some settings (Example 7.12).
  - Vulnerable to collusion among bidders.
  - Can be computationally complex for combinatorial allocations.
- Used in spectrum auctions, online ad auctions (Google's "second-price" design adopted VCG ideas).

Example 7.10 – VCG with one item, 5 bidders

- Valuations: $\theta_1 = 20, \theta_2 = 15, \theta_3 = 12, \theta_4 = 10, \theta_5 = 6$.
- Object goes to bidder 1 (highest).
- **Bidder 1's transfer:**

$$\begin{aligned} t_1 &= \sum_{j \neq 1} u_j(k(\theta), \theta_j) - \sum_{j \neq 1} u_j(k_{-1}(\theta_{-1}), \theta_j) \\ &= (0 + 0 + 0 + 0) - (15 + 0 + 0 + 0) = -15. \end{aligned}$$

Example 7.10 (cont.)

- Bidder 1 pays 15 – the second-highest valuation. This is the externality she imposes: bidder 2 would have won (and enjoyed 15) had she been absent.
- **Other bidders' transfers:**

$$t_i = (20 + 0 + 0 + 0) - (20 + 0 + 0 + 0) = 0 \quad \forall i \neq 1.$$

  A loser changes nobody's allocation, so he imposes no externality and pays nothing.
- VCG = second-price auction here. Sanity check: the designer collects 15 and distributes none of it, so the mechanism runs a **surplus** – VCG is not budget-balanced.

Example 7.11 – VCG with 3 identical items

- Same valuations; now 3 items available; each bidder wants 1.
- Efficient allocation: items go to bidders 1, 2, 3.
- **Bidder 1's transfer:**

$$\begin{aligned} t_1 &= \sum_{j \neq 1} u_j(k(\theta), \theta_j) - \sum_{j \neq 1} u_j(k_{-1}(\theta_{-1}), \theta_j) \\ &= (15 + 12) - (15 + 12 + 10) = -10. \end{aligned}$$

- Bidder 1 pays 10 – the externality on bidder 4, who would have gotten an item if bidder 1 were absent.
- Generalizes VCG to multi-unit settings; same logic.

Section 7.8
Participation Constraints

Three flavors of participation

- Let $\overline{u}_i(\theta_i)$ be agent i's reservation utility (his outside option). A participation constraint (PC) asks that joining beats it. The stage at which we ask sets how strict it is.
- **Ex-ante PC** – before i knows his own type:

$$E_\theta[u_i(g(\theta), \theta_i)] \geq E_{\theta_i}[\overline{u}_i(\theta_i)].$$

  He weighs average utility against his average outside option.
- **Interim PC** – after i knows θ_i but not θ_{-i}:

$$E_{\theta_{-i}}[u_i(g(\theta), \theta_i) \mid \theta_i] \geq \overline{u}_i(\theta_i) \quad \forall \theta_i.$$

  Every type, averaging only over rivals, must want to take part.

Three flavors of participation (cont.)

- **Ex-post PC** – after the full state θ is known:

$$u_i(g(\theta), \theta_i) \geq \overline{u}_i(\theta_i) \quad \forall \theta.$$

  No averaging: the agent must be glad he joined in *every* realization.

- The reading: the three differ only in **what the agent knows** when he decides, which fixes how much uncertainty the constraint must survive.
- More information at the decision stage means fewer states to hide behind, so ex-post is the hardest to satisfy and ex-ante the easiest – the hierarchy we turn to next.

The participation hierarchy

- For any SCF:

$$\text{Ex-post PC} \Rightarrow \text{Interim PC} \Rightarrow \text{Ex-ante PC}.$$

- Ex-post is the most demanding (must hold for *every* realization).
- Ex-ante is the easiest to satisfy (need only hold on average over all uncertainty).

Hierarchy of participation constraints (cont.)

- **Example** – construction firm bidding:
  - Ex-ante: before knowing own costs.
  - Interim: after costs revealed; before opponents' costs.
  - Ex-post: after everything revealed.

Example 7.12 – VCG can violate ex-post PC

- Two agents, $\Theta_i = \{20, 60\}$, project cost \$50, allocation k^* from Example 7.8 (build unless both types are low).
- Valuation: $v_i(k^*, \theta_i) = k^*(\theta_i - 25)$ – each builder bears a \$25 share of the cost.
- Reservation utility $\overline{u}_i = 0$, so ex-post PC asks $v_i(k^*(\theta), \theta_i) \geq 0$ for *every* profile.

Example 7.12 – The failing profile

- **Step 1. Pick the profile** $(\theta_1, \theta_2) = (20, 60)$. Since one type is high, the project is built, $k^* = 1$.
- **Step 2. Evaluate the low type's ex-post payoff:**

$$v_1(k^*(20, 60), 20) = 1 \cdot (20 - 25) = -5 < 0.$$

- **Step 3. Read it.** The low type's \$25 cost share exceeds her \$20 valuation, so she is worse off participating – **ex-post PC fails**.
- The lesson: DSIC and allocative efficiency together still do *not* guarantee that every type wants to take part after the state is realized.

7.8.2 When does Clarke satisfy ex-post PC?

- Clarke (VCG with zero h_i) satisfies ex-post PC if:
  1. Reservation utility is zero: $\overline{u}_i = 0$.
  2. **Choice-set monotonicity**: feasible outcome set X weakly grows in N.
  3. **No negative externality**: $v_i(k^*_{-i}(\theta_{-i}), \theta_i) \geq 0$ for every type.
- **Intuition:** since each agent pays the externality her presence imposes (a positive quantity), her utility cannot drop below what she would get if she were absent – which is non-negative by assumption.
- Auctions, public-good problems with zero outside option, etc. typically satisfy these properties.

Section 7.9
Linear Utility, Myerson, and Revenue Equivalence

The dAGVA mechanism

- d'Aspremont, Gérard-Varet, Arrow mechanism: solves VCG's budget-balance problem.
- Transfer rule:

$$t_i(\hat{\theta}) = \varepsilon(\hat{\theta}_i) - \frac{1}{n-1}\sum_{j \neq i} \varepsilon(\hat{\theta}_j),$$

  where $\varepsilon(\hat{\theta}_i) = E_{\theta_{-i}}[\sum_{j \neq i} u_j(k(\hat{\theta}_i, \theta_{-i}), \theta_j)]$ is the expected externality.
- **Properties:**
  - BIC (not DSIC).
  - Budget-balanced: $\sum_i t_i = 0$.
  - Allocatively efficient.

The dAGVA mechanism (cont.)

- **Reading the transfer.** The first term pays agent i the expected externality his report creates on others; the second charges him an equal share of everyone else's expected externality.
- **Why budget balances.** Summing over i, each agent's expected externality is paid out once and charged back once across the others, so the terms cancel and $\sum_i t_i = 0$.
- **Why only BIC.** The transfers use *expected* externalities (averaged over θ_{-i}), so truth-telling is optimal on average but not report-by-report – a weaker incentive concept than DSIC.
- Trade-off: give up dominant strategies (DSIC) to buy exact budget balance, keeping allocative efficiency.

Linear utility – Setup

- Special case: $v_i(k, \theta_i) = \theta_i v_i(k)$ (separable in type and allocation).
- Types drawn from $[\underline{\theta}_i, \bar{\theta}_i]$, iid with positive density.
- Useful definitions:
  - $\bar{t}_i(\hat{\theta}_i) \equiv E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})]$ – expected transfer.
  - $\bar{v}_i(\hat{\theta}_i) \equiv E_{\theta_{-i}}[v_i(\hat{\theta}_i, \theta_{-i})]$ – expected valuation.
  - $u_i(\theta_i) = \theta_i \bar{v}_i(\theta_i) + \bar{t}_i(\theta_i)$ – expected utility under truth-telling.

7.9.1 Myerson's characterization theorem

- **Theorem (Myerson 1981):** In a linear environment, SCF $f(\cdot)$ is BIC if and only if for every agent i:
  1. $\bar{v}_i(\theta_i)$ is **weakly increasing** in θ_i (monotonicity), and
  2. Expected utility has the **envelope representation**:

$$u_i(\theta_i) = u_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \bar{v}_i(s)\, ds \quad \forall \theta_i \in \Theta_i.$$

- Reading condition 2: a type's utility is the lowest type's utility plus the accumulated expected valuation, so utility rises in θ_i at rate $\bar{v}_i(\theta_i)$.

Myerson's theorem – How to use it

- The theorem turns "design a BIC mechanism" into a **two-step recipe**:
  1. Choose an allocation rule $k(\theta)$ that makes each $\bar{v}_i(\theta_i)$ weakly increasing (monotonicity).
  2. Pick the transfer rule $\bar{t}_i(\theta_i)$ so that expected utility matches the envelope formula.
- Why this helps: monotonicity is a condition on the *allocation alone*, and once it holds the transfers are pinned down up to the constant $u_i(\underline{\theta}_i)$.
- Economically, incentive compatibility ties the payment schedule to the allocation schedule – we are no longer free to choose them separately.

7.9.3 Revenue equivalence theorem

- **Theorem (Myerson):** If two auction formats with risk-neutral bidders (linear utility) and independent valuations have:
  1. The same allocation rule $(\bar{y}_1(\theta_1), \ldots, \bar{y}_n(\theta_n))$, and
  2. The same expected utility for the lowest-type bidder $u_i(\underline{\theta}_i)$,

  then they yield the **same expected revenue** for the seller.
- Implication: with the same allocation, first-price and second-price auctions earn the same revenue.
- **Powerful**: avoid computing equilibrium bids in each format separately.

Proof sketch – Revenue equivalence

- Seller's revenue from bidder i:

$$E[-\bar{t}_i(\theta_i)] = \int_{\underline{\theta}_i}^{\bar{\theta}_i} \underbrace{\bar{y}_i(\theta_i)}_{\text{Term } B} \underbrace{\left[\theta_i - \frac{1-\Phi_i(\theta_i)}{\phi_i(\theta_i)}\right]}_{\text{Term } C} \phi_i(\theta_i) d\theta_i - \underbrace{u_i(\underline{\theta}_i)}_{\text{Term } D}.$$

- Total expected revenue: $\sum_i E[-\bar{t}_i(\theta_i)]$ – depends on:
  - Allocation $\bar{y}_i(\cdot)$ (Term B).
  - Distribution-dependent term $\theta - (1-\Phi)/\phi$ (Term C).
  - $u_i(\underline{\theta}_i)$ (Term D).
- Same allocation $\Rightarrow$ same B; same lowest-type utility $\Rightarrow$ same D; C is determined by primitives. Hence revenue coincides. $\square$

Revenue equivalence – Reading the result

- Two auctions with the same allocation rule and the same expected payment for the lowest type deliver the same expected revenue to the seller.
- **Why this is striking.** First-price and second-price auctions look completely different, yet with i.i.d. private values and risk-neutral bidders they yield the same expected revenue.
- **Practical implication.** Auction *format* is often a red herring; the allocation rule and participation threshold do the real work.
- **When it breaks:** risk-averse bidders, budget constraints, correlated values, or asymmetric distributions – each violates a hypothesis and can favor one format over the other.

Example 7.13 – FPA vs SPA

- Recall: Example 7.1 = first-price-style DRM; Example 7.2 = second-price-style DRM.
- Both:
  - Allocate the object to the bidder with the highest valuation.
  - Yield 0 utility to the bidder with the lowest valuation.
- By revenue equivalence: **both auction formats generate the same expected revenue**.

Example 7.13 – FPA vs SPA (cont.)

- With $\theta_i \sim U[0,1]$ and $n \geq 2$ bidders, the expected revenue equals $\dfrac{n-1}{n+1}$.
- Extends to all-pay auctions, auctions with reserve prices, auctions with entry fees, etc. – as long as allocation and lowest-type utility coincide.
- Also extends to any k-th price auction with $k \geq 3$ by verifying the same two properties (see chapter 6 of Choi and Muñoz-Garcia, 2024).

Section 7.10
Optimal Bayesian Mechanism

Principal-agent with continuous types

- Agent: $u(e, t_1, \theta) = t_1 + \theta g(e)$, with $g' > 0$, $g'' > 0$, $g(0) = 0$ – effort is costly and increasingly so.
- Type $\theta \in [\underline{\theta}, \bar{\theta}]$ with $\bar{\theta} \leq 0$. Because $\theta g(e)$ enters with $\theta < 0$, a *lower* (more negative) θ means a *larger* disutility of effort.
- Assume the monotone-hazard condition: $\theta - \dfrac{1 - \Phi(\theta)}{\phi(\theta)}$ is weakly increasing in θ, so the virtual valuation is well behaved (no ironing needed).

Principal-agent with continuous types (cont.)

- Principal: $u_0 = x(e) + t_0$, with $x' > 0$, $x'' < 0$ – effort raises profit at a decreasing rate.
- Budget balance ties the two transfers: $t_0 = -t_1$, so every dollar the agent receives comes from the principal.
- By the revelation principle we may design a **DRM** in which the agent truthfully reports θ, rather than searching over menus of contracts (the IRM of Chapter 6).
- This is exactly the Chapter 6 screening problem, now cast as a truthful direct mechanism with a continuum of types.

7.10.2 Symmetric information benchmark

- **Step 1. Write the problem.** With θ observed, the principal faces only the PC:

$$\max_{e,t_1} \; x(e(\theta)) - t_1(\theta) \quad \text{s.t.} \quad t_1(\theta) + \theta g(e(\theta)) \geq \bar{u}.$$

- **Step 2. Bind the PC.** She never overpays, so $t_1 = \bar{u} - \theta g(e)$; substituting removes t_1 from the objective.
- **Step 3. Differentiate in** e, giving the first-best rule

$$\underbrace{x'(e^*(\theta))}_{\text{Marginal profit}} = \underbrace{-\theta g'(e^*(\theta))}_{\text{Marginal cost}}.$$

- The reading: push effort until the extra profit equals the extra wage needed to compensate the agent – the standard MRS condition.

Figure 7.4 – SI equilibrium effort

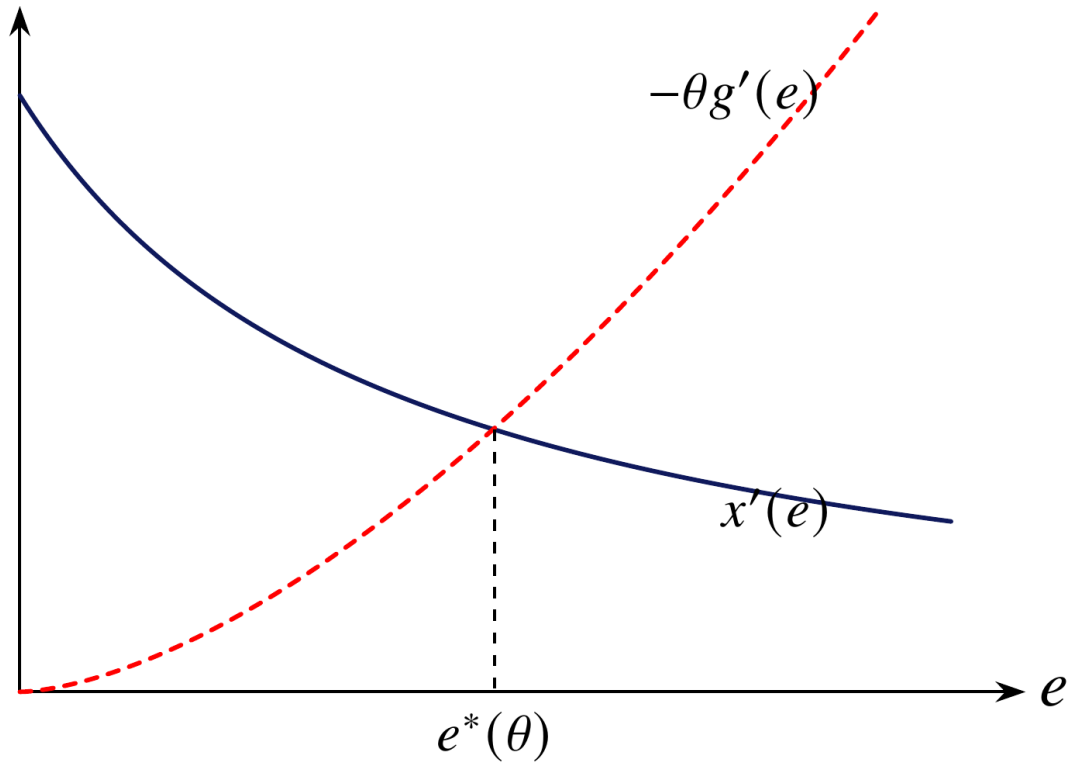


Figure 7.4. Equilibrium effort under symmetric information: marginal profit $x'(e)$ (decreasing) crosses marginal cost $-\theta g'(e)$ (increasing) at $e^*(\theta)$.

7.10.3 Asymmetric information – Setup

- **Step 1. State the problem.** The principal now maximizes expected profit over feasible SCFs:

$$\max_{e(\cdot),t_1(\cdot)} \; E_\theta[x(e(\theta)) - t_1(\theta)] \quad \text{s.t.} \quad f(\cdot) \in F^* \text{ (BIC + PC)}.$$

- **Step 2. Change variable** to the agent's utility $U_1(\theta) = t_1(\theta) + \theta g(e(\theta))$, i.e., $t_1 = U_1 - \theta g(e)$.
- **Step 3. Substitute** to remove t_1:

$$\max_{e(\cdot),U_1(\cdot)} \; E_\theta[x(e(\theta)) - U_1(\theta) + \theta g(e(\theta))].$$

The agent's rent U_1 now appears explicitly as a cost to the principal.

Asymmetric information setup (cont.)

- **Express BIC via Myerson:**
  1. $e(\theta)$ weakly increasing in θ (monotonicity).
  2. $U_1(\theta) = U_1(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} g(e(s))ds$.
- **PC**: $U_1(\theta) \geq \bar{u}\ \forall\theta$.

Simplifying the principal's problem

- **Step 1. Reduce the PC.** Condition 2 makes $U_1(\theta)$ increasing, so PC holds for all θ iff it binds at the *lowest* type: $U_1(\underline{\theta}) = \bar{u}$.
- **Step 2. Substitute and integrate by parts.** Insert $U_1(\theta)$ from condition 2 into the objective and integrate the rent term by parts.
- **Step 3. Collect** $(U_1(\underline{\theta}) = \bar{u})$:

$$\max_{e(\cdot)} \int_{\underline{\theta}}^{\bar{\theta}} \left[x(e(\theta)) + \left(\theta - \frac{1 - \Phi(\theta)}{\phi(\theta)} \right) g(e(\theta)) \right] \phi(\theta) d\theta - \bar{u},$$

  subject to monotonicity of $e(\theta)$.
- The bracket is marginal profit plus a **virtual** marginal cost: the extra term $-\frac{1-\Phi}{\phi}$ is the information rent conceded to higher types.

Optimal-mechanism FOC

- Differentiate pointwise with respect to e:

$$\underbrace{x'(e(\theta))}_{\text{Marginal profit}} = \underbrace{-\left(\theta - \frac{1-\Phi(\theta)}{\phi(\theta)}\right) g'(e(\theta))}_{\text{Marginal cost}}.$$

- Same LHS as SI; RHS now includes the **virtual valuation** $J(\theta) \equiv \theta - \dfrac{1-\Phi(\theta)}{\phi(\theta)}$.
- Extra term $-\dfrac{1-\Phi(\theta)}{\phi(\theta)} < 0$ shifts marginal cost *up* relative to SI.
- Therefore equilibrium effort is **lower** under AI:

$$e^{AI}(\theta) < e^{SI}(\theta) \quad \forall \theta < \bar{\theta}.$$

Figure 7.5 – Comparison SI vs AI

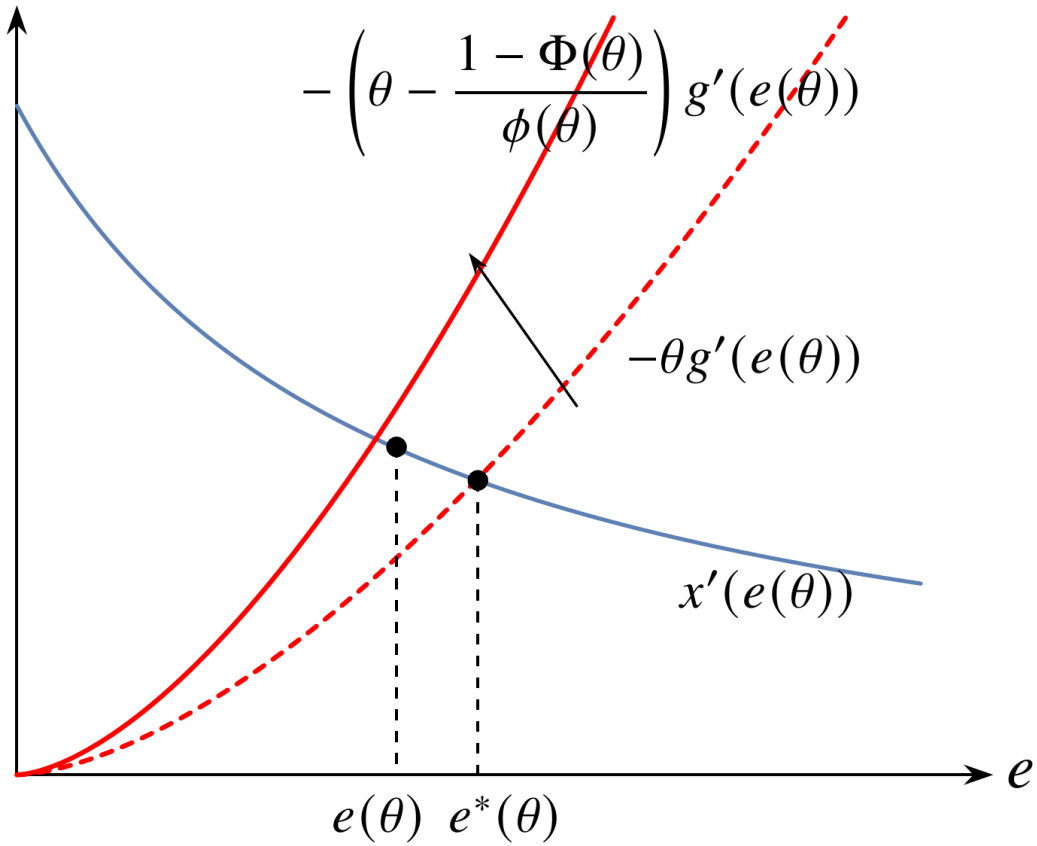


Figure 7.5. Under SI, the marginal cost curve is $-\theta g'(e)$. Under AI, it shifts up to $-(\theta - (1-\Phi)/\phi)g'(e)$ – giving $e^{AI} < e^{SI}$.

7.10.4 No distortion at the top – Again!

- At $\theta = \bar{\theta}$: $\Phi(\bar{\theta}) = 1 \Rightarrow$

$$\bar{\theta} - \frac{1 - \Phi(\bar{\theta})}{\phi(\bar{\theta})} = \bar{\theta}.$$

- FOC simplifies to $x'(e(\bar{\theta})) = -\bar{\theta} g'(e(\bar{\theta}))$ – same as SI.
- **No distortion at the top**: most efficient type's effort is undistorted.
- **Downward distortion** for all other types $\theta < \bar{\theta}$.
- Identical to Chapter 6 results – general principle of adverse-selection problems.

7.10.5 Monotonicity of effort – Verification

- Write FOC as $x'(e(\theta)) + J(\theta)g'(e(\theta)) = 0$.
- Differentiate implicitly w.r.t. θ:

$$e'(\theta) = -\frac{J'(\theta)g'(e(\theta))}{x''(e(\theta)) + J(\theta)g''(e(\theta))}.$$

- Sign analysis:
  - Numerator: $J' > 0$ (by assumption), $g' > 0$. Sign: $+$.
  - Denominator: $x'' < 0$, $J < 0$, $g'' > 0$. Sign: $-$.

Monotonicity of effort (cont.)

- Hence $e'(\theta) > 0$ – monotonicity holds.
- Confirms Myerson's BIC condition (1).

Example 7.14 – Uniform types numerical

- **Step 1. Distribution.** Types $\theta \sim U[-1, 0]$, so $\Phi(\theta) = 1 + \theta$ and $\phi(\theta) = 1$.
- **Step 2. Virtual valuation.**

$$J(\theta) = \theta - \frac{1 - (1 + \theta)}{1} = 2\theta.$$

- **Step 3. Functional forms.** Profit $x(e) = \log e$, cost $g(e) = 2e$, so $x'(e) = 1/e$ and $g'(e) = 2$.
- **Step 4. Symmetric info.** FOC $\frac{1}{e} = -\theta \cdot 2$, hence $e^{SI}(\theta) = -\frac{1}{2\theta}$.

Example 7.14 – Uniform types numerical (cont.)

- **Step 5. Asymmetric info.** Replace θ by $J(\theta) = 2\theta$: FOC $\frac{1}{e} = -J(\theta) \cdot 2 = -4\theta$, hence $e^{AI}(\theta) = -\frac{1}{4\theta}$.
- **Step 6. Compare.** Since $\frac{1}{4} < \frac{1}{2}$, we get $e^{AI}(\theta) < e^{SI}(\theta)$ for every $\theta < 0$.
- **Numeric sanity check.** At $\theta = -\frac{1}{2}$: $e^{SI} = -\frac{1}{2(-1/2)} = 1$ while $e^{AI} = -\frac{1}{4(-1/2)} = \frac{1}{2}$ – effort is halved under asymmetric information.
- Intuitively, the virtual valuation inflates the marginal cost of effort, so the principal buys less effort to hold down the rents she pays higher types.

Where the $J(\theta)$ term comes from

- Virtual valuation:

$$J(\theta) = \theta - \frac{1 - F(\theta)}{f(\theta)}.$$

- Captures the principal's marginal benefit from raising effort adjusted for information rents paid to higher types.
- Total expected information rent cost from increasing effort at θ:

$$[\theta f(\theta) - (1 - F(\theta))]g'(e(\theta)).$$

- Dividing by $f(\theta)$ gives the per-density cost:

$$\left(\theta - \frac{1 - F(\theta)}{f(\theta)}\right) g'(e(\theta)) = J(\theta)g'(e(\theta)).$$

- Hence the FOC has this term in it.

When is virtual valuation increasing?

- $J(\theta)$ is weakly increasing iff the **hazard rate**

$$\frac{f(\theta)}{1 - F(\theta)}$$

  is weakly increasing.
- Common distributions that satisfy this:
  - Uniform: $J(\theta) = 2\theta - 1$.
  - Exponential: $J(\theta) = \theta - 1/\lambda$.
  - $F(\theta) = \theta^{\alpha}$ with $\alpha \geq 1$: $J' > 0$.
- When J has decreasing segments (e.g., $\alpha < 1$ in $F = \theta^{\alpha}$), need "**ironing**" techniques (Bolton-Dewatripont).

Virtual valuation (cont.)

- **Example 7.15.** With $F(\theta) = \theta^{1/3}$, the virtual valuation is $J(\theta) = 4\theta - 3\theta^{2/3}$.
- Differentiating, J falls then rises, reaching a minimum at $\theta = \frac{1}{8}$ – so J is *not* monotone over the whole support.
- When J dips like this, the unconstrained rule would ask effort to fall then rise, violating monotonicity; the fix is **ironing** – flatten J over the offending interval before optimizing.

Section 7.11
Summary

Summary – Model and Implementation

- Mechanisms model: players, types, quasilinear utilities, mechanism designer.
- **IRM** (action sets + outcome function) vs. **DRM** (direct type reports).
- Incentive compatibility:
  - **BIC**: truth-telling is a BNE.
  - **DSIC**: truth-telling is a dominant strategy. Stronger.

Summary – Model and Implementation (cont.)

- **Revelation principle**: WLOG focus on DRMs.
  - M implements f in BNE $\Leftrightarrow$ f is BIC.
  - M implements f in DSE $\Leftrightarrow$ f is DSIC.
- Allocative efficiency: $k(\theta)$ maximizes aggregate utility.

Summary – Mechanisms and PCs

- **Groves** mechanism:

$$t_i = \sum_{j \neq i} u_j(k(\theta), \theta_j) + h_i(\theta_{-i}).$$

  Combined with allocative efficiency $\Rightarrow$ DSIC.
- **VCG / Clarke**: $h_i = -\sum_{j \neq i} u_j(k_{-i}(\theta_{-i}), \theta_j)$. Each agent pays her externality.
- Second-price auction = VCG with one item.
- Three flavors of PC (ex-ante $\subset$ interim $\subset$ ex-post).
- Clarke satisfies ex-post PC under three conditions (reservation utility 0, choice-set monotonicity, no negative externality).

Summary – Linear utility and optimal mechanism

- **Myerson's theorem**: In linear environments, BIC requires monotone $\bar{v}_i(\theta_i)$ and the envelope representation of $u_i(\theta_i)$.
- **Revenue equivalence**: same allocation + same lowest-type utility $\Rightarrow$ same expected revenue across auction formats.
- **Optimal Bayesian mechanism**: FOC

$$x'(e(\theta)) = -J(\theta)g'(e(\theta)), \quad J(\theta) = \theta - \frac{1 - F(\theta)}{f(\theta)}.$$

- No distortion at the top; downward distortion everywhere else.
- Virtual valuation $J(\theta)$ encodes the information-rent cost.

Common pitfalls in mechanism design problems

- **Confusing BIC and DSIC**: DSIC is strictly stronger and more robust.
- **Forgetting allocative efficiency**: Groves transfer alone isn't enough – must also have allocatively efficient $k(\theta)$.
- **Skipping ex-post PC check**: VCG can fail it (Example 7.12); important in voluntary settings.

Common pitfalls (cont.)

- **Misinterpreting** $h_i(\theta_{-i})$: any function of others' types works for Groves – VCG is just one choice.
- **Treating BIC as IC**: BIC requires expectations over θ_{-i}, given truthful behavior by others. Not the same as "truth is best response to any strategies of others."
- **Ignoring virtual valuation issues**: when J is not monotone, need ironing.
- **Forgetting budget balance**: VCG runs surpluses; can't be implemented without an outside funder unless we relax balance.

Mechanism design vs. contract theory

- **Contract theory** (Chs. 2–6): bilateral interactions between principal and agent.
- **Mechanism design** (Ch. 7): one designer, multiple agents – broader.
- Relationship:
  - Principal-agent problem (Ch. 6) = mechanism design with one agent.
  - Auction theory = mechanism design with multiple agents.
  - Public-good provision = mechanism design with collective decisions.

MD vs. contract theory (cont.)

- Same constraints (PC, IC) but generalized.
- Same trade-offs (risk vs. incentives, efficiency vs. rents).
- Mechanism design is the **unifying language** for thinking about any interaction involving private information and choice of rules.

Common student mistakes to avoid (Chapter 7)

- **Confusing DSIC with BIC.** DSIC needs truth-telling to be optimal for *every* report by the others; BIC only needs it in expectation.
- **Applying revenue equivalence outside its hypotheses.** The theorem needs i.i.d. private values, risk-neutrality, and no budget constraints. Drop any of these and formats diverge.
- **Forgetting participation constraints.** A mechanism is useless if no type wants to participate; IR is not optional.

Common student mistakes (Chapter 7) (cont.)

- **Assuming truth-telling means unique equilibrium.** It means truth-telling *is* an equilibrium, not the only one; the revelation principle is silent about uniqueness.
- **Treating mechanism design as pure theory.** Every result here has been used in FCC spectrum auctions, kidney exchange, school choice, and platform pricing.

Key takeaways from Chapter 7

1. Mechanism design = "reverse game theory." Designer chooses the rules.
2. Two ways to model: **IRMs** (actions) vs. **DRMs** (types). By the revelation principle, focus on DRMs.
3. **BIC** (Bayesian) and **DSIC** (dominant strategy) characterize when truth-telling is optimal.

Key takeaways (cont.)

1. **Groves** and **VCG** mechanisms achieve allocative efficiency + DSIC, but may violate ex-post PC.
2. **Myerson's theorem** characterizes BIC SCFs in linear environments via monotone allocation + envelope formula.
3. **Revenue equivalence**: auctions with same allocation and same lowest-type utility yield same expected revenue.
4. **Virtual valuations** drive the optimal Bayesian mechanism: no distortion at top, downward distortion below.

Real-world mechanism design applications

- **Spectrum auctions** (FCC, since 1994):
  - Multi-billion-dollar wireless spectrum sold via SMRA / clock auctions.
  - Combinatorial auctions for related licenses.
- **Online ads** (Google, Bing, Facebook):
  - Search and display ads use generalized second-price auctions.
  - Real-time auctions with billions of impressions per day.
- **Kidney exchange** (Roth-Sonmez):
  - Matching markets to enable kidney transplants among incompatible donor-patient pairs.

Real-world MD applications (cont.)

- **School choice** (Boston, NYC): deferred-acceptance mechanisms.
- **Carbon emissions** (EU ETS, RGGI): auction-based cap-and-trade.
- Nobel Prize 2007: Hurwicz, Maskin, Myerson. Nobel 2012: Roth, Shapley. Nobel 2020: Milgrom, Wilson.

Extensions to the basic framework

- **Multidimensional types**: types in higher dimensions create much harder mechanism design problems (correlated valuations, multiple items).
- **Risk-averse agents**: contracts must also share risk optimally (combining Chapters 2–4 with this chapter).
- **Collusion**: auction designs must be robust to bidder collusion.

Extensions to the basic framework (cont.)

- **Limited commitment**: principal cannot commit not to use revealed information against agents.
- **Dynamic mechanisms**: types and decisions over time (repeated auctions, dynamic pricing).
- **Behavioral mechanisms**: design that accounts for bounded rationality, present bias, etc.
- **Algorithmic mechanism design** (intersection with computer science): mechanisms that are computationally tractable.

Impossibility results

- **Hurwicz (1972):** No mechanism can achieve all three: allocative efficiency, individual rationality, and budget balance with private information.
- **Myerson-Satterthwaite (1983):** In bilateral bargaining with two-sided private information, no efficient mechanism is incentive-compatible, individually rational, and budget-balanced.
- **Gibbard-Satterthwaite (1973-1975):** In settings without quasilinear utility, DSIC mechanisms must be dictatorial.

Impossibility results (cont.)

- **Implication:** mechanism designers face fundamental trade-offs – they cannot achieve every desirable property simultaneously.
- Common compromises:
  - Sacrifice budget balance (VCG) for DSIC + efficiency.
  - Sacrifice DSIC (dAGVA) for BIC + budget balance + efficiency.
  - Sacrifice ex-post PC (Example 7.12) for BIC + efficiency.

Impossibility – What it means for practice

- Myerson-Satterthwaite: no bilateral trading mechanism can be efficient, individually rational, budget-balanced, and Bayesian incentive-compatible all at once.
- Arrow: no non-dictatorial voting rule respects universal domain, Pareto, and independence of irrelevant alternatives on more than two alternatives.
- Gibbard-Satterthwaite: any DSIC voting rule with three or more alternatives is either dictatorial or has a restricted range.

Impossibility – Practice (cont.)

- **Reading these results.** They set the outer boundary of what mechanism design can achieve; every successful mechanism relaxes one of the axioms (subsidies for budget balance, weakened efficiency, partial anonymity).
- **Design lesson.** Don't ask for all-in-one perfection. Pick the two or three properties that matter most for the specific context and accept the trade-offs.

Mathematical toolbox – A recap

- Tools developed across the book:
  - **Lagrangian**: constrained optimization (Chs. 2-3).
  - **First-order approach**: replace argmax with FOC (Ch. 4).
  - **Bayesian updating, perfect Bayesian equilibrium** (Ch. 5).
  - **Self-selection menus**, single-crossing (Ch. 6).
  - **Envelope theorem**, virtual valuations (Ch. 7).
- All these techniques appear throughout modern economic theory:
  - Public economics (taxation).
  - Industrial organization (regulation, auctions).
  - Labor economics (incentive contracts).
  - Macroeconomics (dynamic incentive problems).
  - Finance (security design).

Looking back at the book

- **Chapter 1**: foundational concepts.
- **Chapters 2–4**: moral hazard from discrete to continuous effort.
- **Chapter 5**: adverse selection in markets (lemons, signaling, screening).

Looking back at the book (cont.)

- **Chapter 6**: principal-agent with adverse selection.
- **Chapter 7**: mechanism design – the most general framework.
- Unifying themes:
  - Constraints (PC, IC, BIC, DSIC) shape contract / mechanism design.
  - Information asymmetry creates inefficiency relative to first-best.
  - Information rents accrue to the privately informed party.
  - No distortion at the top; distortion at the bottom.

Suggestions for further study

- **Textbooks:**
  - Bolton & Dewatripont, *Contract Theory* (2005) – advanced graduate text.
  - Salanié, *The Economics of Contracts* (2nd ed., 2005) – concise survey.
  - Laffont & Martimort, *The Theory of Incentives* (2002) – mechanism-design focus.
- **Foundational papers:**
  - Holmström (1979): Moral hazard and observability.
  - Myerson (1981): Optimal auction design.
  - Maskin (1999): Nash implementation.
  - Hart & Moore (1990): Property rights and the nature of the firm.
- **Game theory companion**: Espinola-Arredondo & Muñoz-Garcia, *Game Theory* (2023) for deeper PBE analysis.

Suggestions for further study (cont.)

- **Empirical applications**: Athey, Levin, Roth, etc.

Closing thoughts

- Contract theory and mechanism design are pillars of modern economic theory.
- The recurring lesson: **information asymmetries shape every contract and every market institution**.
- Solutions are subtle:
  - Wage spreads in labor contracts.
  - Deductibles in insurance.
  - Self-selection menus.
  - Truthful direct mechanisms.
  - Optimal allocation rules based on virtual valuations.

Closing thoughts (cont.)

- These ideas have been recognized by multiple Nobel Prizes – and they continue to evolve as new applications and theoretical insights emerge.
- Most importantly: they give you a **conceptual lens** to see the world differently.

End of Chapter 7

End of Book