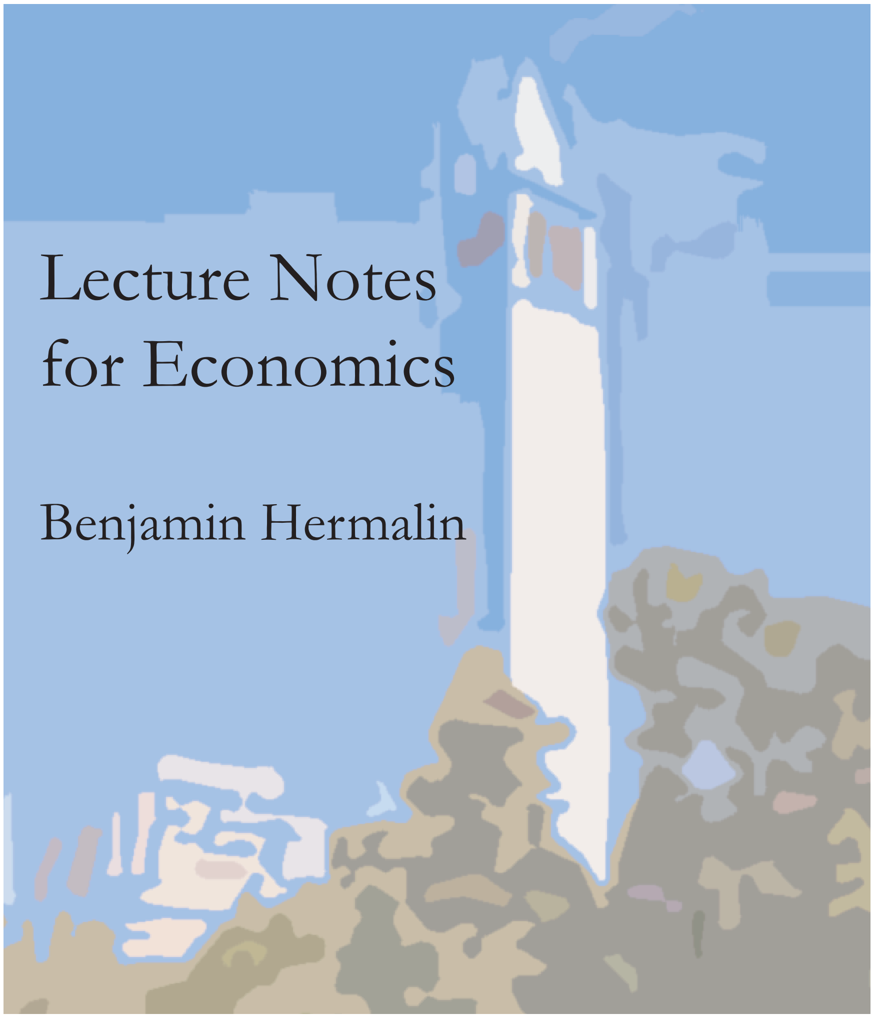

Lecture Notes for Economics

Benjamin Hermalin

Contents

List of Tables v

List of Figures vii

List of Examples ix

List of Propositions, Results, Rules, & Theorems xi

Preface xiii

I Pricing 1

Purpose 3

1 Buyers and Demand 5
- 1.1 Consumer Demand . . . 5
- 1.2 Firm Demand . . . 8
- 1.3 Demand Aggregation . . . 9
- 1.4 Additional Topics . . . 9

2 Linear Pricing 17
- 2.1 Elasticity and the Lerner Markup Rule . . . 19
- 2.2 Welfare Analysis . . . 21
- 2.3 An Application . . . 23

3 First-degree Price Discrimination 25
- 3.1 Two-Part Tariffs . . . 26

4 Third-degree Price Discrimination 31
- 4.1 The Notion of Type . . . 31
- 4.2 Characteristic-based Discrimination . . . 32
- 4.3 Welfare Considerations . . . 33
- 4.4 Arbitrage . . . 35
- 4.5 Capacity Constraints . . . 35

5 Second-degree Price Discrimination **39**
5.1 Quality Distortions 40
5.2 Quantity Discounts 43
5.3 A Graphical Approach to Quantity Discounts 46
5.4 A General Analysis 52

II Mechanism Design **59**

Purpose **61**

6 The Basics of Contractual Screening **63**

7 The Two-type Screening Model **65**
7.1 A simple two-type screening situation 65
7.2 Contracts under Incomplete Information 66

8 General Screening Framework **75**

9 The Standard Framework **81**
9.1 The Spence-Mirrlees Assumption 85
9.2 Characterizing the Incentive-Compatible Contracts 88
9.3 Optimization in the Standard Framework 91
9.4 The Retailer-Supplier Example Revisited 95

10 The Hidden-Knowledge Model **99**

11 Multiple Agents and Implementation **103**
11.1 Public Choice Problems 103
11.2 Mechanisms 104
11.3 Dominant-Strategy Mechanisms 105
11.4 The Revelation Principle 105
11.5 Groves-Clarke Mechanisms 107
11.6 Budget Balancing 110
11.7 Bayesian Mechanisms 111

III Hidden Action and Incentives **115**

Purpose **117**

12 The Moral-Hazard Setting **119**

13 Basic Two-Action Model **123**
13.1 The Two-action Model 123
13.2 The Optimal Incentive Contract 126
13.3 Two-outcome Model 128

13.4 Multiple-outcomes Model 135
13.5 Monotonicity of the Optimal Contract 139
13.6 Informativeness of the Performance Measure 141
13.7 Conclusions from the Two-action Model 142

14 General Framework 145

15 The Finite Model 151
15.1 The "Two-step" Approach 152
15.2 Properties of the Optimal Contract 161
15.3 A Continuous Performance Measure 165

16 Continuous Action Space 167
16.1 The First-order Approach with a Spanning Condition 168

Bibliography 173

Index 176

List of Tables

1 Notation xiv

8.1 The Retailer-Supplier Example in our General Notation 76

2 Examples of Moral-Hazard Problems 117

List of Figures

1.1 Consumer surplus 7

2.1 Marginal revenue 20
2.2 Deadweight loss 22

3.1 General analysis of 2-part tariffs 29

5.1 2nd-degree price discrimination via quality distortions 42
5.2 Ideal 3rd-degree price discrimination 46
5.3 Optimal quantity discount 48

7.1 Two-type screening model 72

9.1 The Spence-Mirrlees Assumption 87

13.1 Indifference curves 132
13.2 The set of feasible contracts. 133

List of Examples

Monopoly Pricing & Welfare 23
Welfare under 3rd-degree Price Discrimination 34
3rd-degree Price Discrimination with a Capacity Constraint 36
Airline Pricing (2nd-degree Price Discrimination) 43
Price discrimination via quantity discounts 50
Retailer-supplier example (part 1) 65
Retailer-supplier example (part 2) 95
Non-monotonic compensation 137

Propositions, Results, Rules, and Theorems

Proposition 1: Factor demand and total net benefit 8
Proposition 2: Aggregate CS equals sum of individual CS 9
Lemma 1: Integral relation between survival function and hazard rate . 12
Corollary 1: Extensions of Lemma 1 12
Lemma 2: Consumer surplus as an average benefit 13
Lemma 3: Concavity implies log concavity 14
Lemma 4: Log concave demand iff MHRP 15
Lemma 5: Log concave demand implies log concave revenue 15
Proposition 3: Sufficiency conditions for $MR = MC$ 18
Proposition 4: Sufficiency conditions for $MR = MC$ 18
Proposition 5: Monopoly pricing yields too little output 21
Proposition 6: Single consumer two-part tariff 26
Proposition 7: Properties of a 2-part tariff 27
Proposition 10: Optimal two-type mechanism 70
Proposition 11: The Revelation Principle 77
Proposition 12: The Taxation Principle 78
Lemma 7: Characterization of Spence-Mirrlees as cross partial 86
Proposition 13: Characterization of direct-revelation mechanisms 89
Lemma 8: Properties of Mechanism 93
Proposition 14: Solution of Standard Framework 93
Corollary 3: First corollary to Proposition 14 94
Corollary 4: Second corollary to Proposition 14 94
Proposition 16: Revelation Principle (multiple agents) 106
Proposition 17: Groves-Clarke mechanism 109
Proposition 18: Bayesian mechanism 112
Proposition 19: Monotonicity of compensation 139
Proposition 20: Summary of the 2-action Model 142
Proposition 21: Simple contracts are sufficient 147
Proposition 22: Individual rationality constraint binds 154
Proposition 23: Conditions for implementability 155
Proposition 24: Cost of implementing least-cost action 156
Proposition 25: Shifting support & implementation cost 157
Proposition 26: Risk neutrality & cost of implementation 158
Proposition 27: Selling the store 158
Corollary 5: corollary to Proposition 27 159
Proposition 28: Existence and uniqueness of optimal contract 160
Proposition 29: Hidden action costly 162

Proposition 30: Monotonicity of payments 163
Proposition 31: The role of information 164
Lemma 9: MLRP implies stochastic dominance 169
Corollary 6: Increasing $S(\cdot)$ implies global concavity 170

Preface

These lecture notes are intended to supplement the lectures and other materials for the first half of Economics 201B at the University of California, Berkeley.

A Word on Notation

Various typographic conventions are used to help guide you through these notes.

Notes in margins: *These denote important "takeaways."*

Text that *looks like this* is an important definition. On the screen or printed using a color printer, such definitions should appear blue.

As the author is an old Fortran programmer, variables i through n inclusive will generally stand for integer quantities, while the rest of the Roman alphabet will be continuous variables (*i.e.,* real numbers). The variable t will typically be used for time, which can be either discrete or continuous.

The ☡ symbol in the margin denotes a paragraph that may be hard to follow and, thus, requires particularly close attention (not that you should read any of these notes without paying close attention).

Vectors are typically represented through bold text (*e.g.,* $\mathbf{x}$ and $\mathbf{w}$ are vectors). Sometimes, when the focus is on the nth element of a vector $\mathbf{x}$, I will write $\mathbf{x} = (x_n, \mathbf{x}_{-n})$. The notation $-n$ indicates the subvector formed by removing the nth element (*i.e.,* all elements *except* the nth).

Derivatives of single-variable functions are typically denoted by primes (*e.g.,* $f'(x) = df(x)/dx$). Table 1 summarizes some of the other mathematical notation.

Symbol	Meaning
$\in$	Element of
$\forall$	For all
$\exists$	There exists
$\cdot$	Dot (vector) multiplication (*i.e.*, $\mathbf{x} \cdot \mathbf{y} = \sum_i x_i y_i$)
s.t.	Such that
a.e.	Almost everywhere (*i.e.*, true everywhere except, possibly, on a set of measure zero)
$\times_{n=1}^{N} \mathcal{X}_n$	The Cartesian product space formed from the spaces $\mathcal{X}_n$, $n = 1, \ldots, N$
$\mathbb{R}$	The set of real numbers. $\mathbb{R}^n = \times_{i=1}^{n} \mathbb{R}$ is the n-dimensional Euclidean space. $\mathbb{R}_+$ are the non-negative reals.
$\mathbb{E}$	The expectations operator. If X is a random variable, then $\mathbb{E}\{X\}$ is the expected value of X.
$\mathcal{X} \backslash \mathcal{Y}$	Set difference; that is, $\mathcal{X} \backslash \mathcal{Y}$ is the set of elements that are in $\mathcal{X}$ that are *not* also in $\mathcal{Y}$. Note $\mathcal{X} \backslash \mathcal{Y} = \mathcal{X} \bigcap \mathcal{Y}^c$, where $\mathcal{Y}^c$ is the complement of $\mathcal{Y}$.

Table 1: Some Mathematical Notation

Pricing

Purpose

If one had to distill economics down to a single-sentence description, one probably couldn't do better than describe economics as the study of how prices are and should be set. This portion of the *Lecture Notes* is primarily focused on the normative half of that sentence, how prices should be set, although I hope it also offers some positive insights as well.

Because I'm less concerned with how prices are set, these notes don't consider price setting by the Walrasian auctioneer or other competitive models. Nor is it concerned with pricing in oligopoly. Our attention will be exclusively on pricing by a single seller who is not constrained by competitive or strategic pressures (*e.g.,* a monopolist).

Now, one common way to price is to set a price, p, per unit of the good in question. So, for instance, I might charge \$10 per coffee mug. You can buy as many or as few coffee mugs as you wish at that price. The *revenue* I receive is \$10 times the number of mugs you purchase. Or, more generally, at price p per unit, the revenue from selling x units is px. Because px is the formula for a line through the origin with slope p, such pricing is called *linear pricing*.

If you think about it, you'll recognize that linear pricing is not the only type of pricing you see. Generically, pricing in which revenue is not a linear function of the amount sold is called *nonlinear pricing*.[1] Examples of nonlinear pricing would be if I gave a 10% discount if you purchased five or mugs (*e.g.,* revenue is \$$10x$ if $x < 5$ and \$$9x$ if $x \geq 5$). Of if I had a "buy one mug, get one free" promotion (*e.g.,* revenue is \$10 if $x = 1$ or 2, \$20 if $x = 3$ or 4, etc.). Or if I gave you a \$3-dollar gift with each purchase (*e.g.,* revenue is \$$10x-3$). Alternatively, the price per mug could depend on some other factor (*e.g.,* I offer a weekend discount or a senior-citizen discount). Or I could let you have mugs at \$5 per mug, but only if you buy at least \$50 worth of other merchandise from my store. Or I could pack 2 mugs in a box with a coffee maker and not allow you to buy mugs separately at all.

[1]Remember in mathematics a function is linear if $\alpha f(x_0) + \beta f(x_1) = f(\alpha x_0 + \beta x_1)$, where α and β are scalars. Note, then, that a linear function from $\mathbb{R}$ to $\mathbb{R}$ is linear only if it has the form $f(x) = Ax$.

Buyers and Demand | 1

A seller sets prices and buyers respond. To understand how they respond, we need to know what their objectives are. If they are consumers, the standard assumption is that they wish to maximize utility. If they are firms, the presumption is they wish to maximize profits.

Consumer Demand | 1.1

In the classic approach to deriving demand,[1] we maximize an individual's utility subject to a budget constraint; that is,

$$\max_{\mathbf{x}} u(\mathbf{x}) \quad \text{subject to } \mathbf{p} \cdot \mathbf{x} \leq I \,, \tag{1.1}$$

where $\mathbf{x}$ is an N-dimensional vector of goods, $\mathbf{p}$ is the N-dimensional price vector, and I is income. Solving this problem yields the individual's demand curve for each good n, $x_n^*(p_n; \mathbf{p}_{-n}, I)$ (where the subscript $-n$ indicates that it is the $N-1$-dimensional subvector of prices other than the price of the nth good). Unfortunately, while this analysis is fine for studying linear pricing, it is hard to utilize for nonlinear pricing: The literature on nonlinear pricing typically requires that the *inverse* of individual demand also represent the consumer's marginal benefit curve (*i.e.,* the benefit the consumer derives from a marginal unit of the good).[2] Unless there are no income effects, this is *not* a feature of demand curves.

For this reason, we will limit attention to *quasi-linear utility*. Assume that each individual j purchases two goods. The amount of the one in which we're interested (*i.e.,* the one whose pricing we're studying) is denoted x. The amount of the other good is denoted y. We can (and will) normalize the price of good y to 1. We can, if we like, consider y to be the amount of consumption other than of good x. The utility function is assumed to have the form

$$u(x, y) = v(x) + y \,. \tag{1.2}$$

[1] As set forth, for instance, in Mas-Colell et al. (1995) or Varian (1992).

[2] Recall that a demand function is a function from price to quantity; that is, for any given price, it tells us the amount the consumer wishes to purchase. Because demand curves slope down—Giffen goods don't exist—the demand function is invertible. Its inverse, which is a function from quantity to price, tells us for any quantity the price at which the consumer would be willing to purchase exactly that quantity (assuming linear pricing).

With two goods, we can maximize utility by first solving the constraint in (1.1) for y, yielding $y = I - px$ (recall y's price is 1), and then substituting that into the utility function to get an unconstrained maximization problem:[3]

$$\max_x v(x) - px + I\,. \tag{1.3}$$

Solving, we have the first-order condition

$$v'(x) = p\,. \tag{1.4}$$

Observe (1.4) also defines the inverse demand curve and, as desired, we have marginal benefit of x equal to inverse demand. If we define $P(x)$ to be the inverse demand curve, then we have

$$\begin{aligned}\int_0^x P(t)dt &= \int_0^x v'(t)dt \\ &= v(x) - v(0)\,.\end{aligned}$$

The second line is the gain in utility—*benefit*—the consumer obtains from acquiring x. Substituting this back into (1.3) we see that utility at the utility-maximizing quantity is equal to

$$\int_0^x P(t)dt - xP(x) \tag{1.5}$$

plus an additive constant, which we are free to ignore. In other words, utility, up to an additive constant, equals the area below the inverse demand curve and above the price of x. See Figure 1.1. You may also recall that (1.5) is the formula for *consumer surplus (CS)*.

Summary 1 *Given quasi-linear utility, the individual's inverse demand curve for a good is his or her marginal benefit for that good. Moreover, his or her utility at the utility-maximizing quantity equals (to an affine transformation) his or her consumer surplus (i.e., the area below inverse demand and above the price).*

Another way to think about this is to consider the first unit the individual purchases. It provides him or her (approximate) benefit $v'(1)$ and costs him or her p. His or her surplus or profit is, thus, $v'(1) - p$. For the second unit the surplus is $v'(2) - p$. And so forth. Total surplus from x units, where $v'(x) = p$, is, therefore,

$$\sum_{t=1}^{x} \left(v'(t) - p\right)\;;$$

[3]Well, actually, we need to be careful; there is an implicit constraint that $y \geq 0$. In what follows, we assume that this constraint doesn't bind.

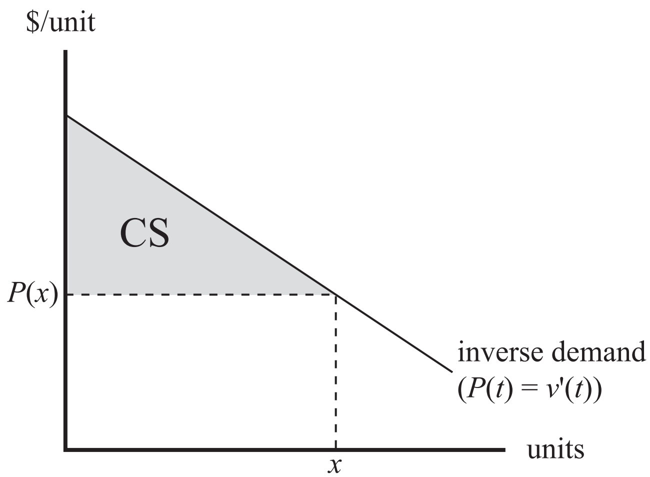

Figure 1.1: Consumer surplus (CS) at quantity x is the area beneath inverse demand curve ($P(t)$) and above inverse demand at x, $P(x)$.

or, passing to the continuum (*i.e.,* replacing the sum with an integral),

$$\int_0^x (v'(t) - p)\, dt = \int_0^x v'(t)dt - px = \int_0^x P(t)dt - px\,.$$

Yet another way to think about this is to recognize that the consumer wishes to maximize his or her surplus (or profit), which is total benefit, $v(x)$, minus his or her total expenditure (or cost), px. As always, the solution is found by equating marginal benefit, $v'(x)$, to marginal cost, p.

Bibliographic Note

One of the best treatments of the issues involved in measuring consumer surplus can be found in Chapter 10 of Varian (1992). This is a good place to go to get full details on the impact that income effects have on measures of consumer welfare.

Quasi-linear utility allows us to be correct in using consumer surplus as a measure of consumer welfare. But even if utility is *not* quasi-linear, the error from using consumer surplus instead of the correct measures, compensating or equivalent variation (see Chapter 10 of Varian), is quite small under assumptions that are reasonable for most goods. See Willig (1976). Hence, as a general rule, we can use consumer surplus as a welfare measure even when there's no reason to assume quasi-linear utility.

Firm Demand | 1.2

Consider a firm that produces $F(\mathbf{x})$ units of a good using inputs $\mathbf{x}$. Let the factor prices be $\mathbf{p}$ and let $R(\cdot)$ be the revenue function. Then the firm maximizes

$$R\bigl(F(\mathbf{x})\bigr) - \mathbf{p} \cdot \mathbf{x} \,. \tag{1.6}$$

The first-order condition with respect to input x_n is

$$R'\bigl(F(\mathbf{x})\bigr)\frac{\partial F}{\partial x_n} - p_n = 0 \,. \tag{1.7}$$

Let $\mathbf{x}^*[\mathbf{p}]$ denote the set of factor demands, which is found by solving the set of equations (1.7). Define the profit function as

$$\pi(\mathbf{p}) = R\bigl(F(\mathbf{x}^*[\mathbf{p}])\bigr) - \mathbf{p} \cdot \mathbf{x}^*[\mathbf{p}] \,.$$

Utilizing the envelope theorem, it follows that

$$\frac{\partial \pi}{\partial p_n} = -x_n^*(p_n; \mathbf{p}_{-n}) \,. \tag{1.8}$$

Consequently, integrating (1.8) with respect to the price of the nth factor, we have

$$-\int_{p_n}^{\infty} \frac{\partial \pi(t; \mathbf{p}_{-n})}{\partial p_n} dt = \int_{p_n}^{\infty} x_n^*(t; \mathbf{p}_{-n}) dt \,. \tag{1.9}$$

The right-hand side of (1.9) is just the area to the left of the factor demand curve that's above price p_n. Equivalently, it's the area below the *inverse* factor demand curve and above price p_n. The left-hand side is $\pi(p_n; \mathbf{p}_{-n}) - \pi(\infty; \mathbf{p}_{-n})$. The term $\pi(\infty; \mathbf{p}_{-n})$ is the firm's profit if it doesn't use the nth factor (which could be zero if production is impossible without the nth factor). Hence, the left-hand side is the increment in profits that comes from going from being unable to purchase the nth factor to being able to purchase it at price p_n. This establishes

Proposition 1 *The area beneath the factor demand curve and above a given price for that factor is the total net benefit that a firm enjoys from being able to purchase the factor at that given price.*

In other words—as we could with quasi-linear utility—we can use the "consumer" surplus that the firm gets from purchasing a factor at a given price as the value the firm places on having access to that factor at the given price.

Observation 1 *One might wonder why we have such a general result with factor demand, but we didn't with consumer demand. The answer is that with factor demands there are* no *income effects. Income effects are what keep consumer surplus from capturing the consumer's net benefit from access to a good at its prevailing price. Quasi-linear utility eliminates income effects, which allows us to treat consumer surplus as the right measure of value or welfare.*

Demand Aggregation 1.3

Typically, a seller sells to more than one buyer. For some forms of pricing it is useful to know total demand as a function of price.

Consider two individuals. If, at a price of \$3 per unit, individual one buys 4 units and individual two buys 7 units, then total or aggregate demand at \$3 per unit is 11 units. More generally, if we have J buyers indexed by j, each of whom has individual demand $x_j(p)$ as a function of price, p, then aggregate demand is $\sum_{j=1}^{J} x_j(p) \equiv X(p)$.

How does *aggregate* consumer surplus (*i.e.,* the area beneath aggregate demand and above price) relate to individual consumer surplus? To answer this, observe that we get the same area under demand and above price whether we integrate with respect to quantity or price. That is, if $x(p)$ is a demand function and $p(x)$ is the corresponding inverse demand, then

$$\int_0^x \big(p(t) - p(x)\big)dt = \int_p^\infty x(t)dt\,.$$

Consequently, if $CS(p)$ is aggregate consumer surplus and $cs_j(p)$ is buyer j's consumer surplus, then

$$\begin{aligned} CS(p) &= \int_p^\infty X(t)dt \\ &= \int_p^\infty \left(\sum_{j=1}^{J} x_j(t)\right) dt \\ &= \sum_{j=1}^{J} \left(\int_p^\infty x_j(t)dt\right) \\ &= \sum_{j=1}^{J} cs_j(p)\,; \end{aligned}$$

that is, we have:

Proposition 2 *Aggregate consumer surplus is the sum of individual consumer surplus.*

Additional Topics 1.4

A Continuum of Consumers

In many modeling situations, it is convenient to imagine a continuum of consumers (*e.g.,* think of each consumer having a unique "name," which is a real number in the interval $[0, 1]$; and think of all names as being used). Rather than thinking of the number of consumers—which would here be uncountably infinite—we think about their *measure*; that is, being rather loose, a function related to the length of the interval.

It might at first seem odd to model consumers as a continuum. One way to think about it, however, is the following. Suppose there are J consumers. Each consumer has demand

$$x(p) = \begin{cases} 1\,, & \text{if } p \leq v \\ 0\,, & \text{if } p > v \end{cases}\,, \tag{1.10}$$

where v is a number, the properties of which will be considered shortly. The demand function given by (1.10) is sometimes referred to as *L-shaped demand* because, were one to graph it, the curve would resemble a rotated L.[4] The demand function also corresponds to the situation in which the consumer wants at most one unit and is willing to pay up to v for it.

Assume, for each consumer, that v is a random draw from the interval $[v_0, v_1]$ according to the distribution function $F : \mathbb{R} \to [0, 1]$. Assume the draws are independent. Each consumer knows the realization of his or her v prior to making his or her purchase decision.

In this case, each consumer's *expected* demand is the probability that he or she wants the good at the given price; that is, the probability that his or her $v \geq p$. That probability is $1 - F(p) \equiv \Sigma(p)$. The function $\Sigma(\cdot)$ is known in probability theory as the *survival function*.[5] Aggregate expected demand is, therefore, $J\Sigma(p)$ (recall the consumers' valuations, v, are independently distributed).

Observe, mathematically, this demand function would be the equivalent of assuming that there are a continuum of consumers living on the interval $[v_0, v_1]$, each consumer corresponding to a fixed valuation, v. Assume further that the measure of consumers on an the interval between v and v' is $JF(v') - JF(v)$ or, equivalently, $J\Sigma(v) - J\Sigma(v')$. As before, consumers want at most one unit and they are willing to pay at most their valuation. Interpret demand at p as being the measure of consumers in $[p, v_1]$; that is,

$$J\Sigma(p) - J\Sigma(v_1) = J\Sigma(p)\,,$$

where the equality follows because $\Sigma(\cdot)$ at the end of the support of the function (the right-end of the interval) is zero. In other words, one can view the assumption of a continuum of consumers as being shorthand for a model with a finite number of consumers, each of whom has stochastic demand.

[4]Note, if one wants to get technical, the curve is really two line segments, one going from $(1, 0)$ to $(1, v)$ (using the usual orientation in which price is on the vertical axis), the other going from $(0, v)$ to $(0, \infty)$ (the latter interval is open on both ends). If, however, one looked at only the former interval and drew a horizontal line at the point of discontinuity, then one gets a rotated L.

[5]The name has an actuarial origin. If the random variable in question is age at death, then $\Sigma(\text{age})$ is the probability of surviving to *at least* that age. Admittedly, a more natural mnemonic for the survival function would be $S(\cdot)$; $S(p)$, however, is "reserved" in economics for the supply function.

A possible objection is that assuming a continuum of consumers on, say, $[v_0, v_1]$ with aggregate demand $J\Sigma(p)$ is a deterministic specification, whereas J consumers with random demand is a stochastic specification. In particular, there is variance in *realized* demand with the latter, but not the former.[6] In many contexts, though, this is not important because other assumptions make the price setter risk neutral.[7]

Demand as a Survival Function

If demand at zero price, $X(0)$, is finite, and if $\lim_{p\to\infty} X(p) = 0$, then any demand function is a multiplicative scalar of a survival function; that is,

$$X(p) = X(0)\Sigma(p)\,, \tag{1.11}$$

where $\Sigma(p) \equiv X(p)/X(0)$. To see that $\Sigma(\cdot)$ is a survival function on $\mathbb{R}_+$, observe that $\Sigma(0) = 1$, $\lim_{p\to\infty} \Sigma(p) = 0$, and, because demand curves slope down, $\Sigma(\cdot)$ in non-decreasing.

Assume $\Sigma(\cdot)$ is differentiable. Let $f(p) = -\Sigma'(p)$. The function $f(\cdot)$ is the *density function* associated with the survival function $\Sigma(\cdot)$ (or, equivalently, with the distribution function $1 - \Sigma(p) \equiv F(p)$). In demography or actuary science, an important concept is the *death rate* at a given age, which is the probability someone that age will die within the year. Treating time continuously, the death rate can be seen as the instantaneous probability of dying at time t conditional on having survived to t. (Why conditional? Because you can't die at time t unless you've lived to time t.) The *unconditional* probability of dying at t is $f(t)$, the probability of surviving to t is $\Sigma(t)$, hence the death rate is $f(t)/\Sigma(t)$. Outside of demographic and actuarial circles, the ratio $f(t)/\Sigma(t)$ is known as the *hazard rate*. Let $h(t)$ denote the hazard rate.

In terms of demand, observe

$$\frac{X'(p)}{X(p)} = \frac{X(0)\Sigma'(p)}{X(0)\Sigma(p)} = -h(p)\,. \tag{1.12}$$

In this context, $h(p)$ is the proportion of all units demanded at price p that will vanish if the price is increased by an arbitrarily small amount; that is, it's the hazard (death) rate of sales that will vanish (die) if the price is increased.

You may recall that the *price elasticity of demand*, ϵ, is minus one times the percentage change in demand per a one-percentage-point change in price.[8] In

[6] For example, if, $J = 2$, $[v_0, v_1] = [0, 1]$, and $F(v) = v$ on $[0, 1]$ (uniform distribution), then realized demand at $p \in (0, 1)$ is 0 with positive probability p^2, 1 with positive probability $2p(1 - p)$, and 2 with positive probability $(1 - p)^2$.

[7] It would be *incorrect* to appeal to the law of large numbers and act as if $J \to \infty$ means realized demand tends to $J\Sigma(p)$ according to some probability-theoretic convergence criterion. If, however, there is a *continuum* of consumers with identical and independently distributed demands, then it can be shown that realized demand is almost surely mean demand (see, *e.g.,* Uhlig, 1996).

[8] Whether one multiplies or not by -1 is a matter of taste; some authors (including this one on occasion) choose not to.

other words

$$\epsilon = -1 \times \left(\frac{\Delta X}{X} \times 100\%\right) \div \left(\frac{\Delta p}{p} \times 100\%\right) , \tag{1.13}$$

where Δ denotes "change in." If we pass to the continuum, we see that (1.13) can be reëxpressed as

$$\epsilon = -\frac{dX(p)}{dp} \times \frac{p}{X} = ph(p) , \tag{1.14}$$

where the last equality follows from (1.12). In other words, price elasticity places a monetary value on the proportion of sales lost from a price increase.

Other Relations

For future reference, let's consider some relations among demand, the associated hazard rate, and price elasticity. In what follows, we continue to assume $X(0)$ is finite and $\lim_{p\to\infty} X(p) = 0$.

Lemma 1 $\Sigma(p) = \exp\Big(-\int_0^p h(z)dz\Big)$.

Proof:

$$\frac{d\log\big(\Sigma(p)\big)}{dz} = \frac{-f(p)}{\Sigma(p)}$$
$$= -h(p) .$$

Solving the differential equation:

$$\log\big(\Sigma(p)\big) = -\int_0^p h(z)dz + \underbrace{\log\big(\Sigma(0)\big)}_{=0} .$$

The result follows by exponentiating both sides. ■

Immediate corollaries of Lemma 1 are

Corollary 1

(i) There exists a constant ξ such that

$$X(p) = \xi \exp\Big(-\int_0^p h(z)dz\Big) = \xi \exp\Big(-\int_0^p \frac{\epsilon(z)}{z}dz\Big)$$

(where $\epsilon(p)$ is the price elasticity of demand at price p).

(ii) Suppose $X_1(p) = \zeta X_2(p)$ for all p, where $X_i(\cdot)$ are demands and ζ is a positive constant. Let $\epsilon_i(\cdot)$ be the price-elasticity function associated with $X_i(\cdot)$. Then $\epsilon_1(p) = \epsilon_2(p)$ for all p.

(iii) Suppose $\epsilon(p) = \frac{pk}{p+p_0}$, where k and p_0 are positive constants. Then $X(p) = \xi(p+p_0)^{-k}$. Observe $\lim_{p_0\to 0}\epsilon(p) = k$ and $\lim_{p_0\to 0} X(p) = \xi p^{-k}$.

Exercise: Prove Corollary 1.

Exercise: Why in Corollary 1(iii) was it necessary to assume $p_0 > 0$?

Exercise: Verify by direct calculation (*i.e.*, using the formula $\epsilon = pX'(p)/X(p)$) that a demand function of the form $X(p) = \xi p^{-k}$ has a constant elasticity.

Lemma 2 *Consumer surplus at price p, $CS(p)$, satisfies*

$$CS(p) = X(0)\int_p^\infty (b-p)f(b)db\,, \tag{1.15}$$

where $f(\cdot)$ is the density function associated with $\Sigma(\cdot)$.

Proof: Recall the definition of consumer surplus is area to the left of demand and above price:

$$\begin{aligned}
CS(p) &= \int_p^\infty X(b)db \\
&= X(0)\int_p^\infty \Sigma(b)db \\
&= X(0)\left(b\Sigma(b)\Big|_p^\infty - \int_p^\infty b\Sigma'(b)db\right) && (1.16) \\
&= X(0)\left(-p\big(1-F(p)\big) + \int_p^\infty bf(b)db\right) && (1.17) \\
&= X(0)\int_p^\infty (b-p)f(b)db\,.
\end{aligned}$$

Observe (1.16) follows by integration by parts.[9] Expression (1.17) follows using the relations $\Sigma(p) = 1 - F(p)$ and $\Sigma'(p) = -f(p)$. ■

The integral in (1.15) is an expected value since $f(\cdot)$ is a density. If we thought of the benefit enjoyed from each unit of the good as a random variable with distribution $F(\cdot)$, then the integral in (1.15) is the expected net benefit (surplus) a given unit will yield a consumer (recalling, because he or she need not purchase, that the net benefit function is $\max\{b-p, 0\}$). The amount $X(0)$ can be thought of as the total number of units (given that at most $X(0)$ units trade). So the overall expression can be interpreted as expected (or average) surplus per item times the total number of items.

[9] To get technical, I've slipped in the assumption that $\lim_{b\to\infty} b\Sigma(b) = 0$. This is innocuous because, otherwise, consumer surplus would be infinite, which is both unrealistic and not interesting; the latter because, then, welfare would always be maximized (assuming finite costs) and there would, thus, be little to study.

Exercise: Prove that if $\lim_{b\to\infty} bX(b) > 0$, then consumer surplus at any price p is infinite. Hints: Suppose $\lim_{b\to\infty} bX(b) = L > 0$. Fix an $\eta \in (0, L)$. Show there is a $\bar{b}$ such that $X(b) \geq \frac{L-\eta}{b}$ for all $b \geq \bar{b}$. Does

$$\int_{\bar{b}}^{\infty} \frac{L-\eta}{b} db$$

converge? Show the answer implies $\int_p^{\infty} X(b)db$ does not converge (*i.e.,* is infinite).

A function, $g(\cdot)$, is *log concave* if $\log\big(g(\cdot)\big)$ is a concave function.

Lemma 3 *If $g(\cdot)$ is a positive concave function, then it is log concave.*[10]

Proof: If $g(\cdot)$ were twice differentiable, then the result would follow trivially using calculus (**Exercise:** do such a proof). More generally, let x_0 and x_1 be two points in the domain of $g(\cdot)$ and define $x_\lambda = \lambda x_0 + (1-\lambda)x_1$. The conclusion follows, by the definition of concavity, if we can show

$$\log\big(g(x_\lambda)\big) \geq \lambda \log\big(g(x_0)\big) + (1-\lambda)\log\big(g(x_1)\big) \tag{1.18}$$

for all $\lambda \in [0, 1]$. Because $\log(\cdot)$ is order preserving, (1.18) will hold if

$$g(x_\lambda) \geq g(x_0)^\lambda g(x_1)^{1-\lambda} . \tag{1.19}$$

Because $g(\cdot)$ is concave by assumption,

$$g(x_\lambda) \geq \lambda g(x_0) + (1-\lambda)g(x_1) .$$

Expression (1.19) will, therefore, hold if

$$\lambda g(x_0) + (1-\lambda)g(x_1) \geq g(x_0)^\lambda g(x_1)^{1-\lambda} . \tag{1.20}$$

Expression (1.20) follows from Pólya's generalization of the arithmetic mean-geometric mean inequality (see, *e.g.,* Steele, 2004, Chapter 2),[11] which states

[10] A stronger result can be established, namely that $r\big(s(\cdot)\big)$ is concave if both $r(\cdot)$ and $s(\cdot)$ are concave and $r(\cdot)$ is non-decreasing. This requires more work than is warranted given what we need.

[11] Pólya's generalization is readily proved. Define $A \equiv \sum_{i=1}^n \lambda_i a_i$. Observe that $x + 1 \leq e^x$ (the former is tangent to the latter at $x = 0$ and the latter is convex). Hence, $\frac{a_i}{A} \leq e^{\frac{a_i}{A}-1}$. Because both sides are positive, $\left(\frac{a_i}{A}\right)^{\lambda_i} \leq e^{\frac{\lambda_i a_i}{A}-\lambda_i}$. We therefore have

$$\prod_{i=1^n} \left(\frac{a_i}{A}\right)^{\lambda_i} \leq \prod_{i=1^n} e^{\frac{\lambda_i a_i}{A}-\lambda_i} = e^{\sum_i^n \left(\frac{\lambda_i a_i}{A}-\lambda_i\right)} .$$

The last term simplifies to 1, so we have

$$\frac{\prod_{i=1^n} a_i^{\lambda_i}}{A^{\sum_{i=1}^n \lambda_i}} \leq 1 .$$

Because the denominator on the left is just A, the result follows.

that if $a_1, \ldots, a_n \in \mathbb{R}_+$, $\lambda_1, \ldots, \lambda_n \in [0,1]$, and $\sum_{i=1}^n \lambda_i = 1$, then

$$\sum_{i=1}^n \lambda_i a_i \geq \prod_{i=1}^n a_i^{\lambda_i} \,. \tag{1.21}$$

■

Observe the converse of Lemma 3 need not hold. For instance, x^2 is log concave ($2\log(x)$ is clearly concave in x), but x^2 is not itself concave. In other words, log-concavity is a weaker requirement than concavity. As we will see, log concavity is often all we need for our analysis, so we gain a measure of generality by assuming log-concavity rather than concavity.

The hazard rate is monotone if it is either everywhere non-increasing or everywhere non-decreasing. The latter case is typically the more relevant (*e.g.,* the death rate for adults increases with age) and is a property of many familiar distributions (*e.g.,* the uniform, the normal, the logistic, among others). When a hazard rate is non-decreasing, we say it satisfies the *monotone hazard rate property*. This is sometimes abbreviated as MHRP.

Lemma 4 *A demand function is (strictly) log concave if and only if the corresponding hazard rate satisfies the (strict) monotone hazard rate property.*

Proof: It is sufficient for $X(\cdot)$ to be log concave that the first derivative of $\log\big(X(\cdot)\big)$—that is,

$$\frac{d\log\big(X(p)\big)}{dp} = \frac{X'(p)}{X(p)} = -h(p)$$

—be non-increasing (decreasing) in p. Clearly it will be if $h(\cdot)$ is non-decreasing (increasing). To prove necessity, assume $X(\cdot)$ is (strictly) log concave, then $-h(\cdot)$ must be non-increasing (decreasing), which is to say that $h(\cdot)$ is non-decreasing (increasing). ■

Lemma 5 *If demand is log concave, then so to is consumer expenditure under linear pricing; that is, $X(\cdot)$ log concave implies $pX(p)$ is log concave for all p.*

Proof: Given that $\log\big(pX(p)\big) \equiv \log(p) + \log\big(X(p)\big)$, the first derivative of $\log\big(pX(p)\big)$ is

$$\frac{1}{p} + \frac{X'(p)}{X(p)} \,.$$

The first term is clearly decreasing in p and the second term is decreasing in p by assumption. ■

Exercise: Suppose the hazard rate is a constant. Prove that $pX(p)$ is, therefore, everywhere strictly log concave.

Exercise: Prove that linear (affine) demand (*e.g.,* $X(p) = a - bp$, a and b positive constants) is log concave.

Exercise: Prove that if demand is log concave, then elasticity is increasing in price (*i.e.,* $\epsilon(\cdot)$ is an increasing function).

Exercise: Prove that if price elasticity is increasing in price, then consumer expenditure is log concave; that is, show $\epsilon(\cdot)$ increasing implies $pX(p)$ is log concave for all p.

Linear Pricing | 2

In this section, we consider a firm that sells all units at a constant price per unit. If p is that price and it sells x units, then its revenue is px. Such linear pricing is also called *simple monopoly pricing*.

Assume this firm incurs a cost of $C(x)$ to produce x units. Suppose, too, that the aggregate demand for its product is $X(p)$ and let $P(x)$ be the corresponding *inverse* demand function. Hence, the maximum price at which it can sell x units is $P(x)$, which generates revenue $xP(x)$. Let $R(x)$ denote the firm's revenue from selling x units; that is, $R(x) = xP(x)$. The firm's profit is revenue minus cost, $R(x) - C(x)$. The profit-maximizing amount to sell maximizes this difference.

If we assume

$$\exists \bar{x} \in [0, \infty) \text{ such that } R(x) - C(x) \leq 0 \; \forall x > \bar{x} \tag{LP1}$$

and

$$\forall x_0 \in \mathbb{R}_+ , \; x \to x_0 \; \Rightarrow R(x) - C(x) \to R(x_0) - C(x_0) \quad \text{(continuity)} , \tag{LP2}$$

then there must exist at least one profit-maximizing quantity.

If we strengthen (LP2) by assuming

$$R(\cdot) \text{ and } C(\cdot) \text{ are differentiable} , \tag{LP2$'$}$$

then

$$R'(x) - C'(x) = 0 \tag{2.1}$$

if x is profit maximizing; or, as it is sometimes written,

$$MR(x) = MC(x)$$

where MR denotes marginal revenue and MC denotes marginal cost.

In many applications, we would like (2.1) to be sufficient as well as necessary. For (2.1) to be sufficient, the maximum cannot occur at a corner. To rule out a corner solution assume

$$\exists \hat{x} \in \mathbb{R}_+ \text{ such that } R(\hat{x}) - C(\hat{x}) > 0 . \tag{LP3}$$

This expression says that positive profit is possible. Given that profit at the "corners," 0 and $\bar{x}$, is zero, this rules out a corner solution. That profit at $\bar{x}$ is zero follows by assumptions (LP1) and (LP2). That profit at 0 is zero follows because $C(0) \equiv 0$ (if one is not doing an activity, then one is not forgoing anything to do it) and there can be no revenue from selling nothing.

For (2.1) to be sufficient it must also be true that it does not define a minimum (or an inflection point) and that there not be more than one local maximum. The following propositions provide conditions sufficient to ensure those conditions are met. Both propositions make use of the fact that

$$MR(x) = P(x) + xP'(x)\,. \tag{2.2}$$

Proposition 3 *Maintain assumptions* LP1, LP2′, *and* LP3. *Assume demand, $X(\cdot)$, is log concave and cost, $C(\cdot)$, is at least weakly convex, then $MR(x) = MC(x)$ is sufficient as well as necessary for the profit-maximizing quantity.*

Proof: Because an interior maximum exists, we know that (2.1) must hold for at least one x. If it holds for only one x, then that x must be the global maximum. Using the fact that $\epsilon = -\frac{P(x)}{xP'(X)}$ and expression (2.2), we can write (2.1) as

$$P(x)\left(1 - \frac{1}{\epsilon}\right) - C'(x) = 0\,. \tag{2.3}$$

Solving (2.3) for $1/\epsilon$ and making the substitutions $p = P\bigl(X(p)\bigr)$ and $x = X(p)$, we have

$$\frac{1}{\epsilon(p)} = \frac{p - C'\bigl(X(p)\bigr)}{p}\,. \tag{2.4}$$

(For future reference, note that expression (2.4) is known as the *Lerner markup rule*.) We are done if we can show that only one p can solve (2.4). As noted, at least one p must solve (2.4). If the left-hand side of (2.4) is decreasing in p and the right-hand side increasing, then the p that solves (2.4) must be unique. Because $X(\cdot)$ is log concave, price elasticity of demand, $\epsilon(\cdot)$, is increasing (see the exercises following Lemma 5); hence, $1/\epsilon(p)$ is decreasing in p. To show the right-hand side of (2.4) is increasing consider p and p', $p > p'$. It is readily seen that (2.4) is greater evaluated at p rather than p' if

$$\frac{C'\bigl(X(p)\bigr)}{p} < \frac{C'\bigl(X(p')\bigr)}{p'}\,.$$

Given $p' < p$, that will be true if $C'\bigl(X(p)\bigr) \leq C'\bigl(X(p')\bigr)$. Because demand curves are non-increasing, $X(p) \leq X(p')$. Because $C(\cdot)$ weakly convex implies $C'(\cdot)$ is non-decreasing, the result follows. ■

Proposition 4 *Maintain assumptions* LP1, LP2′, *and* LP3. *Suppose (i) demand slopes down ($P'(\cdot) < 0$); (ii)* inverse *demand is log concave; (iii) inverse demand and cost are at least* twice *differentiable; and (iv) $P'(x) < C''(y)$ for all x and all $y \in [0, x]$ (i.e., cost is never "too concave"). Then $MR(x) = MC(x)$ is sufficient as well as necessary for the profit-maximizing quantity.*

Proof: Consider an x such that $MR(x) = MC(x)$ (at least one exists because there is an interior maximum). In other words, consider an x such that

$$xP'(x) + P(x) - C'(x) = 0\,. \tag{2.5}$$

If the *derivative* of (2.5) is always negative evaluated at any x that solves (2.5), then that x is a maximum. Moreover, it is the unique maximum because a function cannot have multiple (local) maxima without minima; and if the derivative of (2.5) is always negative at any x that solves (2.5), then $R(x) - C(x)$ has no interior minima (recall (2.5) is also a necessary condition for x to minimize $R(x) - C(x)$). Differentiating (2.5) yields

$$2P'(x) + xP''(x) - C''(x) = xP''(x) + P'(x) + \underbrace{P'(x) - C''(x)}_{<0}\,.$$

It follows we're done if $xP''(x) + P'(x) \leq 0$. Given that demand curves slope down, we're done if $P''(x) \leq 0$. Suppose, therefore, that $P''(x) > 0$. The assumption that $P(\cdot)$ is log concave implies

$$P(x)P''(x) - \big(P'(x)\big)^2 \leq 0\,. \tag{2.6}$$

Because cost is non-decreasing, (2.5) implies $P(x) \geq -P'(x)x$. So (2.6) implies

$$0 \geq -P'(x)xP''(x) - \big(P'(x)\big)^2 \;\Rightarrow\; 0 \geq xP''(x) + P'(x) \tag{2.7}$$

(recall $-P'(x) > 0$). ■

Because demand curves slope down, $P'(x) < 0$; hence, expression (2.2) implies $MR(x) < P(x)$, except at $x = 0$ where $MR(0) = P(0)$. See Figure 2.1. Why is $MR(x) < P(x)$? That answer is that, to sell an additional item, the firm must lower its price (*i.e.,* recall, $P(x + \varepsilon) < P(x)$, $\varepsilon > 0$). So marginal revenue has two components: The price received on the marginal unit, $P(x)$, less the revenue lost on the infra-marginal units from having to lower the price, $|xP'(x)|$ (*i.e.,* the firm gets $|P'(x)|$ less on each of the x infra-marginal units).

2.1 Elasticity and the Lerner Markup Rule

We know that the revenue from selling 0 units is 0. For "sensible" demand curves, $\lim_{x\to\infty} xP(x) = 0$ because eventually price is driven down to zero. In between these extremes, revenue is positive. Hence, we know that revenue must increase over some range of output and decrease over another. Revenue is increasing if and only if

$$P(x) + xP'(x) > 0 \text{ or } xP'(x) > -P(x)\,.$$

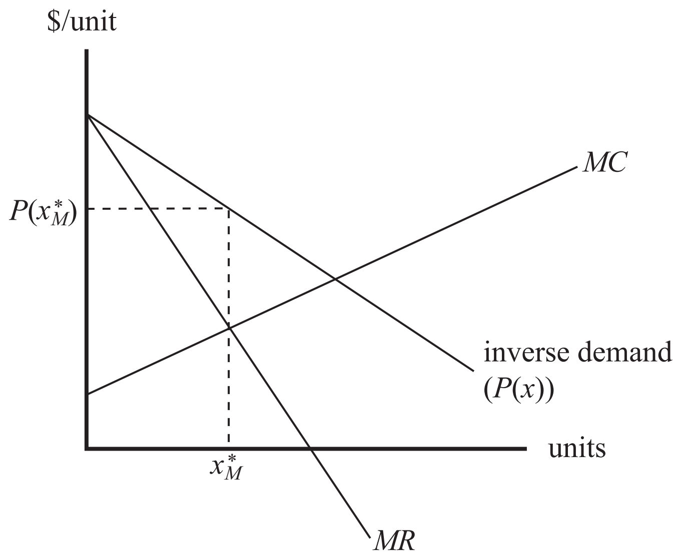

Figure 2.1: Relation between inverse demand, $P(x)$, and marginal revenue, *MR*, under linear pricing; and the determination of the profit-maximizing quantity, x_M^*, and price, $P(x_M^*)$.

Divide both sides by $-P(x)$ to get

$$1 > -\frac{xP'(x)}{P(x)} = \frac{1}{\epsilon}\,. \tag{2.8}$$

Multiplying both sides of (2.8) by ϵ, we have that revenue is increasing if and only if

$$\epsilon > 1\,. \tag{2.9}$$

When ϵ satisfies (2.9), we say that demand is *elastic*. When demand is elastic, revenue is increasing with units sold. If $\epsilon < 1$, we say that demand is *inelastic*. Reversing the various inequalities, it follows that, when demand is *in*elastic, revenue is decreasing with units sold. The case where $\epsilon = 1$ is called *unit elasticity*.

Recall that a firm produces the number of units that equates MR to MC. The latter is positive, which means that a profit-maximizing firm engaged in linear pricing operates only on the *elastic* portion of its demand curve. This makes intuitive sense: If it were on the *in*elastic portion, then, were it to produce less, it would both raise revenue and lower cost; that is, increase profits. Hence, it can't maximize profits operating on the inelastic portion of demand.

Summary 2 *A profit-maximizing firm engaged in linear pricing operates on the* elastic *portion of its demand curve.*

Rewrite the $MR(x) = MC(x)$ condition as

$$P(x) - MC(x) = -xP'(x)$$

and divide both sides by $P(x)$ to obtain

$$\frac{P(x) - MC(x)}{P(x)} = -\frac{xP'(x)}{P(x)} = \frac{1}{\epsilon}, \qquad (2.10)$$

Expression (2.10) is known as the *Lerner markup rule*.[1] In English, it says that the price markup over marginal cost, $P(x) - MC(x)$, as a proportion of the price is equal to $1/\epsilon$. Hence, the less elastic is demand (*i.e.,* as ϵ decreases towards 1), the greater the percentage of the price that is a markup over cost. Obviously, the portion of the price that is a markup over cost can't be greater than the price itself, which again shows that the firm must operate on the elastic portion of demand.

Welfare Analysis | 2.2

Assuming that consumer surplus is the right measure of consumer welfare (*e.g.,* consumers have quasi-linear utility), then total welfare is the sum of firm profits and consumer surplus. Hence, total welfare is

$$\underbrace{xP(x) - C(x)}_{\text{profit}} + \underbrace{\int_0^x (P(t) - P(x))dt}_{\text{CS}} = xP(x) - C(x) + \int_0^x P(t)dt - xP(x)$$

$$= \int_0^x P(t)dt - C(x). \qquad (2.11)$$

Observe, first, that neither the firm's revenue, $xP(x)$, nor the consumers' expenditure, $xP(x)$, appear in (2.11). This is the usual rule that *monetary transfers* made among agents are irrelevant to the amount of total welfare. Welfare is determined by the allocation of the real good; that is, the benefit, $\int P(t)dt$, that consumers obtain and the cost, $C(x)$, that the producer incurs.

Next observe that the derivative of (2.11) is $P(x) - MC(x)$. From (2.5) on page 19 (*i.e.,* from $MR(x_M^*) = MC(x_M^*)$), recall that $P(x_M^*) > MC(x_M^*)$, where x_M^* is the profit-maximizing quantity produced under linear pricing. This means that linear pricing leads to too little output from the perspective of maximizing welfare—if the firm produced more, welfare would increase.

Proposition 5 *Under linear pricing, the monopolist produces too little output from the perspective of total welfare.*

---

[1] Named for the economist Abba Lerner.

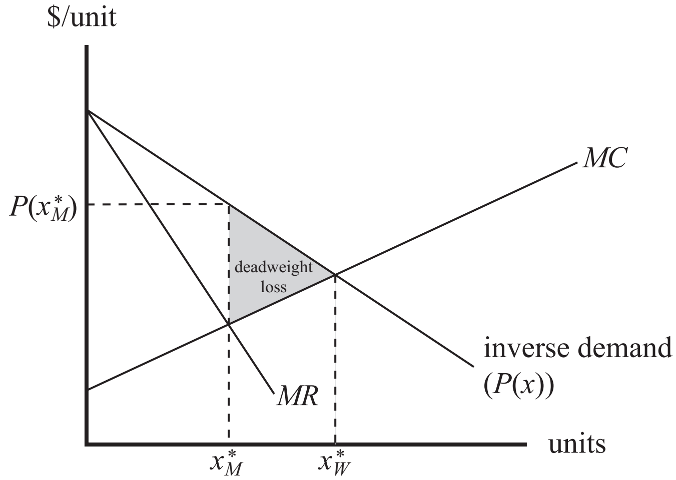

Figure 2.2: The deadweight loss from linear pricing is the shaded triangle.

If we assume—as is generally reasonable given that demand slopes down—that demand crosses marginal cost once from above, then the welfare-maximizing quantity satisfies

$$P(x) - MC(x) = 0\,. \tag{2.12}$$

Let x_W^* be the solution to (2.12). From Proposition 5, $x_W^* > x_M^*$.

What is the welfare loss from linear pricing? It is the amount of welfare forgone because only x_M^* units are traded rather than x_W^* units:

$$\begin{aligned}\left(\int_0^{x_W^*} P(t)dt - C(x_W^*)\right) - \left(\int_0^{x_M^*} P(t)dt - C(x_M^*)\right) \\ &= \int_{x_M^*}^{x_W^*} P(t)dt - \left(C(x_W^*) - C(x_M^*)\right) \\ &= \int_{x_M^*}^{x_W^*} P(t)dt - \int_{x_M^*}^{x_W^*} MC(t)dt \\ &= \int_{x_M^*}^{x_W^*} \big(P(t) - MC(t)\big)dt\,. \end{aligned} \tag{2.13}$$

The area in (2.13) is called the *deadweight loss* associated with linear pricing. It is the area beneath the demand curve and above the marginal cost curve between x_M^* and x_W^*. Because $P(x)$ and $MC(x)$ meet at x_W^*, this area is triangular (see Figure 2.2) and, thus, the area is often called the *deadweight-loss triangle*.

The existence of a deadweight-loss triangle is one reason why governments and antitrust authorities typically seek to discourage monopolization of industries and, instead, seek to encourage competition. Competition tends to drive price toward marginal cost, which causes output to approach the welfare-maximizing quantity.[2]

We can consider the welfare loss associated with linear pricing as a motive to change the industry structure (*i.e.,* encourage competition). We—or the firm—can also consider it as encouragement to change the method of pricing. The deadweight loss is, in a sense, money left on the table. As we will see, in some circumstances, clever pricing by the firm will allow it to pick some, if not all, of this money up off the table.

An Example

To help make all this more concrete, consider the following example. A monopoly has cost function $C(x) = 2x$; that is, $MC = 2$. It faces inverse demand $P(x) = 100 - x$.

Marginal revenue under linear pricing is $P(x) + xP'(x)$, which equals $100 - x + x \times (-1) = 100 - 2x$.[3] Equating MR with MC yields $100 - 2x = 2$; hence, $x^*_M = 49$. The profit-maximizing price is $100 - 49 = 51$.[4] Profit is revenue minus cost; that is, $51 \times 49 - 2 \times 49 = 2401$.[5] Consumer surplus is $\int_0^{49}(100 - t - 51)dt = \frac{1}{2} \times 49^2$.[6]

Total welfare, however, is maximized by equating price and marginal cost: $P(x) = 100 - x = 2 = MC$. So $x^*_W = 98$. Deadweight loss is, thus,

$$\int_{49}^{98}(\underbrace{100 - t}_{P(x)} - \underbrace{2}_{MC})dt = 98t - \frac{1}{2}t^2\Big|_{49}^{98} = 1200.5\,.$$

As an **exercise**, derive the general condition for deadweight loss for affine demand and constant marginal cost (*i.e.,* under the assumptions of footnote 4).

An Application 2.3

We often find simple monopoly pricing in situations that don't immediately appear to be linear-pricing situations. For example, suppose that a risk-neutral

[2] A full welfare comparison of competition versus monopoly is beyond the scope of these notes. See, for instance, Chapters 13 and 14 of Varian (1992) for a more complete treatment.

[3] **Exercise:** Prove that if inverse demand is an affine function, then marginal revenue is also affine with a slope that is twice as steep as inverse demand.

[4] **Exercise:** Prove that, if inverse demand is $P(x) = a - bx$ and $MC = c$, a constant, then $x^*_M = \frac{a-c}{2b}$ and $P(x^*_M) = \frac{a+c}{2}$.

[5] **Exercise:** Prove that profit under linear pricing is $\frac{1}{b}\left(\frac{a-c}{2}\right)^2$ under the assumptions of footnote 4.

[6] **Exercise:** Prove that consumer surplus under linear pricing is $\frac{(a-c)^2}{8b}$ under the assumptions of footnote 4.

seller faces a single buyer. Let the seller have single item to sell (*e.g.,* an artwork). Let the buyer's value for this artwork be v. The buyer knows v, but the seller does not. All the seller knows is that v is distributed according to the differential distribution function $F(\cdot)$. That is, the probability that $v \leq \hat{v}$ is $F(\hat{v})$. Assume $F'(\cdot) > 0$ on the support of v. Let the seller's value for the good—her cost—be c. Assume $F(c) < 1$.

Suppose that the seller wishes to maximize her expected profit. Suppose, too, that she makes a take-it-or-leave-it offer to the buyer; that is, the seller quotes a price, p, at which the buyer can purchase the good if he wishes. If he doesn't wish to purchase at that price, he walks away and there is no trade. Clearly, the buyer buys if and only if $p \leq v$; hence, the probability of a sale, x, is given by the formula $x = 1 - F(p)$. The use of "x" is intentional—we can think of x as the (expected) quantity sold at price p. Note, too, that, because the formula $x = 1 - F(p)$ relates quantity sold to price charged, it is a *demand* curve. Moreover, because the probability that the buyer's value is less than p is increasing in p, this demand curve slopes down. Writing $F(p) = 1 - x$ and inverting F (which we can do because it's monotonic), we have $p = F^{-1}(1-x) \equiv P(x)$. The seller's (expected) cost is cx, so marginal cost is c. The seller's (expected) revenue is $xP(x)$. As is clear, we have a standard linear-pricing problem. Marginal revenue is

$$P(x) + xP'(x) = F^{-1}(1-x) + x\left(\frac{-1}{F'[F^{-1}(1-x)]}\right) .$$

For example, if $c = 1/2$ and v is distributed uniformly on $[0,1]$, then $F(v) = v$, $F'(v) = 1$, and $F^{-1}(y) = y$. So $MR(x)$ is $1 - 2x$. Hence, $x_M^* = 1/4$ and, thus, the price the seller should ask to maximize her expected profit is $3/4$.[7] Note that there is a deadweight loss: Efficiency requires that the good change hands whenever $v > c$; that is, in this example, when $v > 1/2$. But given linear pricing, the good only changes hands when $v > 3/4$—in other words, half the time the good should change hands it doesn't.

[7] An alternative approach, which is somewhat more straightforward in this context, is to solve $\max_p (p-c)\bigl(1-F(p)\bigr)$.

3 First-degree Price Discrimination

We saw in Section 2.2 that linear pricing "leaves money on the table," in the sense that there are gains to trade—the deadweight loss—that are not realized. There is money to be made if the number of units traded can be increased from x_M^* to x_W^*.

Why has this money been left on the table? The answer is that trade benefits both buyer and seller. The seller profits to the extent that the revenue received exceeds cost and the buyer profits to the extent that the benefit enjoyed exceeds the cost. The seller, however, does not consider the positive externality she creates for the buyer (buyers) by selling him (them) goods. The fact that his (their) marginal benefit schedule (*i.e.,* inverse demand) lies above his (their) marginal cost (*i.e.,* the price the seller charges) is irrelevant to the seller insofar as she doesn't capture any of this gain enjoyed by the buyer (buyers). Consequently, she underprovides the good. This is the usual problem with positive externalities: The decision maker doesn't internalize the benefits others derive from her action, so she does too little of it from a social perspective. In contrast, were the action decided by a social planner seeking to maximize social welfare, then more of the action would be taken because the social planner does consider the externalities created. The cure to the positive externalities problem is to change the decision maker's incentives so she effectively faces a decision problem that replicates the social planner's problem.

One way to make the seller internalize the externality is to give her the social benefit of each unit sold. Recall the marginal benefit of the the xth unit is $P(x)$. So let the seller get $P(1)$ if she sells one unit, $P(1) + P(2)$ if she sells two, $P(1) + P(2) + P(3)$ if she sells three, and so forth. Given that her revenue from x units is $\int_0^x P(t)dt$, her *marginal* revenue schedule is $P(x)$. Equating marginal revenue to marginal cost, she produces x_W^*, the welfare-maximizing quantity.

In general, allowing the seller to vary price unit by unit, so as to *march down the demand curve*, is impractical. But, as we will see, there are ways for the seller to effectively duplicate marching down the demand curve. When the seller can march down the demand curve or otherwise capture all the surplus, she's said to be engaging in *first-degree price discrimination*. This is sometimes called *perfect price discrimination*.

Two-Part Tariffs 3.1

Consider a seller who faces a single buyer with inverse demand $p(x)$. Let the seller offer a *two-part tariff*: The buyer pays as follows:

$$T(x) = \begin{cases} 0\,, & \text{if } x = 0 \\ px + f & \text{if } x > 0 \end{cases}, \tag{3.1}$$

where p is price per unit and f is the *entry fee*, the amount the buyer must pay to have access to any units. The scheme in (3.1) is called a two-part tariff because there are two parts to what the buyer pays (the tariff), the unit price and the entry fee.

The buyer will buy only if f is not set so high that he loses all his consumer surplus. That is, he buys provided

$$f \leq \int_0^x (p(t) - p(x))dt = \int_0^x p(t)dt - xp(x)\,. \tag{3.2}$$

Constraints like (3.2) are known as *participation constraints* or *individual-rationality (IR) constraints*. These constraints often arise in pricing schemes or other mechanism design. They reflect that, because participation in the scheme or mechanism is voluntary, it must be induced.

The seller's problem is to choose x (effectively, p) and f to maximize profit subject to (3.2); that is, maximize

$$f + xp(x) - C(x) \tag{3.3}$$

subject to (3.2). Observe that (3.2) must bind: If it didn't, then the seller could raise f slightly, keeping x fixed, thereby increasing her profits without violating the constraint. Note this means that the entry fee is set equal to the consumer surplus that the consumer receives. Because (3.2) is binding, we can substitute it into (3.3) to obtain the unconstrained problem:

$$\max_x \int_0^x p(t)dt - xp(x) + xp(x) - C(x)\,.$$

The first-order condition is $p(x) = MC(x)$; that is, the profit-maximizing quantity is the welfare-maximizing quantity. The unit price is $p(x_W^*)$ and the entry fee is $\int_0^{x_W^*} p(t)dt - x_W^* p(x_W^*)$.

Proposition 6 *A seller who sells to a single buyer with known demand does best to offer a two-part tariff with the unit price set to equate demand and marginal cost and the entry fee set equal to the buyer's consumer surplus at that unit price. Moreover, this solution maximizes welfare.*

Of course, a seller rarely faces a single buyer. If, however, the buyers all have the same demand, then a two-part tariff will also achieve efficiency and allow the seller to achieve the maximum possible profits. Let there be J buyers all of whom are assumed to have the same demand curve. As before, let $P(\cdot)$ denote aggregate inverse demand. The seller's problem in designing the optimal two-part tariff is

$$\max_{f,x} Jf + xP(x) - C(x) \tag{3.4}$$

subject to consumer participation,

$$f \leq cs_j\big(P(x)\big)\,, \tag{3.5}$$

where $cs_j(p)$ denotes the jth buyer's consumer surplus at price p. Because the buyers are assumed to have identical demand, the subscript j is superfluous and constraint (3.5) is either satisfied for all buyers or it is satisfied for no buyer. As before, (3.5) must bind, otherwise the seller could profitably raise f. Substituting the constraint into (3.4), we have

$$\max_{x} J \times cs\big(P(x)\big) + xP(x) - C(x)\,,$$

which, because aggregate consumer surplus is the sum of the individual surpluses (recall Proposition 2 on page 9), can be rewritten as

$$\max_{x} \underbrace{\int_0^x P(t)dt - xP(x)}_{\text{aggregate CS}} + xP(x) - C(x)\,.$$

The solution is x_W^*. Hence, the unit price is $P(x_W^*)$ and the entry fee, f, is

$$\frac{1}{J}\left(\int_0^{x_W^*} P(t)dt - x_W^* P(x_W^*)\right)\,.$$

Proposition 7 *A seller who sells to J buyers, all with* identical *demands, does best to offer a two-part tariff with the unit price set to equate demand and marginal cost and the entry fee set equal to $1/J$th of aggregate consumer surplus at that unit price. This maximizes social welfare and allows the seller to capture all of social welfare.*

We see many examples of two-part tariffs in real life. A classic example is an amusement park that charges an entry fee and a per-ride price (the latter, sometimes, being set to zero). Another example is a price for a machine (*e.g.,* a Polaroid instant camera or a punchcard sorting machine), which is a form of entry fee, and a price for an essential input (*e.g.,* instant film or punchcards), which is a form of per-unit price.[1] Because, in many instances, the per-unit price

[1] Such schemes can also be seen as a way of providing consumer credit: The price of the capital good (*e.g.,* camera or machine) is set lower than the profit-maximizing price (absent liquidity constraints), with the difference being essentially a loan to the consumer that he repays by paying more than the profit-maximizing price (absent a repayment motive) for the ancillary good (*e.g.,* film or punchcards).

is set to zero, some two-part tariffs might not be immediately obvious (*e.g.,* an annual service fee that allows unlimited "free" service calls, a telephone calling plan in which the user pays so much per month for unlimited "free" phone calls, or amusement park that allows unlimited rides with paid admission).

Packaging: *A disguised two-part tariff.*

Packaging is another way to design a two-part tariff. For instance, a grocery story could create a two-part tariff in the following way. Suppose that, rather than being sold in packages, sugar were kept in a large bin and customers could purchase as much or as little as they liked (*e.g.,* like fruit at most groceries or as is actually done at "natural" groceries). Suppose that, under the optimal two-part tariff, each consumer would buy x pounds, which would yield him surplus of cs, which would be captured by the store using an entry fee of $f = cs$. Alternatively, but equivalently, the grocery could package sugar. Each bag of sugar would have x pounds and would cost $px + cs$ *per bag*. Each consumer would face the binary decision of whether to buy 0 pounds or x pounds. Each consumer's total benefit from x pounds is $px + cs$, so each would just be willing to pay $px + cs$ for the package of sugar. Because the entry fee is paid on every x-pound bag, the grocery has devised a (disguised) two-part tariff that is also arbitrage-proof. In other words, *packaging*—taking away consumers ability to buy as much or as little as they wish—can represent an arbitrage-proof way of employing a two-part tariff.

The Two-Instruments Principle

When the seller was limited to just one price parameter, p—that is, engaged in linear pricing—she made less money than when she controlled two parameters, p and f. One way to explain this is that a two-part tariff allows the seller to face the social planner's problem of maximizing welfare and, moreover, capture all welfare. Because society can do no better than maximize welfare and the seller can do no better than capture all of social welfare, she can't do better than a two-part tariff in this context.

But this begs the question of why she couldn't do as well with a single price parameter. Certainly, she could have maximized social welfare; all she needed to do was set $P(x) = MC(x)$. But the problem with that solution is there is no way for her to capture all the surplus she generates. If she had an entry fee, then she could use this to capture the surplus; but with linear pricing we've forbidden her that instrument.

The problem with using just the unit price is that we're asking one instrument to do two jobs. One is to determine allocation. The other is to capture surplus for the seller. Only the first has anything to do with efficiency, so the fact that the seller uses it for a second purpose is clearly going to lead to a distortion. If we give the seller a second instrument, the entry fee, then she has two instruments for the two jobs and she can "allocate" each job an instrument. This is a fairly general idea—efficiency is improved by giving the mechanism designer more instruments—call this the *two-instruments principle*.

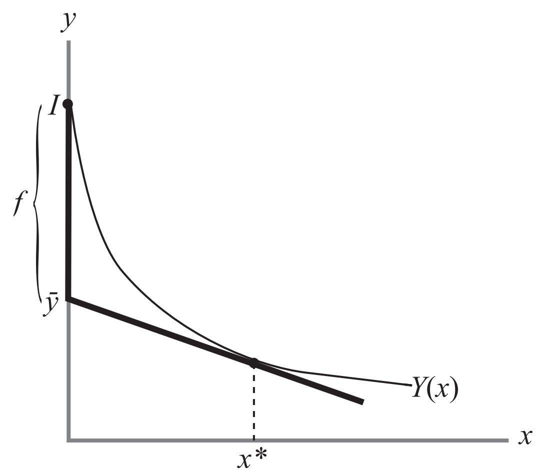

Figure 3.1: A general analysis of a two-part tariff.

Two-Part Tariffs without Apology

It might seem that the analysis of two-part tariffs is dependent on our assumption of quasi-linear utility. In fact, this is *not* the case. To see this, consider a single consumer with utility $u(x, y)$. Normalize the price of y to 1. Assume the individual has income I. Define $Y(x)$ to be the indifference curve that passes through the bundle $(0, I)$; that is, the bundle in which the consumer purchases only the y-good. See Figure 3.1. Assume $MC = c$.

Consider the seller of the x good. If she imposes a two-part tariff, then she transforms the consumer's budget constraint to be the union of the vertical line segment $\{(0, y) | I - f \leq y \leq I\}$ and the line $y = (I - f) - px$, $x > 0$. If we define $\bar{y} = I - f$, then this budget constraint is the thick dark curve shown in Figure 3.1. Given that the consumer can always opt to purchase none of the x good, the consumer can't be put below the indifference curve through $(0, I)$; that is, below $Y(x)$. For a given p, the seller increases profit by raising f, the entry fee. Hence, the seller's goal is to set f so that this kinked budget constraint is just tangent to the indifference $Y(x)$. This condition is illustrated in Figure 3.1, where the kinked budget constraint and $Y(x)$ are tangent at x^*. If the curves are tangent at x^*, then

$$-p = Y'(x^*) \, . \tag{3.6}$$

At x^*, the firm's profit is

$$(p - c)x^* + f \tag{3.7}$$

(recall we've assumed $MC = c$). As illustrated, $f = I - \bar{y}$. In turn, $\bar{y} =$

$Y(x^*) + px^*$.[2] We can, thus, rewrite (3.7) as

$$(p - c)x^* + I - Y(x^*) - px^* = -Y(x^*) - cx^* + I\,. \tag{3.8}$$

Maximizing (3.8) with respect to x^*, we find that $c = -Y'(x^*)$. Substituting for $Y'(x^*)$ using (3.6), we find that $c = p$; that is, as before, the seller maximizes profits by setting the unit price equal to marginal cost. The entry fee is $I - (Y(x^*) + cx^*)$, where x^* solves $c = -Y'(x^*)$. Given that $MC = c$, it is clear this generalizes for multiple consumers.

Summary 3 *The conclusion that the optimal two-part tariff with one consumer or homogeneous consumers entails setting the unit price equal to marginal cost is* not *dependent on the assumption of quasi-linear utility.*

As a "check" on this analysis, observe that $Y'(\cdot)$ is the marginal rate of substitution (MRS). With quasi-linear utility; that is, $u(x, y) = v(x)+y$, the MRS is $-v'(x)$. So x^* satisfies $c = -(-v'(x)) = v'(x) = P(x)$, where the last equality follows because, with quasi-linear utility, the consumer's inverse demand curve is just his marginal benefit (utility) of the good in question. This, of course, corresponds to what we found above (recall Proposition 6).

Bibliographic Note

For more on two-part and multi-part tariffs, see Wilson (1993). Among other topics, Wilson investigates optimal two-part tariffs with *heterogeneous* consumers. Varian (1989) is also a useful reference.

[2]Because the line segment $\bar{y} - px$ is tangent to $Y(\cdot)$ at x^*, we have $\bar{y} - px^* = Y(x^*)$.

Third-degree Price Discrimination[1] | 4

In real life, consumers are rarely homogenous with respect to their preferences and, hence, demands. We wish, therefore, to extend our analysis of price discrimination to accommodate heterogenous consumers.

The Notion of Type | 4.1

To analyze heterogeneous consumers, imagine that we can index the consumers' different utility functions, which here is equivalent to indexing their different demand functions. We refer to this index as the *type space*, a given *type* being a specific index number. All consumers of the same type have the same demand. For example, suppose all consumers have demand of the form

$$x(p) = \begin{cases} 1\,, & \text{if } p \leq \theta \\ 0\,, & \text{if } p > \theta \end{cases} ;$$

that is, a given consumer wants at most one unit and only if price does not exceed θ. Suppose θ varies across consumers. In this context, θ is a consumer's type (index). The set of possible values that θ can take is the type space. For instance, suppose that consumers come in one of two types, $\theta = 10$ or $\theta = 15$. In this case, the type space, Θ, can be written as $\Theta = \{10, 15\}$. Alternatively, we could have a continuous type space; for instance, $\Theta = [a, b]$, a and $b \in \mathbb{R}_+$.

Consider the second example. Suppose there is a continuum of consumers of measure J whose types are distributed uniformly over $[a, b]$. Consequently, at a uniform price of p, demand is $J\frac{b-p}{b-a}$. Suppose the firm's cost of x units is cx; observe $MC = c$. As there would otherwise be no trade, assume $c < b$. If the firm engaged in linear pricing, its profit-maximizing price would be

$$p^* = \begin{cases} \frac{b+c}{2}\,, & \text{if } \frac{b+c}{2} \geq a \\ a\,, & \text{if } \frac{b+c}{2} < a \end{cases} .$$

(**Exercise:** Verify.) Rather than deal with both cases, assume $a < (b + c)/2$. Given this last assumption, there is no further loss of generality in normalizing the parameters so $a = 0$, $b = 1$, and $0 \leq c < 1$.

[1] What happened to *second-degree* price discrimination? Despite the conventional ordering, it makes more sense to cover third-degree price discrimination before second-degree price discrimination.

The firm's profit is

$$\pi_{\text{LP}} = J\frac{(1-c)^2}{4}\,. \tag{4.1}$$

(**Exercise:** Verify.)

Suppose, instead of linear pricing, the firm knew each consumer's type and could base its price to that consumer on his type. Because a consumer is willing to pay up to θ, the firm maximizes its profit by charging each consumer θ (provided $\theta \geq c$). Given that the firm is capturing all the surplus, this is perfect discrimination. The firm's profit is

$$\pi_{\text{PD}} = J\int_c^1 (\theta - c)d\theta = J\frac{(1-c)^2}{2}\,. \tag{4.2}$$

Clearly, this exceeds the profit given in (4.1). Of course, we knew this would be the case without calculating (4.2): linear pricing, because it generates a deadweight loss and leaves some surplus in consumer hands, cannot yield the firm as great a profit as perfect discrimination.

The example illustrates that, with heterogenous consumers, the ideal from the firm's perspective would be to base its prices on consumers' types. In most settings, however, that is infeasible. The question then becomes how closely can the firm approximate that ideal through its pricing.

Characteristic-based Discrimination | 4.2

When a seller cannot observe consumers' types, she has two choices. One, she can essentially ask consumers their type; this, as we will see in the next chapter, is what second-degree price discrimination is all about. Two, she can base her prices on observable characteristics of the consumers, where the observable characteristics are correlated in some way with the underlying types. This is *third-degree price discrimination*.

Examples of third-degree price discrimination are pricing based on observable characteristics such as age, gender, student status, geographic location, or temporally different markets.[2] The idea is that, say, student status is correlated with willingness to pay; on average, students have a lower willingness to pay for an event (*e.g.,* a movie) than do working adults.

Formally, consider a seller who can discriminate on the basis of M observable differences. Let m denote a particular characteristic (*e.g.,* $m = 1$ is student and $m = 2 = M$ is other adult). Based on the distribution of types conditional on m, the firm's demand from those with characteristic m is $X_m(\cdot)$. Let $P_m(\cdot)$ be the corresponding inverse demand. For example, suppose individual demand is the L-shaped demand of the previous section, $\theta|m = 1 \sim 1 - (1-\theta)^2$ on $[0, 1]$, and $\theta|m{=}2 \sim \theta^2$ (observe the second dominates the first in the sense of

[2] Although pricing differently at different times could also be part of second-degree price discrimination scheme.

first-order stochastic dominance). In this case, if J_m is the measure or number of consumers with characteristic m, we have

$$X_1(p) = J_1(1-p)^2 \quad \text{and}$$
$$X_2(p) = J_2\big(1-p^2\big)\,.$$

The seller's problem is

$$\max_{\{x_1,\dots,x_M\}} \sum_{m=1}^{M} x_m P_m(x_m) - C\left(\sum_{m=1}^{M} x_m\right). \tag{4.3}$$

Imposing assumptions sufficient to make the first-order condition sufficient as well as necessary (*e.g.,* assuming (4.3) is concave), the solution is given by

$$P_m + x_m P_m'(x_m) - MC\left(\sum_{m=1}^{M} x_m\right) = 0\,, \text{ for } m = 1,\dots,M\,. \tag{4.4}$$

Some observations based on conditions (4.4):

- If marginal cost is a constant (*i.e., $MC = c$*), then third-degree price discrimination is nothing more than setting profit-maximizing linear prices independently in M different markets.
- If marginal cost is *not* constant, then the markets can*not* be treated independently; how much the seller wishes to sell in one market is dependent on how much she sells in other markets. In particular, if marginal cost is not constant and there is a shift in demand in one market, then the quantity sold in *all* markets can change.
- Marginal revenue across the M markets is the same at the optimum; that is, if the seller found herself with one more unit of the good, it wouldn't matter in which market (to which group) she sold it.

Welfare Considerations | 4.3

Does allowing a seller to engage in third-degree price discrimination raise or lower welfare. That is, if she were restricted to set a single price for all markets, would welfare increase or decrease?

We will answer this question for the case in which $MC = c$ and there are two markets, $m = 1, 2$. Let $v_m(x) = \int_0^x p_\theta(t)dt$; that is, $v_m(x)$ is the gross aggregate benefit enjoyed in market m (by those in group m). Welfare is, therefore,

$$W(x_1, x_2) = v_1(x_1) + v_2(x_2) - (x_1 + x_2)c\,.$$

In what follows, let x_m^* be the quantity traded in market θ under third-degree price discrimination and let x_m^U be the quantity traded in market m if the seller must charge a uniform price across the two markets.[3] Because demand curves slope down, $v_m(\cdot)$ is a concave function, which means

$$\begin{aligned} v_m(x_m^*) &< v_m(x_m^U) + v_m'(x_m^U)\cdot(x_m^* - x_m^U) \\ &= v_m(x_m^U) + p_m(x_m^U)\cdot(x_m^* - x_m^U)\,. \end{aligned} \tag{4.5}$$

Likewise,

$$v_m(x_m^U) < v_m(x_m^*) + p_m(x_m^*)\cdot(x_m^U - x_m^*)\,. \tag{4.6}$$

If we let $\Delta x_m = x_m^* - x_m^U$, $p_m^* = p_m(x_m^*)$, $p^U = p_m(x_m^U)$ (note, by assumption, this last price is common across the markets), and $\Delta v_m = v_m(x_m^*) - v_m(x_m^U)$, then we can combine (4.5) and (4.6) as

$$p^U \Delta x_m > \Delta v_m > p_m^* \Delta x_m\,. \tag{4.7}$$

Going from a uniform price across markets to different prices (*i.e.,* to 3rd-degree price discrimination) changes welfare by

$$\Delta W = \Delta v_1 + \Delta v_2 - (\Delta x_1 + \Delta x_2)c\,.$$

Hence, using (4.7), the change in welfare is bounded by

$$(p^U - c)(\Delta x_1 + \Delta x_2) > \Delta W > (p_1^* - c)\Delta x_1 + (p_2^* - c)\Delta x_2\,. \tag{4.8}$$

Because $p^U - c > 0$, if $\Delta x_1 + \Delta x_2 \leq 0$, then switching from a single price to third-degree price discrimination must reduce welfare. In other words, if aggregate output falls (weakly), then welfare must be reduced. For example, suppose that $c = 0$ and $X_m(p) = a - b_\theta p$, then $x_m^* = a/2$.[4] Aggregate demand across the two markets is $\mathcal{X}(p) = 2a - (b_1 + b_2)p$ and $x_1^U + x_2^U = 2a/2$. This equals $x_1^* + x_2^*$, so there is no increase in aggregate demand. From (4.8), we can conclude that third-degree price discrimination results in a *loss* of welfare relative to a uniform price in this case.

But third-degree price discrimination can also *increase* welfare. The quickest way to see this is to suppose that, at the common monopoly price, one of the two markets is shut out (*e.g.,* market 1, say, has relatively little demand and no demand at the monopoly price that the seller would set if obligated to charge the same price in both markets). Then, if price discrimination is allowed, the already-served market faces the same price as before—so there's no change in its consumption or welfare, but the unserved market can now be served, which increases welfare in that market from zero to something positive.

[3] To determine x_m^U, let $\mathcal{X}(p) = X_1(p) + X_2(p)$ be aggregate demand across the two markets, and let $\mathcal{P}(x) = \mathcal{X}^{-1}(p)$ be aggregate inverse demand. Solve $\mathcal{P}(x) + x\mathcal{P}'(x) = c$ for x (*i.e.,* solve for optimal aggregate production assuming one price). Call that solution x_M^*. Then $x_m^U = X_m\big(\mathcal{P}(x_M^*)\big)$.

[4] One can quickly verify this by maximizing profits with respect to price. Alternatively, observe that inverse demand is

$$P(x) = \frac{a}{b} - \frac{x}{b}\,.$$

Hence, $x^* = a/2$ (see footnote 4 on page 23).

Bibliographic Note

This discussion of welfare under third-degree price discrimination draws heavily from Varian (1989).

4.4 Arbitrage

We have assumed, so far, in our investigation of price discrimination that *arbitrage* is impossible. That is, for instance, a single buyer can't pay the entry fee, then resell his purchases to other buyers, who, thus, escape the entry fee. Similarly, a good purchased in a lower-price market cannot be resold in a higher-price market.

In real life, however, arbitrage can occur. This can make utilizing nonlinear pricing difficult; moreover, the possibility of arbitrage helps to explain why we see nonlinear pricing in some contexts, but not others. For instance, it is difficult to arbitrage amusement park rides to those who haven't paid the entry fee. But is easy to resell supermarket products. Hence, we see two-part tariffs at amusement parks, but we typically don't see them at supermarkets.[5] Similarly, senior-citizen discounts to a show are either handled at the door (*i.e.,* at time of admission), or through the use of color-coded tickets, or through some other means to discourage seniors from reselling their tickets to their juniors.

If the seller cannot prevent arbitrage, then the separate markets collapse into one and there is a single uniform price across the markets. The welfare consequences of this are, as shown in the previous section, ambiguous. Aggregate welfare may either be increased or decreased depending on the circumstances. The seller, of course, is made worse off by arbitrage—given that she could, but didn't, choose a uniform price indicates that a uniform price yields lower profits than third-degree price discrimination.

4.5 Capacity Constraints

Third-degree price discrimination often comes up in the context of discounts for certain groups to some form of entertainment (*e.g.,* a play, movie, or sporting event). Typically, the venue for the event has limited capacity and it's worth considering the implication that has for third-degree price discrimination.

Consider an event for which there are two audiences (*e.g.,* students and non-students). Assume the (physical) marginal cost of a seat is essentially 0. The number of seats sold if unconstrained would be x_1^* and x_2^*, where x_m^* solves

$$P_m(x) + xP_m'(x) = MC = 0\,.$$

[5]Remember, however, that packaging can be a way for supermarkets to use arbitrage-proof two-part tariffs.

If the capacity of the venue, K, is greater than $x_1^*+x_2^*$, then there is no problem. As a convention, assume that $P_2(x_2^*) > P_1(x_1^*)$ (*e.g.,* group 1 are students and group 2 are non-students).

Suppose, however, that $K < x_1^* + x_2^*$. Then a different solution is called for. It might seem, given a binding capacity constraint, that the seller would abandon discounts (*e.g.,* eliminate student tickets), particularly if $x_2^* \geq K$ (*i.e.,* the seller could sell out charging just the high-paying group its monopoly price). This view, however, is naïve, as we will see.

The seller's problem can be written as

$$\max_{\{x_1,x_2\}} x_1 P_1(x_1) + x_2 P_2(x_2)$$

(recall we're assuming no physical costs that vary with tickets sold) subject to

$$x_1 + x_2 \leq K\,.$$

Given that we know the *un*constrained problem violates the constraint, the constraint must bind. Let λ be the Lagrange multiplier on the constraint. The first-order conditions are, thus,

$$P_1(x_1) + x_1 P_1'(x_1) - \lambda = 0 \text{ and}$$
$$P_2(x_2) + x_2 P_2'(x_2) - \lambda = 0\,.$$

Observe that the marginal revenue from each group is set equal to λ, the shadow price of the constraint. Note, too, that the two marginal revenues are equal. This makes intuitive sense: What is the marginal cost of selling a ticket to a group-1 customer? It's the opportunity cost of that ticket, which is the forgone revenue of selling it to a group-2 customer; that is, the marginal revenue of selling to a group-2 customer.

Now we can see why the seller might not want to sell only to the high-paying group. Suppose, by coincidence, that $x_2^* = K$; that is, the seller could sell out the event at price $P_2(x_2^*)$. She wouldn't, however, do so because

$$P_1(0) > 0 = P_2(x_2^*) + x_2^* P_2'(x_2^*)\,;$$

(the equality follows from the definition of x_2^* given that physical marginal cost is 0). The marginal revenue of the Kth seat, if sold to a group-2 customer, is clearly less than its marginal (opportunity) cost.

As an example, suppose that $P_1(x) = 40 - x$ and $P_2(x) = 100 - x$. Suppose $K = 50$. You should readily be able to verify that $x_1^* = 20$ and $x_2^* = 50$; that is, the seller could just sell out if she set a price of \$50, which would yield sales only to group-2 customers (no group-1 customer would pay \$50 for a seat). Her (accounting) profit would be \$2500. This, however, is not optimal. Equating the marginal revenues, we have

$$40 - 2x_1 = 100 - 2x_2\,. \tag{4.9}$$

Substituting the constraint, $x_1 = 50 - x_2$, into (4.9) yields

$$40 - 2(50 - x_2) = 100 - 2x_2\,; \text{ or}$$
$$4x_2 = 160\,.$$

So, optimally, $x_2 = 40$ and, thus, $x_1 = 10$. The seller's profit is $40 \times (100 - 40) + 10 \times (40 - 10) = 2700$ dollars. As claimed, this amount exceeds her take from naïvely pricing only to the group-2 customers.

While the seller's profit is greater engaging in third-degree price discrimination (*i.e.,* charging \$30 for student tickets and \$60 for regular tickets) than it is under uniform pricing (*i.e.,* \$50 per ticket), welfare is less under third-degree price discrimination. We know this, of course, from the discussion in Section 4.3—output hasn't changed (it's constrained to be 50)—so switching from uniform pricing to price discrimination must lower welfare. We can also see this by considering the last 10 tickets sold. Under uniform pricing, they go to group-2 consumers, whose value for them ranges from \$60 to \$50 and whose aggregate gross benefit is $\int_{40}^{50}(100 - t)dt = 550$ dollars. Under price discrimination, they are reserved for group-1 consumers (students), whose value for them ranges from \$40 to \$30 and whose aggregate gross benefit is just $\int_0^{10}(40 - t)dt = 350$ dollars. In other words, to capture more of the total surplus, the seller distorts the allocation from those who value the tickets more to those who value them less.

5 Second-degree Price Discrimination

In many contexts, a seller knows that different types or groups of consumers have different demand, but she can't readily identify from which group any given buyer comes. For example, it is known that business travelers are willing to pay more for most flights than are tourists. But it is impossible to know whether a given flier is a business traveler or a tourist.

A well-known solution is to offer different kinds of tickets. For instance, because business travelers don't wish to stay over the weekend or often can't book much in advance, the airlines charge more for round-trip tickets that don't involve a Saturday-night stayover or that are purchased within a few days of the flight (*i.e.,* in the latter situation, there is a discount for advance purchase). Observe an airline still can't observe which type of traveler is which, but by offering different kinds of service it hopes to *induce revelation* of which type is which. When a firm induces different types to reveal their types for the purpose of differential pricing, we say the firm is engaged in *second-degree price discrimination*.

Restricted tickets are one example of price discrimination. They are an example of second-degree price discrimination via *quality distortions*. Other examples include:

- Different classes of service (*e.g.,* first and second-class carriages on trains). The classic example here is the French railroads in the 19th century, which removed the roofs from second-class carriages to create third-class carriages.

- Hobbling a product. This is popular in high-tech, where, for instance, Intel produced two versions of a chip by "brain-damaging" the state-of-the-art chip. Another example is software, where "regular" and "pro" versions (or "home" and "office" versions) of the same product are often sold.

- Restrictions. Saturday-night stayovers and advance-ticketing requirements are a classic example. Another example is limited versus full memberships at health clubs.

The other common form of second-degree price discrimination is via *quantity discounts*. This is why, for instance, the liter bottle of soda is typically *less* than twice as expensive as the half-liter bottle. Quantity discounts can often be operationalized through multi-part tariffs, so many multi-part tariffs are examples of price discrimination via quantity discounts (*e.g.,* choices in calling plans between say a low monthly fee, few "free" minutes, and a high per-minute charge thereafter versus a high monthly fee, more "free" minutes, and a lower per-minute charge thereafter).

Quality Distortions 5.1

Consider an airline facing N_b business travelers and N_τ tourists on a given round-trip route. Suppose, for convenience, that the airline's cost of flying are essentially all fixed costs (*e.g.,* the fuel, aircraft depreciation, wages of a *fixed*-size crew, etc.) and that the marginal costs per flier are effectively 0. Let there be two possible kinds of round-trip tickets,[1] restricted (*e.g.,* requiring a Saturday-night stayover) and unrestricted (*e.g.,* no stayover requirements); let superscripts r and u refer to these two kinds of tickets, respectively. Let κ denote an arbitrary kind of ticket (*i.e.,* $\kappa \in \{r, u\}$).

A type-θ flier has a valuation (gross benefit or utility) of v_θ^κ for a κ ticket. Assume, consistent with experience, that

- $v_\theta^u \geq v_\theta^r$ for both θ and *strictly* greater for business travelers. That is, fliers prefer unrestricted tickets *ceteris paribus* and business travelers strictly prefer them.
- $v_b^\kappa > v_\tau^\kappa$ for both κ. That is, business travelers value travel more.

Clearly, if the airline offered only one kind of ticket, it would offer *un*restricted tickets given that they cost no more to provide and they can command greater prices. There are, then, two possible prices with one kind of ticket (*i.e.,* when engaged in linear pricing): (i) sell to both types, which means $p = v_\tau^u$; or (ii) sell to business travelers only, which means $p = v_b^u$. Option (i) is as good or better than option (ii) if and only if

$$N_b v_b^u \leq (N_b + N_\tau) v_\tau^u \text{ or, equivalently, } N_b \leq \frac{v_\tau^u}{v_b^u - v_\tau^u} N_\tau \,. \tag{5.1}$$

Alternatively, the airline could offer both kinds of tickets (price discriminate via quality distortion). In this case, the restricted ticket will be intended for one type of flier and the unrestricted ticket will be intended for the other. At some level, which kind of ticket goes to which type of flier needs to be determined, but experience tells us, in this context, that we should target the unrestricted ticket to the business traveler and the restricted ticket to the tourist. Let p^κ be the price of a κ ticket.

The airline seeks to do the following:

$$\max_{\{p^u, p^r\}} N_b p^u + N_\tau p^r \,. \tag{5.2}$$

It, of course, faces constraints. First, fliers need to purchase. This means that buying the kind of ticket intended for you can't leave you with negative consumer

[1] Assume all travel is round-trip.

surplus, because, otherwise, you would do better not to buy. Hence, we have the participation or IR constraints:

$$v_b^u - p^u \geq 0 \tag{5.3}$$
$$v_\tau^r - p^r \geq 0\,. \tag{5.4}$$

But we also have an additional set of constraints because the different types of flier need to be *induced* to reveal their types by purchasing the kind of ticket intended for them. Such *revelation* or *incentive compatibility* (IC) constraints are what distinguish second-degree price discrimination from other forms of price discrimination. Moreover, because such constraints are a feature of *screening* models, they indicate that second-degree price discrimination is a form of screening. The revelation constraints say that a flier type must do better (gain more consumer surplus) purchasing the ticket intended for him rather than purchasing the ticket intended for the other type:

$$v_b^u - p^u \geq v_b^r - p^r \tag{5.5}$$
$$v_\tau^r - p^r \geq v_\tau^u - p^u\,. \tag{5.6}$$

To summarize, the airline's problem is to solve (5.2) subject to (5.3)–(5.6). This is a *linear* programming problem. To solve it, begin by noting that the constraint set is empty if $v_\tau^u - v_\tau^r > v_b^u - v_b^r$ because, then, no price pair could satisfy (5.5) and (5.6). Hence, let's assume:

$$v_\tau^u - v_\tau^r \leq v_b^u - v_b^r\,. \tag{5.7}$$

Most screening problems require an assumption like (5.7) about how the marginal utilities are ordered across types. In many contexts, these are referred to as *single-crossing conditions* because they are or can be interpreted as assumptions about the relative steepness of the types' indifference curves when they cross.

We can visualize the constraints in (p^r, p^u) space (*i.e.,* the space with p^r on the horizontal axis and p^u on the vertical axis). See Figure 5.1. There are two constraints on the maximum p^u: It must lie below the horizontal line $p^u = v_b^u$ (line ① in Figure 5.1) and it must lie below the upward-sloping line $p^u = p^r + v_b^u - v_b^r \equiv \ell(p^r)$ (line ③). The former condition is just (5.3) and the latter is (5.5). From (5.4), we know we must restrict attention to $p^r \leq v_\tau^r$ (line ②). We also need to satisfy the tourist's incentive-compatibility constraint (5.6); that is, prices must lie above line ④. Hence the set of feasible prices is the gray region in Figure 5.1.

Observe we know:

$$\begin{aligned} p^u \leq \ell(p^r) &\leq \ell(v_\tau^r) \\ &= v_b^u - (v_b^r - v_\tau^r) \\ &< v_b^u\,, \end{aligned}$$

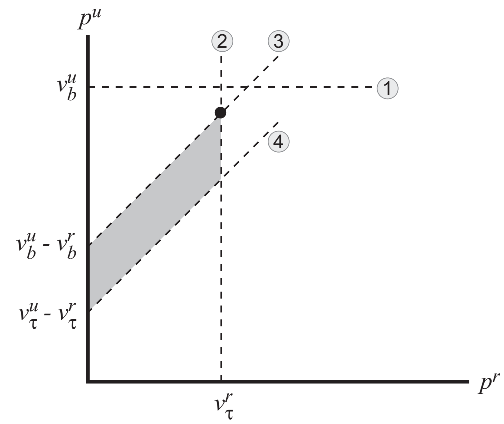

Figure 5.1: The set of prices satisfying the four constraints, (5.3)–(5.6), is the gray region. Line ① corresponds to constraint (5.3). Line ② corresponds to constraint (5.4). Line ③ corresponds to constraint (5.5). Line ④ corresponds to constraint (5.6).

where the last inequality follows because $v_b^\kappa > v_\tau^\kappa$. From this, we see that (5.5) is a binding constraint, while (5.3) is slack. If (5.5) is binding, it's clear that (5.6) is slack. Clearly, we can't set $p^r = \infty$, so we can conclude that (5.4) is binding.

Summary 4 *The incentive compatibility constraint* (5.5) *for the business traveler is binding, but the participation constraint* (5.3) *is not. The incentive compatibility constraint* (5.6) *for the tourist is slack, but the participation constraint* (5.4) *is binding.*

If you think about it intuitively, it is the business traveler who wishes to keep his type from the airline. His knowledge, that he is a business traveler, is valuable information because he's the one who must be induced to reveal his information; that is, he's the one who would have incentive to pretend to be the low-willingness-to-pay type. Hence, it is not surprising that his revelation constraint is binding. Along the same lines, the tourist has no incentive to keep his type from the airline—he would prefer the airline know he has a low willingness to pay. Hence, his revelation constraint is not binding, only his participation constraint is. These are general insights: The type who wishes to conceal information (has valuable information) has a binding revelation constraint and the type who has no need to conceal his information has just a binding participation constraint.

As summarized above, we have just two binding constraints and two unknown parameters, p^r and p^u. We can, thus, solve the maximization problem by solving the two binding constraints. This yields $p_*^r = v_\tau^r$ and $p_*^u = \ell(p_*^r) = v_b^u - \left(v_b^r - v_\tau^r\right)$. Note that the tourist gets no surplus, but the business traveler enjoys $v_b^r - v_\tau^r > 0$ of surplus. This is a general result: The type with the valuable information enjoys some return from having it. This is known as his or her *information rent*. It is no surprise that a type whose information lacks value does not capture any return from it.

Summary 5 *The business traveler enjoys an information rent. The tourist does not.*

Under price discrimination, the airline's profit is

$$N_b \left(v_b^u - \left(v_b^r - v_\tau^r\right)\right) + N_\tau v_\tau^r \,. \tag{5.8}$$

It is clear that (5.8) is dominated by uniform pricing if either N_b or N_τ gets sufficiently small relative to the other. But provided that's not the case, then (5.8)—that is, second-degree price discrimination—can dominate. For instance, if $v_b^u = 500$, $v_b^r = 200$, $v_\tau^u = 200$, and $v_\tau^r = 100$, then $p_*^r = 100$ and $p_*^u = 400$. If $N_b = 60$ and $N_r = 70$, then the two possible uniform prices, v_b^u and v_r^u, yield profits of \$30,000 and \$26,000, respectively; but price discrimination yields a profit of \$31,000.

Observe, too, that, *in this example*, going from a world of profit-maximizing uniform pricing to second-degree price discrimination raises welfare — the tourists would not get to fly under the profit-maximizing uniform price (\$500), but would with price discrimination. Given that tourists value even a restricted ticket more than the marginal cost of flying them (\$0), getting them on board must increase welfare.

5.2 Quantity Discounts

Consider two consumer types, 1 and 2, indexed by θ. Assume the two types occur equally in the population. Assume that each consumer has quasi-linear utility

$$v(x, \theta) - T \,,$$

where x is consumption of a good and T is the payment (transfer) from the consumer to the seller of that good. Assume the following order condition on marginal utility

$$\frac{\partial}{\partial \theta}\left(\frac{\partial v(x,\theta)}{\partial x}\right) > 0 \,. \tag{5.9}$$

Expression (5.9) is called a *Spence-Mirrlees condition*; it is a single-crossing condition. As noted in the previous section, we often impose such an order assumption on the steepness of the indifference curves across types. Another way to state (5.9) is that the marginal utility of consumption is increasing in type for all levels of consumption.

Assume, as is standard, that the reservation utility if 0 units are purchased is the same for both types:

$$v(0, 1) = v(0, 2) \,. \tag{5.10}$$

For convenience assume a constant marginal cost, c. Given this, we can consider the seller's optimal strategy against a representative customer, who is, as previously assumed, as likely to be type 1 as type 2.

In analyzing this problem, we can view the seller's problem as one of designing two "packages." One package will have x_1 units of the good and be sold for T_1 and the other will have x_2 units and be sold for T_2. Obviously, the x_θ-unit package is intended for the type-θ consumer. (One can think of these as being different size bottles of soda with x_θ as the number of liters in the θ bottle.) Hence, the seller's problem is

$$\max_{\{x_1,x_2,T_1,T_2\}} \frac{1}{2}(T_1 - cx_1) + \frac{1}{2}(T_2 - cx_2) \tag{5.11}$$

subject to participation (IR) constraints,

$$v(x_1, 1) - T_1 \geq 0 \text{ and} \tag{5.12}$$

$$v(x_2, 2) - T_2 \geq 0\,, \tag{5.13}$$

and subject to revelation (IC) constraints,

$$v(x_1, 1) - T_1 \geq v(x_2, 1) - T_2 \text{ and} \tag{5.14}$$

$$v(x_2, 2) - T_2 \geq v(x_1, 2) - T_1\,. \tag{5.15}$$

As is often true of mechanism-design problems, it is easier here to work with net utility (in this case, consumer surplus) rather than payments. To that end, let

$$U_\theta = v(x_\theta, \theta) - T_\theta\,.$$

Also define

$$\begin{aligned} I(x) &= v(x, 2) - v(x, 1) \\ &= \int_1^2 \frac{\partial v(x,t)}{\partial \theta} dt\,. \end{aligned}$$

Observe then, that

$$I'(x) = \int_1^2 \frac{\partial^2 v(x,t)}{\partial \theta \partial x} dt > 0\,,$$

where the inequality follows from the Spence-Mirrlees condition (5.9). Note, given condition (5.10), this also implies $I(x) > 0$ if $x > 0$. The use of the letter "I" for this function is not accidental; it is, as we will see, related to the information rent that the type-2 consumer enjoys.

We can rewrite the constraints (5.12)–(5.15) as

$$U_1 \geq 0 \tag{5.16}$$

$$U_2 \geq 0 \tag{5.17}$$

$$U_1 \geq U_2 - I(x_2) \text{ and} \tag{5.18}$$

$$U_2 \geq U_1 + I(x_1)\,. \tag{5.19}$$

We can also rewrite the seller's problem (5.11) as

$$\max_{\{x_1,x_2,U_1,U_2\}} \frac{1}{2}(v(x_1,1)-U_1-cx_1)+\frac{1}{2}(v(x_2,2)-U_2-cx_2)\,. \qquad (5.20)$$

We could solve this problem by assigning four Lagrange multipliers to the four constraints and crank through the problem. This, however, would be way tedious and, moreover, not much help for developing intuition. So let's use a little logic first.

- Unless $x_1 = 0$, one or both of (5.18) and (5.19) must bind. To see this, suppose neither was binding. Then, since the seller's profits are decreasing in U_θ, she would make both U_1 and U_2 as small as possible, which is to say 0. But given $I(x_1) > 0$ if $x_1 > 0$, this would violate (5.19).
- Observe that (5.18) and (5.19) can be combined so that $I(x_2) \geq U_2 - U_1 \geq I(x_1)$. Ignoring the middle term for the moment, the fact that $I(\cdot)$ is increasing means that $x_2 \geq x_1$. Moreover, if $x_1 > 0$, then $U_2 - U_1 \geq I(x_1) > 0$. Hence $U_2 > 0$, which means (5.17) is slack.
- Participation constraint (5.16), however, must bind at the seller's optimum. If it didn't, then there would exist an $\varepsilon > 0$ such that U_1 and U_2 could both be reduced by ε without violating (5.16) or (5.17). Since such a change wouldn't change the difference in the Us, this change also wouldn't lead to a violation of the IC constraints, (5.18) and (5.19). But from (5.20), lowering the Us by ε increases profit, so we're not at an optimum if (5.16) isn't binding.
- We've established that, if $x_1 > 0$, then (5.16) binds, (5.17) is slack, and at least one of (5.18) and (5.19) binds. Observe that we can rewrite the IC constraints as $I(x_2) \geq U_2 \geq I(x_1)$. The seller's profit is greater the *smaller* is U_2, so it is the lower bound in this last expression that is important. That is, (5.19) binds. Given that $I(x_2) \geq I(x_1)$, as established above, we're free to ignore (5.18).

So our reasoning tells us that, provided $x_1 > 0$, we need only pay attention to two constraints, (5.16) and (5.19). Using them to solve for U_1 and U_2, we can turn the seller's problem into the following *un*constrained problem:

$$\max_{\{x_1,x_2\}} \frac{1}{2}(v(x_1,1)-cx_1)+\frac{1}{2}(v(x_2,2)-I(x_1)-cx_2)\,. \qquad (5.21)$$

The first-order conditions are:

$$\frac{\partial v(x_1^*,1)}{\partial x} - I'(x_1^*) - c = 0 \qquad (5.22)$$

$$\frac{\partial v(x_2^*,2)}{\partial x} - c = 0\,. \qquad (5.23)$$

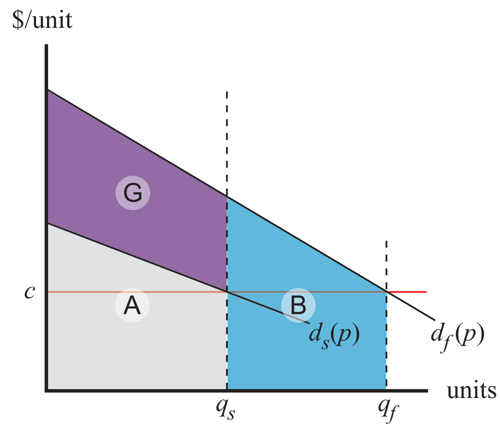

Figure 5.2: The individual demands of the two types of consumers (family and single), $d_f(\cdot)$ and $d_s(\cdot)$, respectively, are shown. Under the ideal *third*-degree price discrimination scheme, a single would buy a package with q_s units and pay an amount equal to area A (gray area). A family would buy a package with q_f units and pay an amount equal to the sum of all the shaded areas (A, B, and G).

Note that (5.23) is the condition for maximizing welfare were the seller selling only to type-2 customers; that is, we have efficiency in the type-2 "market." Because, however, $I'(\cdot) > 0$, we don't have the same efficiency *vis-à-vis* type-1 customers; in the type-1 "market," we see too little output relative to welfare-maximizing amount. This is a standard result—*efficiency at the top* and *distortion at the bottom*.

To make this more concrete, suppose $v(x, \theta) = 5(\theta + 1)\ln(x + 1)$ and $c = 1$. Then $x_2^* = 14$ and $x_1^* = 4$. Consequently, $T_1 \approx 16.1$ and $T_2 = v(x_2^*, 2) - I(x_1^*) \approx 32.6$. *Note the quantity discount:* A type-2 consumer purchases more than three times as much, but pays only roughly twice as much as compared to a type-1 consumer.

5.3 A Graphical Approach to Quantity Discounts

Now we consider an alternative, but ultimately equivalent, analysis of quantity discounts.

Consider a firm that produces some product. Continue to assume the marginal cost of production is constant, c. Suppose the population of potential buyers is divided into families (indexed by f) and single people (indexed by s). Let $d_f(\cdot)$ denote the demand of an *individual* family and let $d_s(\cdot)$ denote the demand of an *individual* single. Figure 5.2 shows the two demands. Note that, at any price, a family's demand exceeds a single's demand.

The ideal would be if the firm could engage in *third*-degree price discrimination by offering two different two-part tariffs to the two populations. That is, if the firm could freely identify singles from families, it would sell to each member of each group the quantity that equated that member's relevant inverse demand to cost (*i.e.,* q_s or q_f in Figure 5.2 for a single or a family, respectively). It could make the per-unit charge c and the entry fee the respective consumer surpluses. Equivalently—and more practically—the firm could use packaging. The package for singles would have q_s units and sell for a single's total benefit, $b_s(q_s)$. This is the area labeled A in Figure 5.2. Similarly, the family package would have q_f units and sell for a family's total benefit of $b_f(q_f)$. This is the sum of the three labeled areas in Figure 5.2.

The ideal is not, however, achievable. The firm can*not* freely distinguish singles from families. It must *induce* revelation; that is, it must devise a *second*-degree scheme. Observe that the third-degree scheme won't work as a second-degree scheme. Although a single would still purchase a package of q_s units at $b_s(q_s)$, a family would not purchase a package of q_f units at $b_f(q_f)$. Why? Well, were the family to purchase the latter package it would, by design, earn no consumer surplus. Suppose, instead, it purchased the package intended for singles. Its total benefit from doing so is the sum of areas A and G in Figure 5.2. It pays $b_s(q_s)$, which is just area A, so it would enjoy a surplus equal to area G. In other words, the family would deviate from the intended package, with q_f units, which yields it no surplus, to the unintended package, with q_s units, which yields it a positive surplus equal to area G.

Observe that the firm could induce revelation—that is, get the family to buy the intended package—if it cut the price of the q_f-unit package. Specifically, if it reduced the price to the sum of areas A and B, then a family would enjoy a surplus equal to area G whether it purchased the q_s-unit package (at price = area A) or it purchased the intended q_f-unit package (at price = area A + area B). Area G is a family's information rent.

Although that scheme induces revelation, it is not necessarily the *profit-maximizing* scheme. To see why, consider Figure 5.3. Suppose that the firm reduced the size of the package intended for singles. Specifically, suppose it reduced it to $\hat{q}_s$ units, where $\hat{q}_s = q_s - h$. Given that it has shrunk the package, it would need to reduce the price it charges for it. The benefit that a single would derive from $\hat{q}_s$ units is the area beneath its inverse demand curve between 0 and $\hat{q}_s$ units; that is, the area labeled A′. Note that the firm is forgoing revenues equal to area J by doing this. But the surplus that a family could get by purchasing a $\hat{q}_s$-unit package is also smaller; it is now the area labeled G′. This means that the firm could raise the price of the q_f-unit package by the area labeled H. Regardless of which package it purchases, a family can only keep surplus equal to area G′. In other words, by reducing the quantity sold to the "low type" (a single), the firm reduces the information rent captured by the "high type" (a family).

Is it worthwhile for the firm to trade area J for area H? Observe that the

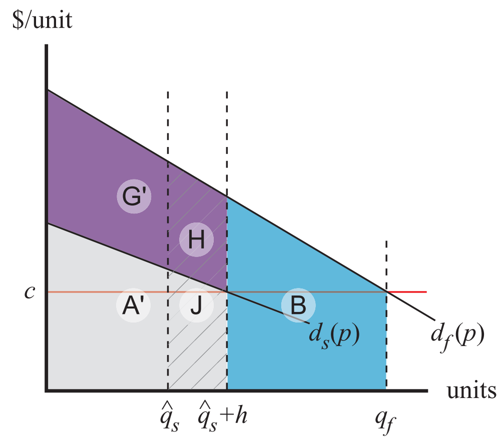

Figure 5.3: By reducing the quantity in the package intended for singles, the firm loses revenue equal to area J, but gains revenue equal to area H.

profit represented by area J is rather modest: While selling the additional h units to a single adds area J in revenue it also adds ch in cost. As drawn, the profit from the additional h units is the small triangle at the top of area J. In contrast, area H represents pure profit—regardless of how many it intends to sell to singles, the firm is selling q_f units to each family (*i.e.,* cq_f is a sunk expenditure with respect to how many units to sell each single). So, as drawn, this looks like a very worthwhile trade for the firm to make.

One caveat, however: The figure only compares a single family against a single single. What if there were lots of singles relative to families? Observe that the total net loss of reducing the package intended for singles by h is

$$(\text{area J} - ch) \times N_s ,$$

where N_s is the number of singles in the population. The gain from reducing that package is

$$\text{area H} \times N_f ,$$

where N_f is the number of families. If N_s is much larger than N_f, then this reduction in package size is not worthwhile. On the other hand if the two populations are roughly equal in size or N_f is larger, then reducing the package for singles by more than h could be optimal.

How do we determine the amount by which to reduce the package intended for singles (*i.e.,* the smaller package)? That is, how do we figure out what h should be? As usual, the answer is that we fall back on our $MR = MC$ rule. Consider a small expansion of the smaller package from $\hat{q}_s$. Because we are using an implicit two-part tariff (packaging) on the singles, the change in revenue—that is, marginal revenue—is the change in a single's benefit (*i.e.,* $mb_s(\hat{q}_s)$) times the number of singles. That is,

$$MR(\hat{q}_s) = N_s mb_s(\hat{q}_s)\,.$$

Recall that the marginal benefit schedule is inverse demand. So if we let $\rho_s(\cdot)$ denote the inverse *individual* demand of a single (*i.e.,* $\rho_s(\cdot) = d_s^{-1}(\cdot)$), then we can write

$$MR(\hat{q}_s) = N_s \rho_s(\hat{q}_s)\,. \tag{5.24}$$

What about MC? Well, if we increase the amount in the smaller package we incur costs from two sources. First, each additional unit raises production costs by c. Second, we increase each family's information rent (*i.e.,* area H shrinks). Observe that area H is the area between the two demand curves (thus, between the two *inverse* demand curves) between $\hat{q}_s$ and $\hat{q}_s + h$. This means that the marginal reduction in area H is

$$\rho_f(\hat{q}_s) - \rho_s(\hat{q}_s)\,,$$

where $\rho_f(\cdot)$ is the inverse demand of an individual family. Adding them together, and scaling by the appropriate population sizes, we have

$$MC(\hat{q}_s) = N_s c + N_f\big(\rho_f(\hat{q}_s) - \rho_s(\hat{q}_s)\big)\,. \tag{5.25}$$

Some observations:

1. Observe that if we evaluate expressions (5.24) and (5.25) at the q_s shown in Figure 5.2, we have

   $$\begin{aligned} MR(q_s) &= N_s \rho_s(q_s) \text{ and} \\ MC(q_s) &= N_s c + N_f\big(\rho_f(q_s) - \rho_s(q_s)\big)\,. \end{aligned}$$

   Subtract the second equation from the first:

   $$\begin{aligned} MR(q_s) - MC(q_s) &= N_s(\rho_s(q_s) - c) - N_f\big(\rho_f(q_s) - \rho_s(q_s)\big) \\ &= -N_f\big(\rho_f(q_s) - \rho_s(q_s)\big) \\ &< 0\,, \end{aligned}$$

   where the second equality follows because, as seen in Figure 5.2, $\rho_s(q_s) = c$ (*i.e.,* q_s is the quantity that equates inverse demand and marginal cost). Hence, provided $N_f > 0$, we see that the profit-maximizing second-degree pricing scheme sells the low type (*e.g.,* singles) less than the welfare-maximizing quantity (*i.e.,* there is a deadweight loss of area J $- ch$). In other words, as we saw previously, there is *distortion at the bottom*.

2. How do we know we want the family package to have q_f units? Well, clearly we wouldn't want it to have more—the marginal benefit we could capture would be less than our marginal cost. If we reduced the package size, we would be creating deadweight loss. Furthermore, because we don't

have to worry about singles' buying packages intended for families (that incentive compatibility constraint is slack) we can't gain by creating such a deadweight loss (unlike with the smaller package, where the deadweight loss is offset by the reduction in the information rent enjoyed by families). We can summarize this as there being no distortion at the top.

3. Do we know that the profit-maximizing $\hat{q}_s$ is positive? That is, do we know that a solution to $MR = MC$ exists in this situation? The answer is no. It is possible, especially if there are a lot of families relative to singles, that it might be profit-maximizing to set $\hat{q}_s = 0$; that is, sell only one package, the q_f-unit package, which only families buy. This will be the case if $MR(0) \leq MC(0)$. In other words, if

$$N_s(\rho_s(0) - c) - N_f\big(\rho_f(0) - \rho_s(0)\big) \leq 0 \tag{5.26}$$

4. On the other hand, it will often be the case that the profit-maximizing $\hat{q}_s$ is positive, in which case it will be determined by equating expressions (5.24) and (5.25).

Extended Example

Consider a cell-phone-service provider. It faces two types of customers, those who seldom have someone to talk to (indexed by s) and those who frequently have someone to talked to (indexed by f). Within each population, customers are homogeneous. The marginal cost of providing connection to a cell phone is 5 cents a minute (for convenience, all currency units are cents). A member of the s-population has demand:

$$d_s(p) = \begin{cases} 450 - 10p\,, & \text{if } p \leq 45 \\ 0\,, & \text{if } p > 45 \end{cases}\,.$$

A member of the f-population has demand:

$$d_f(p) = \begin{cases} 650 - 10p\,, & \text{if } p \leq 65 \\ 0\,, & \text{if } p > 65 \end{cases}\,.$$

There are 1,000,000 f-type consumers. There are N_s s-type consumers.

What is the profit-maximizing second-degree pricing scheme to use? How many minutes are in each package? What are the prices?

It is clear that the f types are the high types (like families in our previous analysis). There is no distortion at the top, so we know we sell an f type the number of minutes that equates demand and marginal cost; that is,

$$q_f^* = d_f(c) = 600\,.$$

We need to find q_s^*. To do this, we need to employ expressions (5.24) and (5.25). They, in turn, require us to know $\rho_s(\cdot)$ and $\rho_f(\cdot)$. Considering the regions of positive demand, we have:

$$q_s = 450 - 10\rho_s(q_s) \text{ and}$$
$$q_f = 650 - 10\rho_f(q_f)\,;$$

hence,

$$\rho_s(q_s) = 45 - \frac{q_s}{10} \text{ and}$$
$$\rho_f(q_f) = 65 - \frac{q_f}{10}\,.$$

Using expression (5.24), marginal revenue from q_s is, therefore,

$$MR(q_s) = N_s\rho_s(q_s) = N_s \times \left(45 - \frac{q_s}{10}\right)\,.$$

Marginal cost of q_s (including forgone surplus extraction from the f type) is

$$\begin{aligned} MC(q_s) &= N_s c + N_f\big(\rho_f(q_s) - \rho_s(q_s)\big) \\ &= 5N_s + 1{,}000{,}000\left(65 - \frac{q_f}{10} - 45 + \frac{q_f}{10}\right) \\ &= 5N_s + 20{,}000{,}000\,. \end{aligned}$$

Do we want to shut out the s-types altogether? Employing expression (5.26), the answer is yes if

$$40N_s - 20{,}000{,}000 < 0\,;$$

that is, if

$$N_s < 500{,}000\,.$$

So, if $N_s < 500{,}000$, then $q_s^* = 0$ and the price for 600 minutes (*i.e.,* q_f^*) is $b_f(600)$, which is

$$\begin{aligned} b_f(600) &= \text{Area under } \rho_f(\cdot) \text{ from 0 to 600} \\ &= \int_0^{600} \rho_f(z)dz \\ &= \int_0^{600} \left(65 - \frac{z}{10}\right) dz \\ &= 21{,}000 \end{aligned}$$

cents or \$210.

Suppose that $N_s \geq 500{,}000$. Then, equating MR and MC, we have

$$N_s \times \left(45 - \frac{q_s}{10}\right) = 5N_s + 20{,}000{,}000\,;$$

hence,

$$q_s^* = 400 - \frac{200,000,000}{N_s} .$$

The low type retains no surplus, so the price for q_s^* minutes is $b_s(q_s^*)$, which equals the area under $\rho_s(\cdot)$ from 0 to q_s^*. This can be shown (see derivation of $b_f(600)$ above) to be

$$\begin{aligned} b_s(q_s^*) &= \text{Area under } \rho_s(\cdot) \text{ from 0 to } q_s^* \\ &= \int_0^{q_s^*} \left(45 - \frac{q}{10}\right) dq \\ &= 45q_s^* - \frac{{q_s^*}^2}{20} . \end{aligned}$$

The price charged the f types for their 600 minutes is $b_f(600)$ less their information rent, which is the equivalent of area G' in Figure 5.3.

$$\text{Area } \mathsf{G}' = \int_0^{q_s^*} \left(65 - \frac{q}{45} - 45 + \frac{q}{45}\right) dq = 20q_s^* .$$

So the price charged for 600 minutes is $21,000 - 20q_s^*$ cents ($\$210 - q_s^*/5$).

To conclude: If $N_s < 500,000$, then the firm sells only a package with 600 minutes for \$210. In this case, only f types buy. If $N_s \geq 500,000$, then the firm sells a package with 600 minutes, purchased by the f types, for $210 - q_s^*/5$ dollars; and it also sells a package with q_s^* minutes for a price of $b_s(q_s^*)$ dollars. For example, if $N_s = 5,000,000$, then the two plans are (i) 600 minutes for \$138; and (ii) 360 minutes for \$97.20.

A General Analysis | 5.4

Although presented in different ways, the various forms of second-degree price discrimination are all examples of mechanism design. As such, they can be modeled in a general way.

Assume there is a measure of consumers, which we can take to be one without loss of generality. Each consumer has a type, θ, where $\theta \in [\underline{\theta}, \bar{\theta}]$. Types are distributed according to the distribution function $F : [\underline{\theta}, \bar{\theta}] \to [0, 1]$. Assume $F(\cdot)$ is differentiable and denote its derivative—the associated density function—by $f(\cdot)$. Assume $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ (*i.e.,* $f(\cdot)$ has *full support*).[2]

A consumer's utility is $u(q, \theta) + y$, where y is the numeraire good and $q \in \mathbb{R}_+$ is either the *quantity* of the other good—the good being sold by the firm engaging in second-degree price discrimination—or the *quality* of that good. In the latter case, assume the consumer buys, at most, a single unit of the good.

[2] A similar analysis could be carried out with a discrete type space, but not as straightforwardly.

Assume that the marginal utility from more q is increasing with type (one can think of this as the definition of type); specifically,

$$\frac{\partial}{\partial\theta}\frac{\partial u(q,\theta)}{\partial q} = \frac{\partial^2 u(q,\theta)}{\partial\theta\partial q} > 0 \tag{5.27}$$

for all q and θ.[3] Assume $u(0,\theta)$ is the same regardless of type; that is, $u(0,\theta) = u(0,\theta')$ for any θ and θ'. When q is quantity, this assumption is relatively innocuous. It could be more of an issue when q is quality—conceivably different types could derive different utility from the lowest-quality product. A way around that is to imagine that a product of quality 0 is the same as not buying the product at all. It is without further loss of generality to assume $u(0,\theta) = 0$ for all θ.

The seller incurs a cost of $c(q)$ to produce either a package with q units in it or a product with quality q. Assume $c'(q) > 0$ for all q. For each type θ, we can consider the seller as determining a package or product, $q(\theta)$, that it will sell for $t(\theta)$ and that only θ-type consumers will buy. Because $q(\theta)$ and $t(\theta)$ could equal zero, there is no loss of generality in assuming all consumers "buy" in equilibrium (*i.e.,* we can consider those who don't buy as "buying" $q = 0$ at price of 0).

The firm's problem then is to choose $\langle q(\cdot), t(\cdot)\rangle$ to maximize profit

$$\int_{\underline{\theta}}^{\bar{\theta}} \Big(t(\theta) - c\big(q(\theta)\big)\Big) f(\theta) d\theta \tag{5.28}$$

subject to the constraint that consumers prefer to buy rather than not buy,

$$u\big(q(\theta),\theta\big) - t(\theta) \geq 0 \ \text{ for all } \theta\,, \tag{IR}$$

and subject to the constraint that each type of consumer buy the package or product intended for him,

$$u\big(q(\theta),\theta\big) - t(\theta) \geq u\big(q(\hat{\theta}),\theta\big) - t(\hat{\theta}) \ \text{ for all } \theta, \hat{\theta}\,. \tag{IC}$$

As written, this is a difficult optimization program to solve because of the nature of the incentive-compatibility (revelation) constraints (IC). To make progress, we need to do some analysis. Consider two types θ_1 and θ_2 with $\theta_1 < \theta_2$. The IC constraints imply:

$$u\big(q(\theta_1),\theta_1\big) - t(\theta_1) \geq u\big(q(\theta_2),\theta_1\big) - t(\theta_2) \ \text{ and}$$
$$u\big(q(\theta_2),\theta_2\big) - t(\theta_2) \geq u\big(q(\theta_1),\theta_2\big) - t(\theta_1)\,.$$

As is often the case in mechanism-design problems, it is easier to work with utilities than payments. To this end, define

$$v(\theta) = u\big(q(\theta),\theta\big) - t(\theta)\,.$$

[3] Expression (5.27) is the Spence-Mirrlees or single-crossing condition for this problem.

$v(\theta)$ is the θ-type consumer's *equilibrium* utility. Observe the utility a θ-type consumer gets from buying the product or package intended for the $\hat{\theta}$-type consumer can be written as

$$v(\hat{\theta}) - u\big(q(\hat{\theta}), \hat{\theta}\big) + u\big(q(\hat{\theta}), \theta\big) .$$

We can, therefore, write the pair of inequalities above as

$$\begin{aligned} v(\theta_1) &\geq v(\theta_2) - u\big(q(\theta_2), \theta_2\big) + u\big(q(\theta_2), \theta_1\big) \text{ and} \\ v(\theta_2) &\geq v(\theta_1) - u\big(q(\theta_1), \theta_1\big) + u\big(q(\theta_1), \theta_2\big) . \end{aligned}$$

These inequalities can be combined as

$$\int_{\theta_1}^{\theta_2} \frac{\partial u\big(q(\theta_1), z\big)}{\partial \theta} dz \leq v(\theta_2) - v(\theta_1) \leq \int_{\theta_1}^{\theta_2} \frac{\partial u\big(q(\theta_2), z\big)}{\partial \theta} dz . \tag{5.29}$$

Expression (5.29) has two consequences. First, ignoring the middle term, it implies

$$\int_{\theta_1}^{\theta_2} \int_{q(\theta_1)}^{q(\theta_2)} \frac{\partial^2 u(q, z)}{\partial q \partial \theta} dq\, dz \geq 0 .$$

From (5.27), we know the cross-partial derivative is everywhere positive. Given $\theta_1 < \theta_2$, this means the double integral can be non-negative only if $q(\theta_1) \leq q(\theta_2)$. Because θ_1 and θ_2 are arbitrary, we can conclude that the quantity/quality function, $q(\cdot)$, is *non-decreasing* in type. That is, if the second-degree price discrimination scheme is incentive compatible, then lower types cannot get more quantity/quality than higher types.

The second consequence of (5.29) is derived as follows. By fixing one end point (*i.e.,* θ_1 or θ_2) and letting the other converge towards it, we see that $v(\cdot)$ is absolutely continuous with respect to the Lebesgue measure and is, thus, almost everywhere differentiable. This derivative is

$$\frac{dv(\theta)}{d\theta} = \frac{\partial u\big(q(\theta), \theta\big)}{\partial \theta}$$

almost everywhere.[4] Integration reveals that

$$v(\theta) = v(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \frac{\partial u\big(q(z), z\big)}{\partial \theta} dz . \tag{5.30}$$

The analysis to this point has established (i) a necessary condition for a second-degree price discrimination scheme to be incentive compatible is that the quality/quantity function $q(\cdot)$ be non-decreasing; and (ii) another necessary

[4]Note the important distinction between $\frac{\partial u(q(\theta),\theta)}{\partial \theta}$ and $\frac{du(q(\theta),\theta)}{d\theta}$. The former is the *partial* derivative of u with respect to to second argument evaluated at $\big(q(\theta), \theta\big)$, while the latter is the total derivative of u.

condition for incentive compatibility is that (5.30) hold. The latter is equivalent, recalling the definition of $v(\cdot)$, to saying the payment function must satisfy

$$t(\theta) = t_L + u\big(q(\theta),\theta\big) - \int_{\underline{\theta}}^{\theta} \frac{\partial u\big(q(z),z\big)}{\partial\theta} dz\,, \tag{5.31}$$

where $t_L = -v(\underline{\theta})$ is a constant.

It turns out that (i) and (ii) are also *sufficient* conditions, which means, given our initial assumptions, that we can restrict attention without loss to pricing schemes in which $q(\cdot)$ is non-decreasing and $t(\cdot)$ is given by (5.31).

Proposition 8 *Under the assumptions of this section, a second-degree price discrimination scheme $\langle q(\cdot), t(\cdot)\rangle$ is incentive compatible if and only if $t(\cdot)$ is as stated in expression* (5.31) *and $q(\cdot)$ is non-decreasing.*

Proof: Necessity was established in the text. Consider a second-degree price-discrimination scheme that satisfies (5.31) (equivalently, (5.30)) and has $q(\cdot)$ non-decreasing. We need to establish that such a scheme induces a type-θ consumer to prefer to purchase $q(\theta)$ at $t(\theta)$ than to pursue another option. His utility if he purchases $q(\theta')$ at $t(\theta')$, $\theta' > \theta$, is

$$\begin{aligned}
&u\big(q(\theta'),\theta\big) - t(\theta') \\
&= -t_L + u\big(q(\theta'),\theta\big) - u\big(q(\theta'),\theta'\big) + \int_{\underline{\theta}}^{\theta} \frac{\partial u\big(q(z),z\big)}{\partial\theta} dz + \int_{\theta}^{\theta'} \frac{\partial u\big(q(z),z\big)}{\partial\theta} dz \\
&\qquad = u\big(q(\theta),\theta\big) - t(\theta) - \int_{\theta}^{\theta'} \left(\frac{\partial u\big(q(\theta'),z\big)}{\partial\theta} - \frac{\partial u\big(q(z),z\big)}{\partial\theta} \right) dz \\
&\qquad\qquad = v(\theta) - \int_{\theta}^{\theta'} \int_{q(z)}^{q(\theta')} \frac{\partial^2 u\big(q,z\big)}{\partial q \partial\theta} dq dz \leq v(\theta)\,,
\end{aligned}$$

where the third line follows from the second by adding and subtracting $u\big(q(\theta),\theta\big)$, the definition of $t(\theta)$, and recognizing that

$$u\big(q(\theta'),\theta\big) - u\big(q(\theta'),\theta'\big) = -\int_{\theta}^{\theta'} \frac{\partial u\big(q(\theta'),z\big)}{\partial\theta} dz\,;$$

the final inequality follows because $q(\cdot)$ is non-decreasing, so the direction of integration in the inner integral of the fourth line is the positive direction and, by (5.27), the function being integrated is positive. Because we established

$$u\big(q(\theta'),\theta\big) - t(\theta') \leq v(\theta)\,,$$

we have shown that the mechanism is incentive compatible insofar as a type θ would not choose a package/product intended for a higher type. **Exercise:** finish the proof of sufficiency for the case $\theta' < \theta$. ■

In light of Proposition 8, the firm's problem is to choose $\langle q(\cdot), t(\cdot)\rangle$ to maximize (5.28) subject to (IR) and $q(\cdot)$ non-decreasing with $t(\cdot)$ being given by (5.31). It is again easier to work with $v(\cdot)$ than $t(\cdot)$, where $v(\cdot)$ is given by (5.30). This makes the firm's problem:

$$\max_{\langle q(\cdot),v(\cdot)\rangle} \int_{\underline{\theta}}^{\bar{\theta}} \Big(\underbrace{u\big(q(\theta),\theta\big) \underbrace{- \int_{\underline{\theta}}^{\theta} \frac{\partial u\big(q(z),z\big)}{\partial\theta} dz - v(\underline{\theta})}_{-v(\theta)}}_{t(\theta)} - c\big(q(\theta)\big)\Big) f(\theta) d\theta \quad (5.32)$$

subject to

$$v(\theta) \geq 0 \;\forall\theta \quad (5.33)$$

and

$$q(\cdot) \text{ non-decreasing} \quad (5.34)$$

Observe that

$$u\big(q(\theta),\theta\big) - c\big(q(\theta)\big) \equiv w\big(q(\theta),\theta\big)$$

is the welfare generated by trade with a type-θ consumer. Comparing that expression with (5.32), we see that the firm's objective is quite different than welfare maximization. Consequently, as will be verified below, we should not expect the firm's choice of a second-degree price-discrimination scheme to maximize welfare.

Using the above definition of $w(\cdot,\cdot)$, the firm's objective function can be rewritten as

$$\begin{aligned} &\int_{\underline{\theta}}^{\bar{\theta}} \left(w\big(q(\theta),\theta\big) - \int_{\underline{\theta}}^{\theta} \frac{\partial u\big(q(z),z\big)}{\partial\theta} dz \right) f(\theta) d\theta - v(\underline{\theta}) \\ &\qquad = \int_{\underline{\theta}}^{\bar{\theta}} \left(w\big(q(\theta),\theta\big) - \frac{1-F(\theta)}{f(\theta)} \frac{\partial u\big(q(\theta),\theta\big)}{\partial\theta} \right) f(\theta) d\theta - v(\underline{\theta}) \,, \end{aligned} \quad (5.35)$$

where the second line follows from integration by parts. The expression

$$\frac{1-F(\theta)}{f(\theta)} \quad (5.36)$$

is the multiplicative inverse of the hazard rate (this inverse is sometimes called the Mills ratio). Many distributions have the property that their hazard rate is non-decreasing (among these distributions are the uniform, the normal, and the exponential). If the hazard rate is non-decreasing, then expression (5.36) is non-increasing in θ. We give a name to this property:

Assumption 1 (MHRP) *Assume the hazard rate associated with the distribution of types exhibits the monotone hazard rate property, which is to say it is a non-decreasing function.*

We will assume, henceforth, that $F(\cdot)$ satisfies the MHRP; that is, that expression (5.36) is non-increasing in θ.

Lemma 6 *The function $u(q,\cdot)$ is increasing for $q>0$.*

Proof: From (5.27)

$$\frac{\partial u(q,\theta')}{\partial q} > \frac{\partial u(q,\theta)}{\partial q}$$

for all q whenever $\theta' > \theta$. Hence,

$$\int_0^q \left(\frac{\partial u(z,\theta')}{\partial q} - \frac{\partial u(z,\theta)}{\partial q}\right) dz > 0$$

for all q. Hence,

$$0 < \Big(u(q,\theta') - u(q,\theta)\Big) - \Big(u(0,\theta') - u(0,\theta)\Big) = u(q,\theta') - u(q,\theta)\,,$$

where the equality follows because, recall, we assumed $u(0,\theta)=0$ for all θ. ■

In light of Lemma 6, we have $\partial u(q,\theta)/\partial\theta > 0$ for all $q>0$. Consequently, from (5.30), we have $v(\theta) \geq v(\underline{\theta})$ for all $\theta > \underline{\theta}$. This means that if the IR constraint is satisfied for type $\underline{\theta}$ it is satisfied for all types. Because, from (5.35), the firm's profits are decreasing in $v(\underline{\theta})$, the firm wants to make $v(\underline{\theta})$ as small as possible. Given the IR constraint, this means that the firm sets $v(\underline{\theta})=0$.

Corollary 2 *Optimally, the firm sets $v(\underline{\theta})=0$. The IR constraint is met for all types.*

Observe this portion of the analysis reveals that the lowest type keeps no consumer surplus; a result that is consistent with our analysis in earlier sections.

Having set $v(\underline{\theta})=0$, the firm's problem boils down to maximizing (5.35) subject to the constraint that $q(\cdot)$ be non-decreasing. In general, this is an optimal-control problem. Fortunately, in many contexts, the problem can be solved simply by maximizing (5.35) pointwise; that is, solve

$$\max_q w\big(q,\theta\big) - \frac{1-F(\theta)}{f(\theta)}\,\frac{\partial u\big(q,\theta\big)}{\partial\theta}\,. \tag{5.37}$$

Provided the solutions yield a non-decreasing $q(\cdot)$, we're done. Observe, except when $\theta=\bar{\theta}$ (and, thus, $1-F(\theta)=0$), the solution to (5.37) will not maximize welfare. In particular, from the Spence-Mirrlees condition—that is, expression (5.27)—we know that derivative of (5.37) lies below the derivative of $w(\cdot,\theta)$ for all $\theta<\bar{\theta}$; hence, the solution to (5.37) will be less than the welfare-maximizing solution. To conclude:

Proposition 9 *The quantity/quality provided the highest type is the welfare-maximizing amount (no distortion at the top). For all other types, however, the quantity/quality provided is less than the welfare-maximizing amount (downward distortion below the top).*

A quantity-discount example

Suppose that $c(q) = cq$, $c > 0$, and $u(q, \theta) = \theta \log(q + 1)$. Observe the welfare-maximizing package for a type-θ consumer is $(\theta - c)/c$. Suppose that types are distributed uniformly from $c + \gamma$ to $c + \gamma + 1$, $\gamma \geq 0$. Observe $f(\theta) = 1$ and $1 - F(\theta) = c + \gamma + 1 - \theta$. Expression (5.37) is, thus,

$$\theta \log(q+1) - cq - (c+\gamma+1-\theta)\log(q+1) = 2\theta \log(q+1) - cq - (c+\gamma+1)\log(q+1)$$

The maximum is

$$q(\theta) = \begin{cases} 0\,, & \text{if } \theta \leq c + \frac{\gamma+1}{2} \\ \frac{2\theta - 2c - \gamma - 1}{c}\,, & \text{if } \theta > c + \frac{\gamma+1}{2} \end{cases}.$$

Because $q(\cdot)$ is non-decreasing, this is indeed the optimum for the firm. From (5.31) we have

$$t(\theta) = \theta \log\big(q(\theta) + 1\big) - \int_{c+\gamma}^{\theta} \log\big(q(z) + 1\big) dz$$

Some observations, if $\gamma < 1$, then some types are shut out of the market; that is, the firm does not offer a package that the lowest types wish to purchase.

If we assume $c = \gamma = 1$, then all types except $\theta = 2$ are served; the size of their package is $2\theta - 4$. A type-θ consumer pays

$$\theta - 2 + \frac{3}{2}\log(2\theta - 3)\,.$$

Dividing the price of each package by its quantity (*i.e.,* calculating $t(\theta)/q(\theta)$) we get

$$\frac{1}{2} + \frac{3\log(2\theta - 3)}{4\theta - 8}\,. \tag{5.38}$$

It is readily verified that (5.38) is a decreasing function of θ; that is, there are quantity discounts.

A quality-distortion example

Suppose that the cost of a product with quality q is $q^2/2$. Suppose types are distributed uniformly on $[1, 2]$. Finally, suppose that $u(q, \theta) = \theta q$. It is readily checked that the welfare-maximizing quality to offer a type-θ consumer is $q = \theta$. The firm's problem, expression (5.37), is

$$\max_q \theta q - \frac{q^2}{2} - (2 - \theta)q \equiv \max_q 2(\theta - 1)q - \frac{q^2}{2}\,.$$

The solution is $q(\theta) = 2(\theta - 1)$, which is increasing in θ and, thus, a valid solution. Observe that type 1 is just shut out (*i.e.,* $q(1) = 0$). Note there is downward distortion, except at the top: $2(\theta - 1) < \theta$ for all $\theta < 2$ and $2(\theta - 1) = \theta$ if $\theta = 2$. From (5.31),

$$t(\theta) = \theta\big(2(\theta - 1)\big) - \int_1^{\theta} 2(z - 1)dz = \theta^2 - 1\,.$$

Mechanism Design

Purpose

Our purpose is to consider the problem of hidden information; that is, a game between two economic actors, one of whom possesses *mutually* relevant information that the other does not. This is a common situation: The classic example—covered in the previous part of these lecture notes—being the "game" between a monopolist, who doesn't know the consumer's willingness to pay, and the consumer, who obviously does. Within the realm of contract theory, relevant situations include a seller who is better informed than a buyer about the cost of producing a specific good; an employee who alone knows the difficulty of completing a task for his employer; a divisional manager who can conceal information about his division's investment opportunities from headquarters; and a leader with better information than her followers about the value of pursuing a given course of action. In each of these situations, having private information gives the player possessing it a potential strategic advantage in his dealings with the other player. For example, consider a seller who has better information about his costs than his buyer. By behaving as if he had high costs, the seller can seek to induce the buyer to pay him more than she would if she knew he had low costs. That is, he has an incentive to use his superior knowledge to capture an *"information rent."* Of course, the buyer is aware of this possibility; so, if she has the right to propose the contract between them, she will propose a contract that works to reduce this information rent. Indeed, how the contract proposer—the principal—designs contracts to mitigate the informational *dis*advantage she faces will be a major focus of this part of the lecture notes.

Bibliographic Note

This part of the lecture notes draws heavily from a set of notes that I co-authored with Bernard Caillaud.

Not surprisingly, given the many applications of the screening model, this coverage cannot hope to be fully original. The books by Laffont and Tirole (1993) and Salanié (1997) include similar chapters. Surveys have also appeared in journals (*e.g.,* Caillaud et al., 1988). Indeed, while there are idiosyncratic aspects to the approach pursued here, the treatment is quite standard.

The Basics of Contractual Screening | 6

Let us begin by broadly describing the situation in which we are interested. We shall fill in the blanks as we proceed through this reading.

- Two players are involved in a strategic relationship; that is, each player's well being depends on the play of the other player.

- One player is better informed (or will become better informed) than the other; that is, he has *private information* about some state of nature relevant to the relationship. As is typical in information economics, we will refer to the player with the private information as the *informed player* and the player without the private information as the *uninformed player*.

- Critical to the analysis of these situations is the bargaining game that determines the contract. Call the contract proposer the *principal* and call the player who receives the proposal the *agent*. Moreover, assume contracts are proposed on a take-it-or-leave-it (TIOLI) basis: The agent's only choices are to accept or reject the contract proposed by the principal. Rejection ends the relationship between the players. A key assumption is that the principal is the *un*informed player. Models like this, in which the uninformed player proposes the contract, are referred to as *screening models*. In contrast, were the informed player the contract proposer, we would have a type of signaling model.

- A contract can be seen as setting the rules of a secondary game to be played by the principal and the agent.

The *asymmetry* of information that exists in this game results because prior experience or expertise, location, or other factors give the agent free access to information about the state of nature; while the absence of expertise, different experience or location, or other factors exclude the principal from this information (make it prohibitively expensive for her to acquire it). For example, past jobs may tell a seller how efficient he is—and thus what his costs will be—while ignorance of these past jobs means the buyer has a less precise estimate of what his costs will be. The reason for this asymmetry of information is taken to be *exogenous*. In particular, the informed player is simply assumed to be endowed with his information. Here, note, only one player is better informed; that is, we are ruling *out* situations where *each* player has his or her own private information.[1]

Given this information structure, the two parties interact according to some specified rules that constitute the extensive form of a game. In this two-person game, the players must contract with each other to achieve some desired outcome. In particular, there is no ability to rely on some exogenously fixed and anonymous market mechanism. My focus will be on instances of the game where the informed player can potentially benefit from his informational advantage (*e.g.,* perhaps inducing a buyer to pay more for a good than necessary because she fears the seller is high cost). But, because the informed player doesn't have the first move—the uninformed player gets to propose the contract—this informational advantage is not absolute: Through her design of the contract, the uninformed player will seek to offset the informed player's inherent advantage.

[1]Put formally, the uninformed player's information partition is coarser than the informed player's information partition.

The Two-type Screening Model | 7

I will begin to formalize these ideas in as simple a model as possible, namely the *two-type model.* In the two-type model, the state of nature can take one of two possible values. As is common in this literature, I will refer to the realized state of nature as the agent's *type*. As there are only two possible states, the agent can have one of just two types.

Before proceeding, however, it needs to be emphasized that such simplicity in modeling is not without cost. The two-type model is "treacherous," in so far as it may suggest conclusions that seem general, but which are not. For example, the conclusion that we will shortly reach with this model that the optimal contract implies distinct outcomes for distinct states of nature—a result called *separation*—is not as general as it may seem. Moreover, the assumption of two types conceals, in essence, a variety of assumptions that must be made clear. It similarly conceals the richness of the screening problem in complex, more realistic, relationships. Few *economic* prescriptions and predictions should be reached from considering just the two-type model. Keeping this admonition in mind, we now turn to a simple analysis of private procurement in a two-type model.

A simple two-type screening situation | 7.1

A large retailer (the principal) wishes to purchase units of some good for resale. Assume its size gives it all the bargaining power in its negotiations with the one firm capable of supplying this product (the agent). Let $x \in \mathbb{R}_+$ denote the units of this good and let $r(x)$ denote the retailer's revenues from x units.[1] Assume that $r(\cdot)$ is strictly concave and differentiable everywhere. Assume, too, that $r'(0) > 0$. (Because $r(\cdot)$ is a revenue function, $r(0) = 0$.)

The retailer is uncertain about the efficiency of the supplier. In particular, the retailer knows an *in*efficient supplier has production costs of $C_I(x)$, but an efficient supplier has production costs of $C_E(x)$. Let the retailer's prior belief be that the supplier is inefficient with probability f, where—reflecting its uninformed status—$0 < f < 1$. The supplier, in contrast, knows its type; that is, whether it is efficient or not.

[1] For convenience, assume that the retailer incurs no costs other than those associated with acquiring the x units from the supplier. Alternatively, one could simply imagine that these other costs have been subtracted from revenues, so that $r(x)$ is profit gross of the cost of purchasing the x units.

Assume $C_t(\cdot)$ is increasing, everywhere differentiable, and convex for both types, t. (Because $C_t(\cdot)$ is a cost function, we know $C_t(0) = 0$.) Consistent with the ideas that the two types correspond to different levels of efficiency, we assume $C_I'(x) > C_E'(x)$ for all $x > 0$—the inefficient type's marginal-cost schedule lies above the efficient type's. Observe, necessarily then, that $C_I(x) > C_E(x)$ for all $x > 0$.

The retailer and the supplier have to agree on the quantity, x, of the good to trade and on a payment, s, for this *total* quantity. Please note that s is *not* the per-unit price, but the payment for *all* x units. Profits for retailer and supplier are, then, $r(x) - s$ and $s - C_t(x)$ respectively. The retailer makes a take-it-or-leave-it offer, which the supplier must either accept or refuse. If the supplier refuses the offer, there is no trade and each firm's payoff from this "transaction" is zero. This outcome, no agreement, is equivalent to agreeing to trade 0 units for a 0 payment. Hence, it is without loss of generality to assume that the parties must reach some agreement in equilibrium.

We begin our analysis with the *benchmark* case of symmetric or full information. That is, for the moment, assume the retailer knew the supplier's type (*i.e.,* $f = 0$ or $f = 1$). We may immediately characterize the Pareto optimal allocation: x_t^F units are traded, where

$$x_t^F = \arg\max_{x \geq 0} \{r(x) - C_t(x)\}.$$

Because the problem is otherwise uninteresting if trade is never desirable, assume that $r'(0) > C_E'(0)$ so that, with an efficient supplier at least, some trade is desirable. Pareto optimality, also referred to as *ex post efficiency*, then reduces to

$$r'(x_E^F) = C_E'(x_E^F) \text{ and } \left[r'\left(x_I^F\right) - C_I'\left(x_I^F\right)\right] x_I^F = 0$$

(where we take the larger non-negative root of the second equation). As it is optimal to produce a greater amount when marginal costs are lower, our assumptions about $C_t'(\cdot)$ imply $0 \leq x_I^F < x_E^F$. In making its contract offer, the retailer sets $x = x_t^F$ and it offers a payment, s_t^F, no larger than necessary to induce the supplier to accept; that is, s_t^F satisfies

$$s_t^F - C_t(x_t^F) = 0.$$

Contracts under Incomplete Information 7.2

This symmetric-information (benchmark) solution collapses when the retailer is uninformed about the state of nature. To see why, suppose that the retailer offered the supplier its choice of $\langle x_E^F, s_E^F \rangle$ or $\langle x_I^F, s_I^F \rangle$ with the expectation that the supplier would choose the one appropriate to its type (*i.e.,* the first contract if it were efficient and the second if it were not). Observe that the retailer is relying on the supplier to honestly disclose its type. Suppose, moreover, that the true state of nature is E. By truthfully revealing that the state of nature is E, the supplier would just be compensated for its cost of supplying x_E^F units; that is, it would earn a profit of $s_E^F - C_E(x_E^F) = 0$. On the other hand, if the supplier pretends to have high costs—claims the state of nature is I—it receives compensation s_I^F, while incurring cost $C_E(x_I^F)$ for supplying x_I^F units. This yields the supplier a profit of

$$\begin{aligned} s_I^F - C_E\left(x_I^F\right) &= \\ C_I(x_I^F) - C_E(x_I^F) &> 0 \end{aligned}$$

(recall $s_I^F = C_I\left(x_I^F\right)$). Clearly, then, the efficient-type supplier cannot be relied on to disclose its type honestly.

This difference or profit, $C_I(x_I^F) - C_E(x_I^F)$, which motivates the supplier to lie, is called an *information rent*. This is a loss to the retailer but a gain to the supplier. There is, however, an additional loss suffered by the retailer that is *not* recaptured by the supplier: Lying means inefficiently little is produced; that is, a real deadweight loss of

$$\left[r\left(x_E^F\right) - C_E\left(x_E^F\right)\right] - \left[r\left(x_I^F\right) - C_E\left(x_I^F\right)\right]$$

is suffered.

Given this analysis, it would be surprising if the retailer would be so naïve as to rely on the supplier to freely reveal its type. In particular, we would expect the retailer to seek a means of improving on this *ex post in*efficient outcome by devising a more sophisticated contract. What kind of contracts can be offered? Because the retailer does *not* know the supplier's level of efficiency, it may want to delegate the choice of quantity to the supplier under a *payment schedule* that implicitly rewards the supplier for not pretending its costs are high when they are truly low. This payment schedule, $S(\cdot)$, specifies what payment, $s = S(x)$, is to be paid the supplier as a function of the units, x, *it chooses* to supply. Wilson (1993) provides evidence that such payment schedules are common in real-world contracting.

If the supplier accepts such a contract, the supplier's choice of quantity, x_t, is given by

$$x_t \in \arg\max_{x \geq 0} \left\{S(x) - C_t(x)\right\}. \tag{7.1}$$

Assume for the moment that this program has a unique solution. Let u_t denote the value of this maximization program and let $s_t = S(x_t)$ be the supplier's payment under the terms of the contract. By definition,

$$u_t = s_t - C_t(x_t).$$

Observe that this means we can write the equilibrium payment, s_t, as

$$s_t = u_t + C_t\left(x_t\right).$$

Define

$$R\left(\cdot\right) = C_I\left(\cdot\right) - C_E\left(\cdot\right)$$

as the *information-rent function*. Our earlier assumptions imply that $R(\cdot)$ is positive for $x > 0$, zero for $x = 0$, and strictly increasing.

Revealed preference in the choice of x necessarily implies the following about x_I and x_E:

$$u_E = s_E - C_E(x_E) \geq s_I - C_E(x_I) = u_I + R(x_I) \tag{7.2}$$
$$u_I = s_I - C_I(x_I) \geq s_E - C_I(x_E) = u_E - R(x_E). \tag{7.3}$$

These inequalities are referred by many names in the literature: *incentive-compatibility* constraints, *self-selection* constraints, *revelation constraints*, and *truth-telling* constraints. Regardless of name, they simply capture the requirement that (x_I, s_I) and (x_E, s_E) be the preferred choices for the supplier in states I and E, respectively.

What can we conclude from expressions (7.2) and (7.3)? First, rewriting them as

$$R(x_I) \leq u_E - u_I \leq R(x_E), \tag{7.4}$$

we see that

$$x_I \leq x_E, \tag{7.5}$$

because $R(\cdot)$ is strictly increasing. Observe, too, that expression (7.2) implies $u_E > u_I$ (except if $x_I = 0$, in which case we only know $u_E \geq u_I$). Finally, expressions (7.2) and (7.5) implies $s_E > s_I$ (unless $x_E = x_I$, in which case (7.2) and (7.3) imply $s_E = s_I$).

Of course the contract—payment schedule $S(\cdot)$—must be acceptable to the supplier, which means

$$u_I \geq 0; \text{ and} \tag{7.6}$$
$$u_E \geq 0\,. \tag{7.7}$$

If these did not both hold, then the contract would be rejected by one or the other or both types of supplier. The constraints (7.6) and (7.7) are referred to as the agent's *participation* or *individual-rationality* constraints. They simply state that, without any bargaining power, the supplier accepts a contract if and only if accepting does not entail suffering a loss.

The retailer's problem is to determine a price schedule $S(\cdot)$ that maximizes its *expected* profit ("expected" because, recall, it knows only the probability that a give type will be realized). Specifically, the retailer seeks to maximize

$$f \times [r(x_I) - s_I] + (1-f) \times [r(x_E) - s_E]\,; \text{ or, equivalently,}$$
$$f \times [r(x_I) - C_I(x_I) - u_I] + (1-f) \times [r(x_E) - C_E(x_E) - u_E]\,,$$

where (x_t, u_t) are determined by the supplier's optimization program (7.1) in response to $S(\cdot)$.

Observe that only two points on the whole price schedule enter the retailer's objective function: (x_I, s_I) and (x_E, s_E); or, equivalently, (x_I, u_I) and (x_E, u_E). The maximization of the principal's objectives can be performed with respect to just these two points provided that we can recover a general payment schedule afterwards such that the supplier would accept this schedule and choose the appropriate point for its type given this schedule. For this to be possible, we know that the self-selection constraints, (7.2) and (7.3), plus the participation constraints, (7.6) and (7.7), must hold.

In fact, the self-selection constraints and the participation constraints on (x_I, s_I) and (x_E, s_E) are necessary and sufficient for there to exist a payment schedule such that the solution to (7.1) for type t is (x_t, s_t). To prove this assertion, let (x_I, s_I) and (x_E, s_E) satisfy those constraints and construct the rest of the payment schedule as follows:

$$\begin{aligned} S(x) &= 0 && \text{if } 0 \le x < x_I \\ &= s_I && \text{if } x_I \le x < x_E \\ &= s_E && \text{if } x_E \le x, \end{aligned}$$

when $0 < x_I < x_E$.[2] Given that $C_t(\cdot)$ is increasing in x, no supplier would ever choose an x other than 0, x_I, or x_E (the supplier's marginal revenue is zero except at these three points). The participation constraints ensure that (x_t, s_t) is (weakly) preferable to $(0, 0)$ and the self-selection constraints ensure that a type-t supplier prefers (x_t, s_t) to $(x_{t'}, s_{t'})$, $t \neq t'$. That is, we've shown that faced with this schedule, the type-I supplier's solution to (7.1) is (x_I, s_I)—as required—and that the type-E supplier's solution to (7.1) is (x_E, s_E)—as required.

The retailer's problem can thus be stated as

$$\max_{\{x_I, x_E, u_I, u_E\}} f \times [r(x_I) - C_I(x_I) - u_I] + (1-f) \times [r(x_E) - C_E(x_E) - u_E] \quad (7.8)$$

subject to (7.2), (7.3), (7.6), and (7.7). Solving this problem using the standard Lagrangean method is straightforward, albeit tedious. Because, however, such a mechanical method provides little intuition, we pursue a different, though equivalent, line of reasoning.

- One can check that ignoring the self-selection constraints (treating them as *not* binding) leads us back to the symmetric-information arrangement; and we know that at least one self-selection constraint is then violated. We can, thus, conclude that in our solution to (7.8) at least one of the self-selection constraints is binding.

- The self-selection constraint in state E implies that: $u_E \geq R(x_I) + u_I \geq u_I$. Therefore, if the supplier accepts the contract in state I, it will also accept it in state E. We can, thus, conclude that constraint (7.7) is slack and can be ignored.

- It is, however, the case that (7.6) must be binding at the optimum: Suppose not, then we could lower both utility terms u_L and u_H by some

[2] If $x_I = 0$, then $s_I = 0$. If $x_I = x_E$, then $s_I = s_E$.

$\varepsilon > 0$ without violating the participation constraints. Moreover, given the two utilities have been changed by the same amount, this can't affect the self-selection constraints. But, from (7.8), lowering the utilities raises the principal's profits—which means our "optimum" wasn't optimal.

- Using the fact that (7.6) is binding, expression (7.4)—the pair of self-selection constraints—reduces to

$$R(x_I) \leq u_E \leq R(x_E).$$

Given a pair of quality levels (x_I, x_E), the retailer wants to keep the supplier's rent as low as possible and will, therefore, choose to pay him the smallest possible information rent; that is, we can conclude that $u_E = R(x_I)$. The self-selection constraint (7.2) is, thus, slack, provided the necessary monotonicity condition (7.5) holds.

Plugging our findings, $u_I = 0$ and $u_E = R(x_I)$, into the retailer's objectives yields the following reduced program:

$$\max_{\{(x_I, x_E) |\ x_I \leq x_E\}} \{f \times [r(x_I) - C_I(x_I)] + (1-f) \times [r(x_E) - C_E(x_E) - R(x_I)]\}.$$

The solution is

$$x_E = x_E^F = \arg\max_{x \geq 0} \{r(x) - C_E(x)\} \tag{7.9}$$

$$x_I = x_I^*(f) \equiv \arg\max_{x \geq 0} \left\{r(x) - C_I(x) - \frac{1-f}{f} R(x)\right\}. \tag{7.10}$$

The only step left is to verify that the monotonicity condition (7.5) is satisfied for these values. If we consider the last two terms in the maximand of (7.10) to be cost, we see that the effective marginal cost of output from the inefficient type is

$$C_I'(x) + \frac{1-f}{f} R'(x) > C_I'(x) > C_E'(x)$$

for $x > 0$.[3] The greater the marginal-cost schedule given a fixed marginal-revenue schedule, the less is traded; that is, it must be that $x_I^*(f) < x_E^F$—the monotonicity condition (7.5) is satisfied.

It is worth summarizing the nature and properties of the optimal price schedule for the retailer to propose:

Proposition 10 *The optimal (non-linear) payment schedule for the principal induces two possible outcomes depending upon the state of nature such that:*

- *the supplier trades the ex post efficient quantity, x_E^F, when it is an efficient producer, but trades* less than *the efficient quantity when it is an inefficient producer (i.e., $x_I^*(f) < x_I^F$);*

[3] Because $x_E^F > 0$, this is the relevant domain of output to consider.

- *an inefficient supplier makes no profit ($u_I = 0$), but an efficient supplier earns an information rent of $R\left[x_I^*(f)\right]$;*
- *the revelation constraint is binding in state E, slack in state I;*
- *the participation constraint is binding in state I, slack in state E;*
- *$x_I^*(f)$ and $R\left[x_I^*(f)\right]$ are non-decreasing in the probability of drawing an inefficient producer (i.e., are non-decreasing in f);*
- *and, finally, $\lim_{f\downarrow 0} x_I^*(f) = 0$, $\lim_{f\uparrow 1} x_I^*(f) = x_I^F$, $\lim_{f\downarrow 0} R\left[x_I^*(f)\right] = 0$, and $\lim_{f\uparrow 1} R\left[x_I^*(f)\right] = R(x_I^F)$.*

To see that the last two points hold, note first that the effective marginal cost of production from an inefficient supplier,

$$C_I'(x) + \frac{1-f}{f} R'(x)\,,$$

is falling in f. By the usual comparative statics, this means that $x_I^*(f)$ is non-decreasing. Since $R(\cdot)$ is an increasing function, $R\left[x_I^*(f)\right]$ must be similarly non-decreasing. As $f \downarrow 0$, this effective marginal cost tends to $+\infty$ for $x > 0$, which means the optimal level of trade falls to zero. As $f \uparrow 1$, this effective marginal cost tends to the symmetric-information marginal cost, hence $x_I^*(f)$ tends to the symmetric-information level, x_I^F.

Intuition for these results can be gained from Figure 7.1. This figure shows one indifference curve for an inefficient (type-I) supplier and three indifference curves for an efficient (type-E) supplier in output-payment space. The type-I indifference curve is that type's zero-profit curve (hence, by necessity, it passes through the origin). Correspondingly, the lowest and darkest of the type-E indifference curves is that type's zero-profit curve. The faint dash-dot lines are iso-profit curves for the retailer (to minimize clutter in the figure, they're sketched as straight lines, but this is not critical for what follows). Observe that an iso-profit curve is tangent to type-I's zero-profit indifference curve at point A. Likewise, we have similar tangency for type-E at point B. Hence—*under symmetric information*—points A and B would be the contracts offered. Under asymmetric information, however, contract B is *not* incentive compatible for type-E: Were it to lie and claim to be type-I (*i.e.,* move to point A), then it would be on a higher (more profitable) indifference curve (the highest of its three curves). Under asymmetric information, an incentive compatible pair of contracts that induce the symmetric-information levels of trade are A and C. The problem with this solution, however, is that type-E earns a large information rent, equal to the distance between B and C. The retailer can reduce this rent by distorting downward the quantity asked from a type-I supplier. For example, by lowering quantity to $x_I^*(f)$, the retailer significantly reduces the information rent (it's now the distance between B and E). How much distortion in quantity the retailer will impose depends on the likelihood of the two types. When f is small, the expected savings in information rent is large, while the

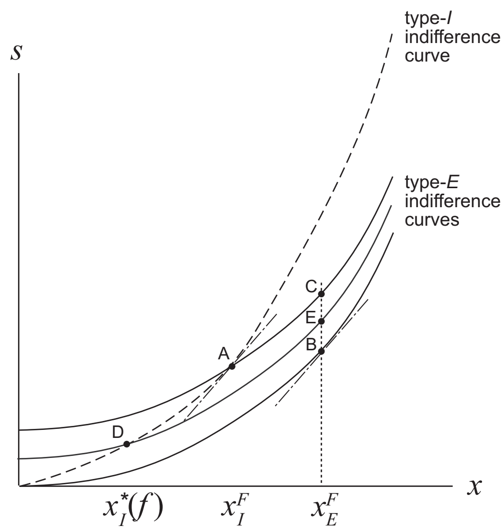

Figure 7.1: The symmetric-information contracts, points A and B, are *not* incentive compatible. The symmetric-information quantities, x_E^F and x_I^F, are too expensive because of the information rent (the distance from B to C). Consequently, with asymmetric information, the principal trades off a distortion in type-I's output (from x_I^F to $x_I^*(f)$) to reduce type-E's information rent (from BC to BE).

expected cost of too-little output is small, so the downward distortion in type-I's output is big. Conversely, when f is large, the expected savings are small and the expected cost is large, so the downward distortion is small. The exact location of point D is determined by finding where the expected marginal cost of distorting type-I's output, $f \times [r'(x_I) - C_I'(x_I)]$, just equals the expected marginal reduction in type-E's information rent, $(1-f) \times R'(x_I)$.

From Figure 7.1, it is clear that the retailer loses from being uninformed about the supplier's type: Point D lies on a worse iso-profit curve than does point A and point E lies on a worse iso-profit curve than does point B.[4] Put another way, if the retailer draws a type-I supplier, then it gets a non-optimal (relative to symmetric information) quantity of the good. While if it draws a type-E supplier, then it pays more for the optimal quantity (again, relative to symmetric information). Part—but only part—of the retailer's loss is the supplier's gain. In expectation, the supplier's profit has increased by $(1-f)R[x_I^*(f)]$. But part of the retailer's loss is also deadweight loss: Trading $x_I^*(f)$ units instead of x_I^F

[4]Since the retailer likes more output (in the relevant range) and smaller payments to the supplier, its utility (profits) are greater on iso-profit curves toward the southeast of the figure and less on iso-profit curves toward the northwest.

units is simply inefficient. In essence, our retailer is a monopsonist and, as is typical of monopsony, there is a deadweight loss.[5] Moreover, the deadweight loss arises here for precisely the same reason it occurs in monopsony: Like a monopsonist, our retailer is asking the payment (price) schedule to play two roles. First, it asking it to allocate goods and, second, it is asking it to preserve its rents from having all the bargaining power. Since only the first role has anything to do with allocative efficiency, giving weight to the second role can only create allocative *in*efficiency. As often happens in economics, a decision maker has one instrument—here, the payment schedule—but is asking it to serve multiple roles. Not surprisingly then, the ultimate outcome is less than first best (recall our discussion of the two-instruments principle page 28).

Because the first best is *not* achieved, it is natural to ask whether the retailer could improve on the outcome in Proposition 10 by using some more sophisticated contract? The answer is no and the proof is, as we will see later, quite general. Whatever sophisticated contract the retailer uses, this contract will boil down to a pair of points, (x_I, s_I) and (x_E, s_E), once it is executed; that is, a final quantity traded and a final payment for each possible state of nature. Consequently, whatever complicated play is induced by the contract, both parties can see through it and forecast that the equilibrium outcomes correspond to these two points. Moreover, by mimicking the strategy it would play in state t, the supplier can generate either of the two outcomes regardless of the true state. In addition, if it can't profit from (x_t, s_t) in state t, it can simply not participate. Necessary equilibrium conditions are, then, that the supplier choose (x_t, s_t) in state t rather than $(x_{t'}, s_{t'})$, $t \neq t'$, and that it choose to participate anticipating that it will choose (x_t, s_t) in state t. But these are precisely the revelation and participation constraints (7.2), (7.3), (7.6), and (7.7). Therefore, whatever the contractual arrangement, the final outcome can always be generated by a simple (non-linear) payment schedule like the one derived above. We've, thus, established that the outcome described in Proposition 10 cannot be improved on by using more sophisticated or alternative contracts.[6]

Finally, note that we don't need an entire payment schedule, $S(\cdot)$. In particular, there is a well-known alternative: a *direct-revelation contract* (mechanism). In a direct-revelation contract, the retailer commits to pay the supplier s_E for x_E or s_I for x_I depending on the supplier's announcement of its type. Failure by the supplier to announce its type (*i.e.,* failure to announce a $\hat{t} \in \{E, I\}$) is equivalent to the supplier rejecting the contract. Finally, if, after announcing its type, the supplier produces a quantity other than $x_{\hat{t}}$, the supplier is punished (*e.g.,* paid nothing). It is immediate that this direct-revelation contract is equivalent to the optimal payment schedule derived above. It is also simpler, in

[5] Monopsony is a situation in which the buyer has all the market power (in contrast to monopoly, in which the seller has all the market power).

[6] This is not to say that *another* contract couldn't do as well. This is rather obvious: For instance, suppose $\tilde{S}(x) = 0$ for all x except x_E and x_I, where it equals s_E or s_I, respectively. We've merely established that no other contract can do strictly better than the $S(\cdot)$ derived in the text.

that it only deals with the relevant part of the payment schedule. Admittedly, it is not terribly realistic,[7] but as this discussion suggests we can transform a direct-revelation contract into a more realistic contract (indeed, we will formalize this below in Proposition 12). More importantly, as we will see, in terms of determining what is the optimal feasible outcome, there is no loss of generality in restricting attention to direct-revelation contracts.

[7]Although see Gonik (1978) for a real-life example of a direct-revelation mechanism.

8 General Screening Framework

The two-type screening model yielded strong results. But buried within it is a lot of structure and some restrictive assumptions. If we are really to use the screening model to understand economic relationships, we need to deepen our understanding of the phenomena it unveils, the assumptions they require, and the robustness of its conclusions. Our approach in this section is, thus, to start from a very general formalization of the problem and to motivate or discuss the assumptions necessary for making this model "work."

A principal and an agent are involved in a relationship that can be characterized by an *allocation* $x \in \mathcal{X}$ and a real-valued monetary transfer $s \in \mathcal{S} \subset \mathbb{R}$ between the two players. A transfer-allocation pair, (x, s), is called an *outcome*. The space of possible allocations, $\mathcal{X}$, can be quite general: Typically, as in our analysis of the retailer-supplier problem, it's a subspace of $\mathbb{R}$; but it could be another space, even a multi-dimensional one. In what follows, we assume that outcomes are verifiable.

The agent's information is characterized by a parameter $\theta \in \Theta$. As before, we'll refer to this information as the agent's type. The *type space*, Θ, can be very general. Typically, however, it is either a discrete set (*e.g.,* as in the retailer-supplier example where we had $\Theta = \{I, E\}$) or a compact interval in $\mathbb{R}$. Nature draws θ from Θ according to a commonly known probability distribution. While the agent learns the value of θ perfectly, the principal only knows that it was drawn from the commonly known probability distribution.

Both players' preferences are described by von Neumann-Morgenstern utility functions, $\mathcal{W}(x, s, \theta)$ for the principal and $\mathcal{U}(x, s, \theta)$ for the agent, where both are defined over $\mathcal{X} \times \mathcal{S} \times \Theta$. Because we interpret s as a transfer *from* the principal *to* the agent, we assume that $\mathcal{U}$ increases in s, while $\mathcal{W}$ decreases in s. For convenience, we assume these utility functions are smooth; more precisely, that they are three-times continuously differentiable.[1]

To have a fully general treatment, we need to extend this analysis to allow for the possibility that the actual outcome is chosen randomly from $\mathcal{X} \times \mathcal{S}$. To this end, let σ denote a generic element of the set of probability distributions, $\Delta\left(\mathcal{X} \times \mathcal{S}\right)$, over the set of possible outcomes, $\mathcal{X} \times \mathcal{S}$. We extend the utility functions to $\Delta\left(\mathcal{X} \times \mathcal{S}\right)$ through the expectation operator:

$$\mathcal{W}(\sigma, \theta) \equiv \mathbb{E}_\sigma\left[\mathcal{W}(x, s, \theta)\right] \text{ and } \mathcal{U}(\sigma, \theta) \equiv \mathbb{E}_\sigma\left[\mathcal{U}(x, s, \theta)\right].$$

[1] One could relax this smoothness assumption, but the economic value of doing so is too small to warrant time on this issue.

Table 8.1: The Retailer-Supplier Example in our General Notation

Description	General Notation	Specific Value
Allocation space	$\mathcal{X}$	$\mathbb{R}_+$
Transfer space	$\mathcal{S}$	$\mathbb{R}$
Outcome function	σ	All mass on $(x, S(x))$
Principal's strategy space	$\mathcal{M}$	$\{\text{pay } S(x)\}$
Agent's strategy space	$\mathcal{N}$	$\mathcal{X} = \mathbb{R}_+$

By adding an element to $\mathcal{X}$ if necessary, we assume that there exists a no-trade outcome $(x_0, 0)$; that is, $(x_0, 0)$ is the outcome if no agreement is reached (*e.g.,* in the retailer-supplier example this was $(0, 0)$). The values of both $\mathcal{W}$ and $\mathcal{U}$ at this no-agreement point play an important role in what follows, so we give them special notation:

$$W_R(\theta) \equiv \mathcal{W}(x_0, 0, \theta) \text{ and } U_R(\theta) \equiv \mathcal{U}(x_0, 0, \theta).$$

These will be referred to as the *reservation utilities* of the players (alternatively, their individual-rationality payoffs). Obviously, an agent of type θ accepts a contract if and only if his utility from doing is not less than $U_R(\theta)$.

It is convenient, at this stage, to offer a formal and general definition of what a contract is:

Definition 1 *A contract in the static contractual screening model is a game form, $\langle \mathcal{M}, \mathcal{N}, \sigma \rangle$, to be played by the principal and the agent, $\mathcal{M}$ denotes the agent's strategy set, $\mathcal{N}$ the principal's strategy set, and σ an outcome function that maps any pair of strategies (m, n) to a probability mapping on $\mathcal{X} \times \mathcal{S}$. That is, $\sigma : \mathcal{M} \times \mathcal{N} \rightarrow \Delta(\mathcal{X} \times \mathcal{S})$.*

To make this apparatus somewhat more intuitive, consider Table 8.1, which "translates" our retailer-supplier example into this more general framework. Observe that, in the example, the contract fixes a trivial strategy space for the principal: She has no discretion, she simply pays $S(x)$.[2] Moreover, there is no randomization in that example: σ simply assigns probability one to the pair $(x, S(x))$. As this discussion suggests, generality in notation need not facilitate understanding—the generality contained here is typically greater than we need. Fortunately, we will be able to jettison much of it shortly.

A *direct mechanism* is a mechanism in which $\mathcal{M} = \Theta$; that is, the agent's action is limited to making announcements about his type. The physical consequences of this announcement are then built into the outcome function, σ.

[2] We could, alternatively, expand her strategy space to $\mathcal{S}$—she can attempt to pay the agent whatever she wants. But, then, the outcome function would have to contain a punishment for not paying the agent appropriately: That is, $\sigma(x, s) = (x, S(x))$ if $s = S(x)$ and equals (x, ∞) if $s \neq S(x)$ (where ∞ is shorthand for some large transfer sufficient to deter the principal from not making the correct payment).

For instance, as we saw at the end of the previous section, we can translate our retailer-supplier contract into a direct mechanism: $\mathcal{M} = \{E, I\}$, $\mathcal{N}$ is a singleton (so we can drop n as an argument of σ), and

$$\sigma(m) = \begin{cases} (x_I, s_I) = (x_I^*(f), C_I[x_I^*(f)]) & \text{if } m = I \\ (x_E, s_E) = \left(x_E^F, R[x_I^*(f)] + C_E\left(x_E^F\right)\right) & \text{if } m = E \end{cases}.$$

A *direct-revelation mechanism* (alternatively, a direct truthful mechanism*)* is a direct mechanism where it is an equilibrium strategy for the agent to tell the truth: Hence, if $m(\cdot) : \Theta \to \Theta$ is the agent's strategy, we have $m(\theta) = \theta$ in equilibrium for all $\theta \in \Theta$. That is, for any θ and θ' in Θ,

$$\mathcal{U}(\sigma(m[\theta]), \theta) \geq \mathcal{U}(\sigma(m[\theta']), \theta).$$

Note that not every direct mechanism will be a direct-revelation mechanism. Being truthful in equilibrium is a property of a mechanism; that is, it depends on $\sigma(\cdot)$.

Observe that the design of a contract means choosing $\mathcal{M}$, $\mathcal{N}$, and σ. In theory, the class of spaces and outcome functions is incomprehensibly large. How can we find the optimal contract in such a large class? Indeed, given the inherent difficulties in even characterizing such a large class, how can we ever be sure that we've found the optimal contract? Fortunately, two simple, yet subtle, results—the *revelation principle* and the *the taxation principle*—allow us to avoid these difficulties.[3] From the revelation principle, the search for on optimal contract reduces *without loss of generality* to the search for the optimal direct-revelation mechanism. Moreover, if the outcome in the direct-revelation mechanism is a *deterministic* function of the agent's announcement, then, from the taxation principle, we may further restrict attention to a payment schedule that is a function of the allocation x (as we did in the retailer-supplier example).

Proposition 11 (The Revelation Principle)[4] *For any general contract* $(\mathcal{M}, \mathcal{N}, \sigma)$ *and associated Bayesian equilibrium, there exists a direct-revelation mechanism such that the associated truthful Bayesian equilibrium generates the same equilibrium outcome as the general contract.*

Proof: The proof of the revelation principle is standard but informative. A Bayesian equilibrium of the game $(\mathcal{M}, \mathcal{N}, \sigma)$ is a pair of strategies $(m(\cdot), n)$.[5] Let us consider the following direct mechanism: $\hat{\sigma}(\cdot) = \sigma(m(\cdot), n)$. Our claim is

[3]It is unfortunate that these two fundamental results are called principles, since they are not, as their names might suggest, premises or hypotheses. They are, as we will show, deductive results.

[4]The revelation principle is often attributed to Myerson (1979), although Gibbard (1973) and Green and Laffont (1977) could be identified as earlier derivations. Suffice it to say that the revelation principle has been independently derived a number of times and was a well-known result before it received its name.

[5]Observe that the agent's strategy can be conditioned on θ, which he knows, while the principal's cannot be (since she is ignorant of θ).

that $\hat{\sigma}(\cdot)$ induces truth-telling (is a direct-*revelation* mechanism). To see this, suppose it were not true. Then there must exist a type θ such that the agent does better to lie—announce some $\theta' \neq \theta$—when he is type θ. Formally, there must exist θ and $\theta' \neq \theta$ such that

$$\mathcal{U}(\hat{\sigma}(\theta'), \theta) > \mathcal{U}(\hat{\sigma}(\theta), \theta).$$

Using the definition of $\hat{\sigma}(\cdot)$, this means that

$$\mathcal{U}(\sigma\left[m(\theta'), n\right], \theta) > \mathcal{U}(\sigma\left[m\left(\theta\right), n\right], \theta);$$

but this means the agent prefers to play $m(\theta')$ instead of $m(\theta)$ in the *original* mechanism against the principal's equilibrium strategy n. This, however, can't be since $m(\cdot)$ is an equilibrium best response to n in the original game. Hence, truthful revelation must be an optimal strategy for the agent under the constructed direct mechanism. Finally, when the agent truthfully reports the state of nature in the direct truthful mechanism, the same outcome $\hat{\sigma}(\theta) = \sigma(m(\theta), n)$ is implemented in equilibrium. ■

An intuitive way to see the revelation principle is imagine that before he plays some general mechanism, the agent could delegate his play to some trustworthy third party. There are two *equivalent* ways this delegation could work. One, the agent could tell the third party to play m. Alternatively, if the third party knows the agent's equilibrium strategy—the mapping $m : \Theta \to \mathcal{M}$—then the agent could simply reveal (announce) his type to the third party with the understanding that the third party would choose the appropriate actions, $m\left(\theta\right)$. But, because we can build this third party into the design of our direct-revelation mechanism, this equivalence means that there is no loss of generality in restricting attention to direct-revelation mechanisms.

The taxation principle requires a little more structure: It assumes that there is a possibility of punishing the agent so as to deter him from violating the contractual rules

Assumption 2 (Existence of a punishment) *There exists an* $\underline{s} \in \mathbb{R}$ *such that:*

$$\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \mathcal{U}(x, \underline{s}, \theta) \leq \inf_{(x,s,\theta)\in\mathcal{X}\times\mathcal{S}\times\Theta} \mathcal{U}(x, s, \theta).$$

In other words, there exists a punishment so severe that the agent would always prefer not to suffer it.

With this assumption, one can construct a payment schedule that generates the same outcome as any *deterministic* direct-revelation mechanism and is, therefore, as general as any contract in this context.

Proposition 12 (The Taxation Principle) *Under Assumption 2, the equilibrium outcome under any deterministic direct-revelation mechanism,* $\sigma\left(\cdot\right) =$

$(x(\cdot), s(\cdot))$, *is also an equilibrium outcome of the game where the principal proposes the payment schedule* $S(\cdot)$ *defined by*

$$S(x) = \begin{cases} s(\theta)\,, & \text{if } \theta \in x^{-1}(x) \text{ (i.e., such that } x = x(\theta) \text{ for some } \theta \in \Theta\text{)} \\ \underline{s}, & \text{otherwise} \end{cases} .$$

Proof: Let's first establish that $S(\cdot)$ is unambiguously defined: Suppose there existed θ_1 and θ_2 such that

$$x = x(\theta_1) = x(\theta_2),$$

but $s(\theta_1) \neq s(\theta_2)$. We're then free to suppose that $s(\theta_1) > s(\theta_2)$. Then, because $\mathcal{U}$ is increasing in s,

$$\mathcal{U}(x(\theta_1), s(\theta_1), \theta_2) > \mathcal{U}(x(\theta_2), s(\theta_2), \theta_2).$$

But this means the agent would prefer pretending that the state of nature is θ_2 when it's actually θ_1; the mechanism would *not* be truthful. Hence, if $\theta_1, \theta_2 \in x^{-1}(x)$, we must have $s(\theta_1) = s(\theta_2)$; that is, the payment schedule $S(\cdot)$ is unambiguously defined.

Now, the agent's problem when faced with the payment schedule $S(\cdot)$ is simply to choose the allocation x that maximizes $\mathcal{U}(x, S(x), \theta)$. Given the severity of the punishment, the agent will restrict his choice to $x \in x(\Theta)$. But since our original mechanism was a direct-revelation mechanism, we know

$$\mathcal{U}(x(\theta), s(\theta), \theta) \geq \mathcal{U}(x(\theta'), s(\theta'), \theta)$$

for all θ and θ'. So no type θ can do better than to choose $x = x(\theta)$. ■

The economic meaning of the taxation principle is straightforward: When designing a contract, the principal is effectively free to focus on "realistic" compensation mechanisms that pay the agent according to his achievements. Hence, as we argued above, there is no loss of generality in our solution to the retailer-supplier problem.

Although payment schedules involve no loss of generality and are realistic, they are often nonlinear, which creates difficulties in working with them.[6] In particular, when looking for the optimal non-linear price schedule, one must be able to compute the functional mapping that associates to each schedule $S(\cdot)$ the action choice $x(\theta)$ that maximizes $\mathcal{U}(x, S(x), \theta)$. Even assuming $S(\cdot)$ is differentiable—which is not ideal because one should refrain from making assumptions about *endogenous* variables—solving this problem can be difficult. Direct-revelation mechanisms, on the other hand, allow an easier mathematical treatment of the problem using standard convex analysis: The revelation constraints simply consist of writing $\theta' = \theta$ is a maximum of $\mathcal{U}(x(\theta'), s(\theta'), \theta)$

[6] Admittedly, if the derived payment schedule is sufficiently nonlinear, one could begin to question its realism, as real-world contracts are often linear or, at worst, piecewise linear. Although see footnote 7 on page 74 for an example of where reality "matches" theory.

and writing that a given point is a maximum is easier that characterizing an unknown maximum. For this reason, much of the mechanism-design literature has focussed on direct-revelation mechanisms over optimal-payment schedules despite the latter's greater realism.

The Standard Framework | 9

As we've already hinted, the general framework introduced in the previous section is more general than what is commonly used. In this section, we will therefore considered a restricted framework known as the *"standard"* framework, Much of the contractual screening literature can be placed within this standard framework. We begin by defining this framework and contrasting it to the more general framework introduced above. At the end of this note, I offer less conventional views on the screening model, which require departing from the standard framework. In order to provide a road map to the reader, I spell out each assumption in this section and suggest what consequences would arise from alternative assumptions.

In the standard framework, the allocation space, $\mathcal{X}$, is $\mathbb{R}_+$.[1] The type space, Θ, is $[\theta_L, \theta_H] \subset \mathbb{R}$, where both bounds, θ_L and θ_H, are finite.[2] The most critical assumptions in going from the general framework to the standard framework involve the utility functions. Henceforth, we assume they are additively separable in the transfer and the allocation. Moreover, we assume the marginal value of money is type *in*dependent (*i.e.,* $\partial\left(\partial\mathcal{U}/\partial s\right)/\partial\theta = \partial\left(\partial\mathcal{W}/\partial s\right)/\partial\theta = 0$). These assumptions simplify the analysis by eliminating income effects from consideration. Formally, the agent's and principal's utility functions are

$$\mathcal{U}(x, s, \theta) = s + u(x, \theta) \text{ and}$$
$$\mathcal{W}(x, s, \theta) = w(x, \theta) - s,$$

respectively. Essentially for convenience, we take $u(\cdot,\cdot)$ and $w\left(\cdot,\cdot\right)$ to be three-times continuously differentiable. The aggregate (full-information) surplus is defined as:

$$\Omega(x, \theta) = w(x, \theta) + u(x, \theta).$$

We also assume, mainly for convenience, that

1. **Some trade is desirable:** For all $\theta \in (\theta_L, \theta_H]$, $\partial\Omega\left(0,\theta\right)/\partial x > 0$.

2. **There can be too much of a good thing:** For all $\theta \in [\theta_L, \theta_H]$ $\exists\bar{x}\left(\theta\right)$ such that $\Omega\left(x,\theta\right) \leq \Omega\left(0,\theta\right)$ for all $x > \bar{x}\left(\theta\right)$.

[1] That the space be bounded below at 0 is not critical—any lower bound would do. Alternatively, by appropriate changes to the utility functions, we could allow the allocation space to be unbounded. Zero is simply a convenience.

[2] This constitutes no loss of *economic* (as opposed to mathematical) generality, since we can set the bounds as far apart as we need.

Observe these two assumptions entail that $\Omega(x, \theta)$ has an interior maximum for all $\theta \in (\theta_L, \theta_H]$. If the first assumption didn't hold for at least some types, then trade—contracting—would be pointless. Extending the desirability of trade to almost all types saves from the bookkeeping headache of distinguishing between types with which trade is efficient and those with which it is not. The second assumption is just one of many ways of expressing the sensible economic idea that, beyond some point, welfare is reduced by trading more.

Observe that the standard framework is restrictive in several potential important ways:

- The type space is restricted to be one-dimensional. In many applications, such as the retailer-supplier example, this is a natural assumption. One can, however, conceive of applications where it doesn't fit: Suppose, *e.g.,* the retailer cared about the quantity and quality of the goods received and the supplier's type varied on both an efficiency dimension and a conscientiousness-of-employees dimension (the latter affecting the cost of providing quality). Not surprisingly, restricting attention to one dimension is done for analytic tractability: Assuming a single dimension make the order properties of the type space straightforward (*i.e.,* greater and less than are well-defined on the real line). As we will see, the order properties of the type space are critical to our analysis.

- The set of possible allocations is one-dimensional. Again, this is sufficient for some applications (*e.g.,* the quantity supplied to the retailer), but not others (*e.g.,* when the retailer cares about both quantity and quality). The difficulty in expanding to more than one dimension arise from difficulties in capturing how the agent's willingness to make tradeoffs among the dimensions (including his income) varies with his type. The reader interested in this extension should consult Rochet and Choné (1998).

- As noted, the utility functions are separable in money and allocation; the marginal utility of income is independent of the state of nature; and the marginal utility of income is constant, which means both players are risk neutral with respect to gambles over money. The gains from these assumptions are that we can compute the transfer function $s(\cdot)$ in terms of the allocation function $x(\cdot)$, which means our optimization problem is a standard optimal-control problem with a unique control, $x(\cdot)$. In addition, risk neutrality insulates us from problems that *exogenously* imposed risk might otherwise create (*e.g.,* the need to worry about mutual insurance). On the other hand, when the agent is risk averse, the ability to threaten him with *endogenously* imposed risk (from the contract itself) can provide the principal an additional tool with which to improve the ultimate allocation. For a discussion of some of these issues see Edlin and Hermalin (2000). Note we still have the flexibility to endogenously impose risk over the *allocation* (the x), we discuss the desirability of doing so below. See also Maskin (1981).

Continuing with our development of the standard framework, we assume that nature chooses the agent's type, θ, according to the distribution function $F(\cdot) : [\theta_L, \theta_H] \to [0,1]$. Let $f(\cdot)$ be the associated density function, which we assume to be continuous and to have full support (*i.e.,* $f(\theta) > 0$ for all $\theta \in [\theta_L, \theta_H]$). Assuming a continuum of types and a distribution without mass points is done largely for convenience. It also generalizes our analysis from just two types. Admittedly, we could have generalized beyond two types by allowing for a finite number of types greater than two. The conclusions we would reach by doing so would be economically similar to those we'll shortly obtain with a continuum of types.[3] The benefit of going all the way to the continuum is it allows us to employ calculus, which streamlines the analysis.

Recall that, at its most general, a direct-revelation mechanism is a mapping from the type space into a *distribution* over outcomes. Given the way that money enters both players utility functions, we're free to replace a distribution over payments with an expected payment, which means we're free to assume that the payment is fixed deterministically by the agent's announcement.[4] What about random-*allocation* mechanisms? The answer depends on the risk properties of the two players' utilities over allocation. If we postulate that $w(\cdot,\theta)$ and $u(\cdot,\theta)$ are concave for all $\theta \in [\theta_L, \theta_H]$, then, *absent incentive concerns*, there would be no reason for the principal to prefer a random-allocation mechanism; indeed, if at least one is strictly concave (*i.e.,* concave, but not affine), then she would strictly prefer not to employ a random-allocation mechanism absent incentive concerns: Her expected utility is greater with a deterministic mechanism and, since the agent's expected utility is greater, her payment to him will be less (a benefit to her). Hence, we would only expect to see random-allocation mechanisms if the randomness somehow relaxed the incentive concerns. Where does this leave us? At this point, consistent with what is standardly done, we will assume that both $w(\cdot,\theta)$ and $u(\cdot,\theta)$ are concave, with one at least being strictly concave, for all $\theta \in [\theta_L, \theta_H]$ (note this entails that $\Omega(\cdot,\theta)$, the social surplus function, is also strictly concave). Hence, absent incentive concerns, we'd be free to ignore random-allocation mechanisms. For the time being, we'll also ignore random-allocation mechanisms with incentive concerns. Later, we'll consider the circumstances under which this is appropriate. Since we're ignoring random mechanisms, we'll henceforth write $\langle x(\cdot), s(\cdot)\rangle$ instead of $\sigma(\cdot)$ for the mechanism.

Within this framework, the route that we follow consists of two steps. First, we will characterize the set of direct-revelation contracts; that is, the set of contracts $\langle x(\cdot), s(\cdot)\rangle$ from $[\theta_L, \theta_H]$ to $\mathbb{R} \times \mathbb{R}_+$ that satisfy truthful revelation. This truthful-revelation—or incentive compatibility—condition can be expressed as:

$$s(\theta) + u\big(x(\theta),\theta\big) \geq s(\tilde{\theta}) + u\big(x(\tilde{\theta}),\theta\big) \tag{9.1}$$

[3]See, *e.g.,* Caillaud and Hermalin (1993), particular §3, for a finite-type-space analysis under the standard framework.

[4]That is, the mechanism that maps θ to a distribution $G(\theta)$ over payments is equivalent to a mechanism that maps θ to the deterministic payment $\hat{s}(\theta) = \mathbb{E}_{G(\theta)}\{s\}$.

for all $(\theta,\tilde{\theta}) \in [\theta_L,\theta_H]^2$. After completing this first step, the second step is identifying from within this set of incentive-compatible contracts the one that maximizes the principal's expected utility subject to the agent's participation. Whether the agent participates depends on whether his equilibrium utility exceeds his reservation utility, $U_R(\theta)$. Observe that the requirement that the agent accept the contract imposes an additional constraint on the principal in designing the optimal contract. As before, we refer to this constraint as the *participation* or *individual-rationality* (IR) constraint. Recall that $\mathcal{X}$ is assumed to contain a no-trade allocation, x_0, $s = 0$ is a feasible transfer, and $U_R(\theta) \equiv \mathcal{U}(x_0, 0, \theta)$. Hence there is no loss of generality in requiring the principal to offer a contract that is individually rational for all types (although the "contract" for some types might be no trade).

We assume that the agent acts in the principal's interest when he's otherwise indifferent. In particular, he accepts a contract when he is indifferent between accepting and rejecting it and he tells the truth when indifferent between being honest and lying. This is simply a necessary condition for an equilibrium to exist and, as such, should not be deemed controversial.

In our treatment of the retailer-supplier example, we assumed that both types of supplier had the same reservation utility (*i.e.,* recall, $U_R(\theta) = \mathcal{U}(0,0,\theta) = -C_\theta(0) = 0$). It is possible, however, to imagine models in which the reservation utility varies with θ. For instance, suppose that an efficient supplier could, if *not* employed by the retailer, market its goods directly to the ultimate consumers (although, presumably, not as well as the retailer could). Suppose, in fact, it would earn a profit of $\pi_E > 0$ from direct marketing. Then we would have $U_R(E) = \pi_E$ and $U_R(I) = 0$. A number of authors (see, *e.g.,* Lewis and Sappington, 1989, Maggi and Rodriguez-Clare, 1995, and Jullien, 1996) have studied the role of such type-*dependent* reservation utilities in contractual screening models. Type dependence can, however, greatly complicate the analysis. We will, therefore, adopt the more standard assumption of type-*in*dependent reservation utilities; that is, we assume—as we did in our retailer-supplier example—that

$$U_R(\theta) = U_R \text{ and } W_R(\theta) = W_R$$

for all $\theta \in [\theta_L, \theta_H]$. As a further convenience, we will interpret $x = 0$ as the no-trade allocation. Observe that these last two assumptions imply

$$u(0,\theta) = u(0,\theta') = U_R \text{ and}$$
$$w(0,\theta) = w(0,\theta') = W_R$$

for all θ, $\theta' \in [\theta_L, \theta_H]$. Given these assumptions, we can express the agent's participation constraint as

$$s(\theta) + u\big(x(\theta),\theta\big) \geq U_R. \tag{9.2}$$

Although our treatment of reservation utilities and no trade is standard, we have, nevertheless added to the assumptions underlying the standard framework.

At last, we can state the problem that we seek to solve: Find the optimal contract $\langle x(\cdot), s(\cdot)\rangle$ that maximizes

$$\int_{\theta_L}^{\theta_H} \Big(w\big(x(\theta),\theta\big) - s(\theta)\Big) f(\theta)d\theta \tag{9.3}$$

subject to (9.1) and (9.2) holding.

Before solving this program, it's valuable to consider the full (symmetric) information benchmark: *Ex post* efficiency corresponds to adopting the allocation

$$x^F(\theta) \in \arg\max_{x\in\mathbb{R}_+} \Omega(x,\theta)$$

for each θ (recall $\Omega(x,\theta)$ is the aggregate surplus from the relationship). Our earlier assumptions ensure that $\Omega(\cdot,\theta)$ is strictly concave for each θ; so the *ex post* efficient allocation is uniquely defined by the first-order condition:

$$\frac{\partial\Omega}{\partial x}(x^F(\theta),\theta) = \frac{\partial w}{\partial x}(x^F(\theta),\theta) + \frac{\partial u}{\partial x}(x^F(\theta),\theta) = 0.$$

Our earlier assumptions also entail that $x^F(\cdot)$ is uniformly bounded from above. Observe that any sharing of the surplus can, then, be realized by the appropriate transfer function.[5] It follows that, if the principal knows θ—that is, the parties are playing under full (symmetric) information—then the contracting game can be easily solved: In equilibrium, the principal offers a contract $\big\langle x^F(\cdot), s^F(\cdot)\big\rangle$ such that the agent's utility is exactly equal to his reservation utility; that is,

$$s^F(\theta) = U_R - u(x^F(\theta),\theta).$$

In other words, the principal captures the entire surplus—a consequence of endowing her with all the bargaining power—leaving the agent at his outside (non-participation) option.

9.1 The Spence-Mirrlees Assumption

Before we proceed to solve (9.3), we need to introduce one more assumption. Given this assumption's importance, it is worth devoting a short section to it.

In order to screen types, the principal must be able to exploit differences across the *tradeoffs* that different types are willing to make between money and allocation. Otherwise a strategy, for instance, of decreasing the x expected from the agent in exchange for slightly less pay wouldn't work to induce one type to reveal himself to be different than another type. Recall, for instance, because the marginal cost of output differed between the efficient and inefficient types

[5] Note that this is an instance of where we're exploiting the additive separability (lack of income effects) assumed of the utility functions. Were there income effects, we couldn't define $x^F(\cdot)$ independently of the transfers.

in our retailer-supplier example, the buyer could design a contract to induce revelation. Different willingness to make tradeoffs means we require that the indifference curves of the agent in x-s space differ with his type. In fact, we want, for any point in x-s space, that these slopes vary monotonically with whatever natural order applies to the type space. Or, when, no natural order applies, we want it to be possible to define an order, $\succ$, over the types so that $\theta \succ \theta'$ if and only if the slope of θ's indifference curve is greater (alternatively less) than the slope of θ''s indifference at *every* point $(x, s) \in \mathcal{X} \times \mathcal{S}$. Such a monotonicity-of-indifference-curves condition is known as a Spence-Mirrlees condition and, correspondingly, the assumption that this condition is met is known as the Spence-Mirrlees assumption.

The slope of the indifference curve in x-s space is equal to $-\partial u/\partial x$. Hence, we require that $-\partial u/\partial x$ or, equivalently and more naturally, $\partial u/\partial x$ vary monotonically in θ. Specifically, we assume:

Assumption 3 (Spence-Mirrlees) *For all $x \in \mathcal{X}$,*

$$\frac{\partial u(x, \theta)}{\partial x} > \frac{\partial u(x, \theta')}{\partial x}$$

if $\theta > \theta'$.

That is, if $\theta > \theta'$—θ is a *higher* type than θ'—then -1 times the slope of type θ's indifference curve is, at any point, greater than -1 times the slope of type θ''s indifference curve. Observe that a consequence of Assumption 3 is that a given indifference curve for one type can cross a given indifference curve of another type at most once. For this reason, the Spence-Mirrlees Assumption is sometimes called a *single-crossing* condition. Figure 9.1 illustrates.

Lemma 7 *Assuming (as we are) that $u(\cdot, \cdot)$ is at least twice differentiable in both arguments, then the Spence-Mirrlees assumption (Assumption refass:S-M1) is equivalent to assuming*

$$\frac{\partial^2 u(x, \theta)}{\partial \theta \partial x} > 0\,. \tag{9.4}$$

Proof: As a differentiable function cannot be everywhere increasing unless its derivative is positive, the fact that the Spence-Mirrlees assumption implies (9.4) is obvious. To go the other way, observe (9.4) implies

$$\begin{aligned} 0 &< \int_{\theta'}^{\theta} \frac{\partial^2 u(x, t)}{\partial \theta \partial x} dt \\ &= u(x, \theta) - u(x, \theta')\,, \end{aligned} \tag{9.5}$$

where (9.5) follows from the fundamental theorem of calculus. ■

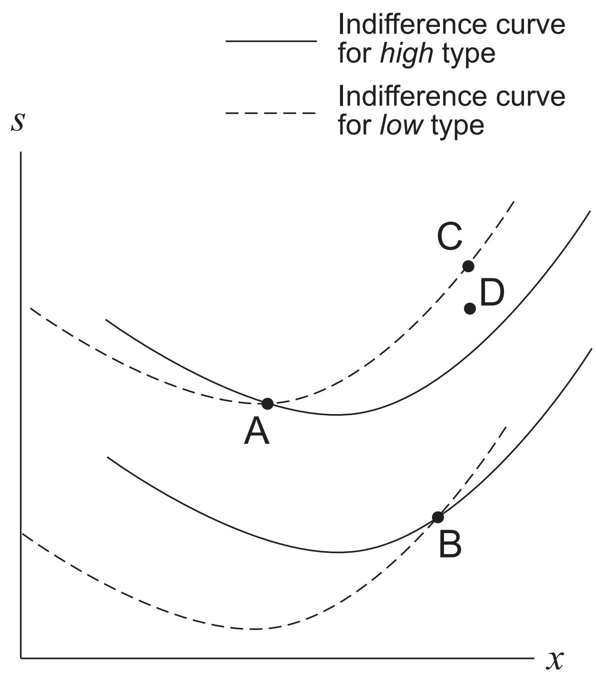

Figure 9.1: The *Spence-Mirrlees* Assumption: Through any point (*e.g.,* A or B), the indifference curve through that point for the high type cross the indifference curve through that point for the low point from *above.*

Before leaving the Spence-Mirrlees assumption, it is important to understand that the Spence-Mirrlees assumption is an assumption about order. Consequently, differentiability of u with respect to either x or θ is not necessary. Nor, in fact, is it necessary that $\mathcal{U}$ be additively separable as we've been assuming. At its most general, then, we can state the Spence-Mirrlees assumption as

Assumption 3′ (generalized Spence-Mirrlees): *There exists an order $\succ_\theta$ on the type space, Θ, such that if $\theta \succ_\theta \theta'$ and $x \succ_x x'$ (where $\succ_x$ completely orders $\mathcal{X}$), then the implication*

$$\mathcal{U}(x, s, \theta') \geq \mathcal{U}(x', s', \theta') \Longrightarrow \mathcal{U}(x, s, \theta) > \mathcal{U}(x', s', \theta)$$

is valid.

This generalized Spence-Mirrlees assumption states that we can order the types so that if a low type (under this order) prefers, at least weakly, an outcome with more x (with "more" being defined by the order $\succ_x$) than a second outcome, then a higher type must strictly prefer the first outcome to the second. Figure

9.1 illustrates: Since the low type prefers point C to A (weakly), the high type must strictly prefer C to A, which the figure confirms. Similarly, since the low type prefers C to B (strictly), the high type must also strictly prefer C to B, which the figure likewise confirms. See Milgrom and Shannon (1994) for a more complete discussion of the relationship between Assumption 3 and Assumption 3′.

As suggested at beginning of this section, a consequence of the Spence-Mirrlees assumption is that it is possible to *separate* any two types; by which we mean it is possible to find two outcomes (x_1, s_1) and (x_2, s_2) such that a type-θ_1 agent prefers (x_1, s_1) to (x_2, s_2), but a type-θ_2 agent has the opposite preferences. For instance, in Figure 9.1, let point A be (x_1, s_1) and let D be (x_2, s_2). If θ_2 is the high type and θ_1 is the low type, then it is clear that given the choice between A and D, θ_1 would select A and θ_2 would select D; that is, this pair of contracts separates the two types. Or, for example, back in Figure 7.1, contracts D and E separate the inefficient and efficient types of supplier.

9.2 Characterizing the Incentive-Compatible Contracts

Our approach to solving the principal's problem (9.3) is a two-step one. First, we will find a convenient characterization of the set of incentive-compatible mechanisms (*i.e.,* those that satisfy (9.1)). This is our objective here. Later, we will search from *within* this set for those that maximize (9.3) subject to (9.2).

Within the standard framework it is relatively straightforward to derive the necessary conditions implied by the self-selection constraints. Our approach is standard (see, *e.g.,* Myerson, 1979, among others). Consider any direct-revelation mechanism $\langle x(\cdot), s(\cdot)\rangle$ and consider any pair of types, θ_1 and θ_2, with $\theta_1 < \theta_2$. Direct revelation implies, among other things, that type θ_1 won't wish to pretend to be type θ_2 and *vice versa.* Hence,

$$s(\theta_1) + u\left[x(\theta_1), \theta_1\right] \geq s(\theta_2) + u\left[x(\theta_2), \theta_1\right] \text{ and}$$
$$s(\theta_2) + u\left[x(\theta_2), \theta_2\right] \geq s(\theta_1) + u\left[x(\theta_1), \theta_2\right].$$

As is often the case in this literature, it is easier to work with utilities than payments. To this end, define

$$v(\theta) = s(\theta) + u\left[x(\theta), \theta\right].$$

Observe that $v(\theta)$ is the type-θ agent's *equilibrium* utility. The above pair of inequalities can then be written as:

$$v(\theta_1) \geq v(\theta_2) - u\left[x(\theta_2), \theta_2\right] + u\left[x(\theta_2), \theta_1\right] \text{ and}$$
$$v(\theta_2) \geq v(\theta_1) - u\left[x(\theta_1), \theta_1\right] + u\left[x(\theta_1), \theta_2\right].$$

Or, combining these two inequalities, as

$$\int_{\theta_1}^{\theta_2} \frac{\partial u\left[x(\theta_1),\theta\right]}{\partial\theta} d\theta \le v(\theta_2) - v(\theta_1) \le \int_{\theta_1}^{\theta_2} \frac{\partial u\left[x(\theta_2),\theta\right]}{\partial\theta} d\theta. \tag{9.6}$$

This double inequality has two consequences. First, ignoring the middle term, it implies

$$\int_{\theta_1}^{\theta_2} \int_{x(\theta_1)}^{x(\theta_2)} \frac{\partial^2 u}{\partial x \partial\theta}(x,\theta) dx d\theta \ge 0.$$

The Spence-Mirrlees assumption means the integrand is positive. Given $\theta_1 < \theta_2$, this means the integral can be non-negative only if $x(\theta_1) \le x(\theta_2)$. Since this is true for any $\theta_1 < \theta_2$, we may conclude that *the allocation function $x(\cdot)$ is non-decreasing.* Note that this necessarily implies that $x(\cdot)$ is almost everywhere continuous.

The second consequence of (9.6) is as follows: By fixing one end point and letting the other converge towards it, we see that $v(\cdot)$ is absolutely continuous with respect to Lebesgue measure and is, thus, almost everywhere differentiable.[6] This derivative is

$$\frac{dv\left(\theta\right)}{d\theta} = \frac{\partial u(x(\theta),\theta)}{\partial\theta}$$

almost everywhere.[7] Consequently, one can express equilibrium utility as

$$v(\theta) = v(\theta_L) + \int_{\theta_L}^{\theta} \frac{\partial u}{\partial\theta}(x(t),t)dt. \tag{9.7}$$

Expression (9.7) and the monotonicity of the allocation function $x(\cdot)$ are, thus, necessary properties of a direct-revelation mechanism. In particular, we've just proved that a necessary condition for an allocation function $x\left(\cdot\right)$ to be implementable is that

$$s(\theta) = v_L - u(x(\theta),\theta) + \int_{\theta_L}^{\theta} \frac{\partial u}{\partial\theta}(x(t),t)dt.$$

where v_L is an arbitrary constant.

It turns out that these properties are also *sufficient*:

---

[6]To be precise, one should make assumptions to ensure that $\partial u(x(\cdot),\cdot)/\partial\theta$ is integrable on Θ. This can be done by requiring that $\mathcal{X}$ be bounded or that $\partial u(\cdot,\cdot)/\partial\theta$ be bounded on $\mathcal{X} \times \Theta$. Both assumptions are natural in most economic settings and simply extend the assumptions that bound $x^F(\cdot)$. Henceforth, we assume $\partial u/\partial\theta$ is bounded.

[7]Note the important difference between $\frac{\partial u[x(\theta),\theta]}{\partial\theta}$ and $\frac{du[x(\theta),\theta]}{d\theta}$. The former is the *partial* derivative of u with respect to its second argument evaluated at $\left(x\left(\theta\right),\theta\right)$, while the latter is the total derivative of u.

Proposition 13 (Characterization) *Within the standard framework and under Assumption 3 (Spence-Mirrlees), a direct mechanism* $\langle x(\cdot), s(\cdot)\rangle$ *is truthful if and only if there exists a real number* v_L *such that:*

$$s(\theta) = v_L - u(x(\theta),\theta) + \int_{\theta_L}^{\theta} \frac{\partial u}{\partial \theta}(x(t),t)dt \tag{9.8}$$

$$\text{and } x(\cdot) \text{ is non-decreasing.} \tag{9.9}$$

Consequently, an allocation function $x(\cdot)$ *is implementable if and only if it is non-decreasing.*

Proof: Since we established necessity in the text, we need only prove sufficiency here. Let $\langle x(\cdot), s(\cdot)\rangle$ satisfy (9.8) and (9.9). Consider the agent's utility when the state of nature is θ, but he claims that it is $\theta' > \theta$:

$$s(\theta') + u(x(\theta'),\theta) = \overbrace{v_L - u(x(\theta'),\theta') + \int_{\theta_L}^{\theta'} \frac{\partial u}{\partial \theta}(x(t),t)dt}^{s(\theta')} + u(x(\theta'),\theta)$$

$$= \underbrace{v_L - u(x(\theta),\theta) + \int_{\theta_L}^{\theta} \frac{\partial u}{\partial \theta}(x(t),t)dt}_{s(\theta)} + u(x(\theta),\theta) \tag{9.10}$$

$$+ \left[u(x(\theta'),\theta) - u(x(\theta'),\theta')\right] + \left[\int_{\theta_L}^{\theta'} \frac{\partial u}{\partial \theta}(x(t),t)dt - \int_{\theta_L}^{\theta} \frac{\partial u}{\partial \theta}(x(t),t)dt\right] \tag{9.11}$$

$$= v(\theta) + \int_{\theta}^{\theta'} \left[\frac{\partial u}{\partial \theta}(x(t),t) - \frac{\partial u}{\partial \theta}(x(\theta'),t)\right] dt$$

Where the second equality (beginning of (9.10)) derives from adding and subtracting $s(\theta)$. In the last line, the first term, $v(\theta)$, is (9.10) and the second term, the integral, is (9.11). Since we've assumed (9.9), $x(t) \leq x(\theta')$ for $t \in [\theta, \theta']$. Moreover, A1 implies $\partial u/\partial\theta$ is increasing in x. Hence, the integral in the last line is non-positive; which means we may conclude

$$s(\theta') + u(x(\theta'),\theta) \leq v(\theta).$$

That is, under this mechanism, the agent does better to tell the truth than exaggerate his type. An analogous analysis can be used for $\theta' < \theta$ (*i.e.,* to show the agent does better to tell the truth than understate his type). Therefore, the revelation constraints hold and the mechanism is indeed truthful. ■

This characterization result is, now, a well-known result and can be found, implicitly at least, in almost every mechanism design paper. Given its importance, it is worth understanding how our assumptions drive this result. In particular, we wish to call attention to the fact that neither the necessity of (9.7) nor (9.8) depends on the Spence-Mirrlees assumption. The Spence-Mirrlees assumption's role is to establish (i) that a monotonic allocation function is necessary and (ii) that, if $x(\cdot)$ is monotonic, then (9.8) is sufficient to ensure a truth-telling equilibrium.

This discussion also demonstrates a point that was implicit in our earlier discussion of the Spence-Mirrlees assumption: What is critical is not that $\frac{\partial^2 u}{\partial\theta\partial x}$ be positive, but rather that it keep a constant sign over the relevant domain. If, instead of being positive, this cross-partial derivative were negative everywhere, then our analysis would remain valid, except that it would give us the inverse monotonicity condition: $x(\cdot)$ would need to be non-increasing in type. But with a simple change of the definition of type, $\tilde{\theta} = -\theta$, we're back to our original framework. Because, as argued above, the definition of type is somewhat arbitrary, we see that our conclusion of a non-*de*creasing $x(\cdot)$ is simply a consequence of the assumption that different types of agent have different marginal rates of substitution between money and allocation and that an ordering of these marginal rates of substitution by type is invariant to which point in $\mathcal{X} \times \mathcal{S}$ we're considering.

What if the Spence-Mirrlees assumption is violated (*e.g.*, $\frac{\partial^2 u}{\partial x\partial\theta}$ changes sign)? As our discussion indicates, although we still have necessary conditions concerning incentive-compatible mechanisms, we no longer have any reason to expect $x(\cdot)$ to be monotonic. Moreover—and more critically if we hope to characterize the set of incentive-compatible mechanisms—we have no sufficiency results. It is not surprising, therefore, that little progress has been made on the problem of designing optimal contracts when the Spence-Mirrlees condition fails.

9.3 Optimization in the Standard Framework

The previous analysis has given us, within the standard framework at least, a complete characterization of the space of possible (incentive-compatible) contracts. We can now concentrate on the principal's problem of designing an optimal contract.

Finding the optimal direct-revelation mechanism for the principal means maximizing the principal's expected utility over the set of mechanisms that induce truthful revelation of the agent's type and full participation. From page 84, the participation constraint is (9.2); while, from the previous section, truthful revelation is equivalent to (9.8) and (9.9).[8] We can, thus, express the principal's problem as

$$\max_{x(\cdot),s(\cdot)} \int_{\theta_L}^{\theta_H} [w(x(\theta),\theta) - s(\theta)]\, f(\theta)d\theta$$

subject to (9.2), (9.8), and (9.9).

[8] Assuming Assumption 3 holds, which we will do henceforth.

Once again, it's more convenient to work with $v(\cdot)$ than $s(\cdot)$. Observe that

$$\begin{aligned} w[x(\theta), \theta] - s(\theta) &= w[x(\theta), \theta] - v(\theta) + u[x(\theta), \theta] \\ &= \Omega(x(\theta), \theta) - v(\theta). \end{aligned}$$

Moreover, (9.8) can be used to compute $v(\theta)$ using only $x(\cdot)$ and a number v_L. Hence, we're free to write the principal's problem as

$$\begin{aligned} \int_{\theta_L}^{\theta_H} [w(x(\theta), \theta) - s(\theta)] f(\theta) d\theta &= \int_{\theta_L}^{\theta_H} [\Omega(x(\theta), \theta) - v(\theta)] f(\theta) d\theta \\ &= \int_{\theta_L}^{\theta_H} \left[\Omega(x(\theta), \theta) - \int_{\theta_L}^{\theta} \frac{\partial u}{\partial \theta}(x(t), t) dt\right] f(\theta) d\theta - v_L. \end{aligned}$$

Integration by parts implies

$$-\int_{\theta_L}^{\theta_H} \left[\int_{\theta_L}^{\theta} \frac{\partial u}{\partial \theta}(x(t), t) dt\right] f(\theta) d\theta = -\int_{\theta_L}^{\theta_H} [1 - F(\theta)] \frac{\partial u}{\partial \theta}(x(\theta), \theta) d\theta,$$

which allows us to further transform the principal's objective function:

$$\begin{aligned} &\int_{\theta_L}^{\theta_H} [w(x(\theta), \theta) - s(\theta)] f(\theta) d\theta \\ &\qquad = \int_{\theta_L}^{\theta_H} \left[\Omega(x(\theta), \theta) - \frac{1 - F(\theta)}{f(\theta)} \frac{\partial u}{\partial \theta}(x(\theta), \theta)\right] f(\theta) d\theta - v_L. \end{aligned}$$

Observation 2 *From this last expression, we can see that it is unreasonable to expect to achieve the first best: The principal's objective function differs from the first-best objective function,* $\max \mathbb{E}_\theta \{\Omega(x(\theta), \theta)\}$*, by*

$$-\int_{\theta_L}^{\theta_H} [1 - F(\theta)] \frac{\partial u [x(\theta), \theta]}{\partial \theta} d\theta.$$

Consequently, because the principal wishes to maximize something other than expected social surplus and the principal proposes the contract, we can't expect the contract to maximize social surplus.

Define

$$\Sigma(x, \theta) \equiv \Omega(x, \theta) - \frac{[1 - F(\theta)]}{f(\theta)} \frac{\partial u}{\partial \theta}(x, \theta).$$

Observe that our earlier assumptions ensure that $\Sigma(x, \theta)$ is bounded and at least twice-differentiable. We will refer to $\Sigma(x, \theta)$ as the *virtual surplus* (this follows Jullien, 1996).[9]

[9]Guesnerie and Laffont (1984) and Caillaud et al. (1988) use the term *surrogate welfare function*.

The principal's problem can now be restated in a tractable and compact form:

$$\max_{x(\cdot),v_L} \left\{ \int_{\theta_L}^{\theta_H} \Sigma(x(\theta),\theta) f(\theta) d\theta - v_L \right\} \tag{9.12}$$

$$\text{subject to } v_L + \int_{\theta_L}^{\theta} \frac{\partial u}{\partial \theta}(x(t),t)dt \geq U_R$$

and $x(\cdot)$ is non-decreasing.

Although this problem is solvable, we will focus on only a special (but widely employed) case. Specifically, we make three additional assumptions:

1. For all x, $\partial u/\partial\theta \geq 0$ (*i.e.,* utility is non-decreasing in type);
2. $\Sigma(\cdot,\theta)$ is strictly quasi-concave for all $\theta \in [\theta_L, \theta_H]$; and
3. $\partial\Sigma/\partial x$ is non-decreasing in θ for all x.

What are the consequences of these assumptions? The first entails that the participation constraint holds for all types if it holds for the lowest type, θ_L. Consequently, we can ignore this constraint for all but the lowest type. Moreover, for this type, the constraint reduces to $v_L \geq U_R$. Since v_L is a direct transfer to the agent without incentive effects, we know the principal will set it as low as possible. That is, we can conclude that, optimally, $v_L = U_R$. Note, too, this means the participation constraint is binding for the lowest type (similarly to what we saw in the retailer-supplier example). To summarize:

Lemma 8 *If utility is non-decreasing in type for all allocations (i.e., $\partial u/\partial\theta \geq 0$ for all x), then (i) $v_L = U_R$; (ii) the participation constraint is binding for the lowest type, θ_L; and (iii) the participation constraint holds trivially for all higher types (i.e., for $\theta > \theta_L$).*

In light of this lemma, we can be emboldened to try the following solution technique for (9.12): Ignore the monotonicity constraint and see if the *un*constrained problem yields a monotonic solution. The solution to the unconstrained problem, $x^*(\cdot)$, is to solve (9.12) pointwise; that is, to set $x^*(\theta) = X(\theta)$, where

$$X(\theta) \equiv \arg\max_x \Sigma(x,\theta).$$

Note that the second assumption means $X(\cdot)$ is uniquely defined. Finally, the third assumption—the marginal-benefit schedule, $\partial\Sigma/\partial x$, is non-decreasing in θ—means the point at which $\partial\Sigma(x,\theta)/\partial x$ crosses zero is non-decreasing in θ. But this point is $X(\theta)$; hence, monotonicity is ensured. To conclude:

Proposition 14 *If*

- *for all x, $\partial u/\partial\theta \geq 0$;*

- $\Sigma(\cdot, \theta)$ *is strictly quasi-concave; and*
- $\partial\Sigma/\partial x$ *is non-decreasing in* θ*;*

then the solution to (9.12) is $x^*(\theta) = X(\theta)$ *and* $v_L = U_R$*.*

How does the solution in Proposition 14 compare to the full-information benchmark? The answer is given by the following corollaries:

Corollary 3 $x^*(\theta) < x^F(\theta)$ *for all* $\theta \in [\theta_L, \theta_H)$ *and* $x^*(\theta_H) = x^F(\theta_H)$*.*

Proof: At $x = x^F(\theta)$,

$$\frac{\partial\Sigma}{\partial x} = -\frac{1-F(\theta)}{f(\theta)} \times \frac{\partial^2 u}{\partial x \partial\theta}.$$

Since (i) $1 - F(\theta) > 0$ (except for $\theta = \theta_H$) and (ii) the cross-partial derivative is strictly positive by A1, the right-hand side is negative for all θ, except θ_H. Consequently, since $\Sigma(\cdot, \theta)$ is strictly quasi-concave, we can conclude that $x^*(\theta) < x^F(\theta)$ for all $\theta \in [\theta_L, \theta_H)$. For $\theta = \theta_H$, we've just seen that

$$\frac{\partial\Sigma\left[x^F(\theta_H), \theta_H\right]}{\partial x} = 0;$$

hence, the strict quasi-concavity of $\Sigma(\cdot, \theta_H)$ ensures that $x^F(\theta_H)$ is the maximum. ■

Corollary 4 $v'(\theta) \geq 0$ *and* $v(\theta_L) = U_R$*.*

Proof: We've already established the second conclusion. The first follows since

$$v(\theta) = v_L + \int_{\theta_L}^{\theta} \frac{\partial u}{\partial\theta}(x(t), t)dt;$$

so

$$v'(\theta) = \frac{\partial u}{\partial\theta}[x(\theta), \theta] \geq 0$$

by the first assumption. ■

In much of the literature, the three additional assumptions supporting Proposition 14 hold, so Proposition 14 and its corollaries might be deemed the "standard solution" to the contractual screening problem within the standard framework. These conclusions are sometimes summarized as

Observation 3 *Under the* standard solution*, there's a downward distortion in allocation (relative to full information) for all types but the highest, the lowest type earns* no *information rent, but higher types may.*

Observe the "may" at the end of the last remark becomes a "do" if $\partial u[X(\theta), \theta]/\partial\theta > 0$ for all $\theta > \theta_L$.[10]

[10] There are, of course, many assumptions that will change the "may" to a "do."

9.4 The Retailer-Supplier Example Revisited

Before returning to our fairly abstract analysis of contractual screening, let's consider an extension of our earlier retailer-supplier example. Specifically, let's imagine that there are a continuum of efficiency types, which we normalize to be the interval $[1, 2]$. Instead of $C_t(x)$, write the supplier's cost function as $C(x, \theta)$, and suppose that

$$\frac{\partial^2 C}{\partial\theta\partial x} < 0; \tag{9.13}$$

that is, higher (more efficient) types have lower marginal costs. Because

$$u(x, \theta) = -C(x, \theta),$$

(9.13) implies that the Spence-Mirrlees assumption is met (recall Lemma 7). In addition to these assumptions, maintain all the assumptions from our earlier model. In particular, we assume the revenue function, $r(\cdot)$, is concave and bounded. Hence, because $C(\cdot, \theta)$ is strictly convex,

$$\lim_{x\to\infty} \Omega(x, \theta) = -\infty \text{ for all } \theta.$$

Assume, too, that $\partial\Omega(0, \theta)/\partial x > 0$ for all θ; *i.e.*, $x^F(\theta) > 0$ for all θ. It is readily checked that all the assumptions of the standard framework are, therefore, satisfied.

Since $C(\cdot, \theta)$ is a cost function, we necessarily have $C(0, \theta) = 0$ for all θ. Combined with (9.13), this entails that $C(x, \theta) > C(x, \theta')$ if $\theta < \theta'$.[11] Hence, we may conclude

$$\frac{\partial u}{\partial\theta}(x, \theta) > 0. \tag{9.14}$$

Observe that

$$\Sigma(x, \theta) = r(x) - C(x, \theta) - \frac{1 - F(\theta)}{f(\theta)}\left[\frac{-\partial C}{\partial\theta}\right].$$

Hence,

$$\frac{\partial\Sigma(x, \theta)}{\partial x} = r'(x) - \frac{\partial C}{\partial x} - \frac{1 - F(\theta)}{f(\theta)}\left[\frac{-\partial^2 C}{\partial x\partial\theta}\right].$$

It is clear, therefore, that to take advantage of Proposition 14, we need to know something about the shape of $\partial C(\cdot, \theta)/\partial\theta$ (*e.g.*, is it at least quasi-concave)

[11] **Proof:** The initial condition, $C(0, t) = 0$ for all $t \in [1, 2]$, means we can write

$$C(x, \theta') - C(x, \theta) = \int_0^x \int_\theta^{\theta'} \frac{\partial^2 C}{\partial\theta\partial x} d\theta dx;$$

the result follows from (9.13).

and how the Mills ratio[12] and the cross-partial derivative change with respect to θ. To this end, let's impose two frequently made assumptions:

- $C(x,\theta) = \beta(\theta)\, c(x)$, where $\beta(\cdot)$ is positive, strictly *de*creasing, and convex—higher types have lower marginal costs, but this marginal cost advantage may be less pronounced when moving up from one high type to another than it is when moving up from one low type to another. The function $c(\cdot)$ is strictly increasing, strictly convex, and $c(0) = 0$.

- Let $M(\theta)$ denote the Mills ratio. Assume $M'(\theta) \leq 1$. Observe this is our usual monotone hazard rate property.

Given these assumptions, we may conclude that $\Sigma(\cdot,\theta)$ is globally strictly concave and that

$$\begin{aligned}\frac{\partial^2 \Sigma(x,\theta)}{\partial\theta\partial x} &= -\beta'(\theta)\, c'(x) + M'(\theta)\, \beta'(\theta)\, c'(x) + M(\theta)\, \beta''(\theta)\, c'(x) \\ &\propto \beta'(\theta)\left[M'(\theta) - 1\right] + M(\theta)\, \beta''(\theta) \geq 0.\end{aligned}$$

That is, we may conclude that $\Sigma(\cdot,\theta)$ admits a unique maximum, which is non-decreasing with type. Combined with (9.14), this means we can apply Proposition 14.

For example, suppose $r(x) = x$, $\beta(\theta) = 1/\theta$, $c(x) = x^2/2$, and $F(\theta) = \theta - 1$ (*i.e.,* the uniform distribution on $[1,2]$). Both bullet points are met, so we can apply Proposition 14. This yields $x^*(\theta) = X(\theta)$, where

$$\begin{aligned} X(\theta) \text{ solves } 1 - \frac{x}{\theta} - (2-\theta)\frac{x}{\theta^2} &= 0; \\ \text{or } X(\theta) &= \frac{1}{2}\theta^2. \end{aligned}$$

From Proposition 14, we may set $v_L = U_R$, which is zero in this model. Noting that $\partial u/\partial\theta = x^2/2\theta^2$, we thus have

$$\begin{aligned} v(\theta) &= 0 + \int_{\theta_L}^{\theta} \frac{x^*(t)^2}{2t^2} dt \\ &= \int_1^{\theta} \frac{t^2}{8} dt \\ &= \frac{1}{24}\theta^3 - \frac{1}{24}. \end{aligned}$$

[12] The Mills ratio is the ratio of a survival function (here, $1 - F(\theta)$) to its density function (here, $f(\theta)$). Because the Mills ratio is the inverse of the hazard rate, it is also known as the *inverse* hazard rate (this formal distinction between the inverse and "regular" hazard rate is not always respected by economic theorists).

Hence, the *transfer* function is

$$\begin{aligned} s(\theta) &= v(\theta) - u(x(\theta), \theta) \\ &= \frac{1}{24}\theta^3 - \frac{1}{24} + \frac{\theta^3}{8} \\ &= \frac{\theta^3}{6} - \frac{1}{24}. \end{aligned}$$

Observe that

$$x^*(\theta) = \frac{1}{2}\theta^2 < \theta = x^F(\theta)$$

for all $\theta < 2$; that $x^*(2) = 2 = x^F(2)$; that $v(1) = 0$; and that $v(\theta) > 0$ for all $\theta > 1$—all consistent with the corollaries to Proposition 14.

Finally, although there's no need to do it, we can check that the incentive-compability constraints are indeed satisfied by the mechanism $\left\langle \frac{1}{2}\theta^2, \frac{1}{6}\theta^3 - \frac{1}{24} \right\rangle$:

$$\begin{aligned} \max_{\hat{\theta}} s(\hat{\theta}) - \beta(\theta)c\big(x^*(\hat{\theta})\big) \quad &= \quad \max_{\hat{\theta}} \frac{1}{6}\hat{\theta}^3 - \frac{1}{24} - \frac{\left[\frac{1}{2}\hat{\theta}^2\right]^2}{2\theta} \\ &\Longrightarrow \quad \frac{\hat{\theta}^2}{2} - \frac{\hat{\theta}^3}{2\theta} = 0. \end{aligned}$$

Clearly, $\hat{\theta} = \theta$ satisfies the first-order condition (since the program is clearly concave, the second-order conditions are also met).

By the taxation principle (Proposition 12), an alternative to this direct-revelation contract is a payment schedule. Noting that $x^{*-1}(x) = \sqrt{2x}$, an optimal payment schedule, $S(x)$, is

$$S(x) = \begin{cases} s\left[x^{*-1}(x)\right] = \frac{x^{3/2}\sqrt{2}}{3} - \frac{1}{24} \text{ for } x \in \left[\frac{1}{2}, 2\right] \\ 0 \text{ for } x \notin \left[\frac{1}{2}, 2\right] \end{cases} .$$

The Hidden-Knowledge Model | 10

In this lecture, we consider a model that, at first, seems quite different than contractual screening, but which ultimately shares many similarities to it. In particular, we consider the *hidden-knowledge* model. In this model, unlike above, the principal and agent are *symmetrically* informed at the time they enter into a contractual arrangement. *After* contracting, the agent acquires private information (his hidden knowledge). As an example, suppose the principal employs the agent to do some task—for instance, build a well on the principal's farm—initially, both parties could be symmetrically informed about the difficulty of the task (*e.g.,* the likely composition of the rock and soil, how deep the water is, etc.). However, once the agent starts, he may acquire information about how hard the task really is (*e.g.,* he alone gains information that better predicts the depth of the water).

This well-digging example reflects a general problem. In many employment situations, the technological, organizational, market, and other conditions that an employee will face will become known to him only after he's been employed by the firm. This information will affect how difficult his job is, and, thus, his utility. Similarly, think of two firms that want to engage in a specific trade (*e.g.,* a parts manufacturer and an automobile manufacturer who contract for the former to supply parts meeting the latter's unique specifications). Before the contract is signed, the supplier may not know much about the cost of producing the specific asset and the buyer may have little knowledge about the prospects of selling the good to downstream consumers. These pieces of information will flow in during the relationship—*but after contracting*—and once again a hidden-knowledge framework is more appropriate for studying such a situation.

This difference in timing is reflected in the participation constraint for this problem. When considering a contract, $\langle x(\cdot), s(\cdot)\rangle$, the agent compares his *expected utility* if he accepts the contract,

$$\mathbb{E}\left\{u(\theta)\right\} = \mathbb{E}\left\{s(\theta) + u\left[x(\theta), \theta\right]\right\},$$

to his expected utility if he refuses the contract; that is, to $U_R^a \equiv \mathbb{E}\left\{U_R(\theta)\right\}$. Acceptation or refusal of the contract cannot depend upon the, as yet, unrealized state of nature: Unlike the screening model, the participation decision is *not* contingent on type.

Why is this distinction so important? Because it turns out that, at least in the standard framework, the hidden-knowledge model has an extremely simple solution. To see this, we invoke the assumptions of the standard framework, including the Spence-Mirrlees assumption. In addition, we assume—consistent with the assumptions of the standard framework—that $\partial\Omega/\partial x$ is non-decreasing in θ. Consequently, we know the first-best allocation, $x^F(\cdot)$, is non-decreasing. We can then be sure from Proposition 13 that there exists a transfer function, $\hat{s}^F(\cdot)$, such that $\left\langle x^F(\cdot), \hat{s}^F(\cdot)\right\rangle$ is a direct-revelation mechanism.[1] Note that because $\hat{s}^F(\cdot)$ is defined by (9.8), it is defined up to a constant that can be chosen by the principal to ensure the agent meets his participation constraint. In particular, it can be chosen so that the agent's expected utility equals his non-participation expected utility:

$$\int_{\theta_L}^{\theta_H} \left[\hat{s}^F(\theta) + u\left(x^F(\theta), \theta\right)\right] f(\theta) d\theta = \int_{\theta_L}^{\theta_H} U_R(\theta) f(\theta) d\theta.$$

With this mechanism, the principal's expected utility becomes:

$$\begin{aligned}\int_{\theta_L}^{\theta_H} \left[w(x^F(\theta), \theta) - \hat{s}^F(\theta)\right] f(\theta) d\theta &= \int_{\theta_L}^{\theta_H} \left[\Omega(x^F(\theta), \theta) - U_R(\theta)\right] f(\theta) d\theta \\ &= \int_{\theta_L}^{\theta_H} \Omega(x^F(\theta), \theta) f(\theta)\, d\theta - U_R^a.\end{aligned}$$

Given that the *ex post* efficient allocation—that is, $x^F(\cdot)$—maximizes the integrand in the last integral, the principal obtains the highest possible expected utility with this mechanism. Hence, $\left\langle x^F(\cdot), \hat{s}^F(\cdot)\right\rangle$ is optimal and we see, therefore, that the first-best allocation will be achieved with hidden-knowledge, in contrast to the less desirable equilibrium of the screening model:

Proposition 15 *In a hidden-knowledge model, in which the agent fully commits to the contract* before *learning his type, which satisfies the assumptions of the standard framework, and in which $\partial\Omega/\partial x$ is non-decreasing in θ, then the equilibrium allocation is the* ex post *efficient allocation.*

To gain intuition for this result, return to our two-type example from Lecture Note 7, but suppose now that the agent doesn't learn his type until *after* contracting with the principal. Observe that we can reduce the agent's information rent *on average* by paying him less than his full-information payment if he announces he is the low type (*i.e.,* type I) and more than his full-information payment if he announces he is the high type (*i.e.,* type E): Now set the payments to be

$$\begin{aligned}\hat{s}_I^F &= C_I\left(x_I^F\right) - \gamma; \text{ and} \\ \hat{s}_E^F &= C_E\left(x_E^F\right) + \eta.\end{aligned}$$

Since the agent doesn't learn his type until after contracting, the participation constraint is

$$\begin{aligned}f \times \left(\hat{s}_I^F - C_I\left(x_I^F\right)\right) + (1-f) \times \left(\hat{s}_E^F - C_E\left(x_E^F\right)\right) &= -f\gamma + (1-f)\,\eta \\ &\geq 0.\end{aligned}$$

[1] The transfer function $\hat{s}^F(\cdot)$ is *a priori* different from the full-information transfer function, $s^F(\cdot)$, because the latter is not subject to revelation constraints.

We also need direct revelation, which, for type I, means

$$\begin{aligned}\hat{s}_I^F - C_I\left(x_I^F\right) &\geq \hat{s}_E^F - C_I\left(x_E^F\right) \text{; or}\\ -\gamma &\geq \eta + \left[C_E\left(x_E^F\right) - C_I\left(x_E^F\right)\right].\end{aligned}$$

Treating these two constraints as equalities, we can solve for γ and η:

$$\begin{aligned}\eta &= f \times \left[C_I\left(x_E^F\right) - C_E\left(x_E^F\right)\right] \text{; and}\\ \gamma &= (1-f) \times \left[C_I\left(x_E^F\right) - C_E\left(x_E^F\right)\right].\end{aligned}$$

Provided these also satisfy type E's revelation constraint, we're done. But they do, since

$$\begin{aligned}\hat{s}_E^F - C_I\left(x_E^F\right) &= \eta\\ &= -\gamma + C_I\left(x_E^F\right) - C_E\left(x_E^F\right)\\ &> -\gamma + C_I\left(x_I^F\right) - C_E\left(x_I^F\right)\\ &= \hat{s}_I^F - C_E\left(x_I^F\right)\end{aligned}$$

(the inequality follows because, recall, $C_I(\cdot) - C_E(\cdot)$ is increasing).

Note the phrase *"in which the agent fully commits to the contract"* that constitutes one of the assumptions in Proposition 15. Why this assumption? Well suppose that, after learning his type, the agent could quit (a reasonable assumption if the agent is a person who enjoys legal protections against slavery). If his payoff would be less than $U_R(\theta)$ if he played out the contract, he would do better to quit.[2] To keep the agent from quitting, the principal would have to design the contract so that the agent's equilibrium utility was not less than $U_R(\theta)$ for all θ. But then this is just the screening model again! In other words, the hidden-knowledge model reverts to the screening model—with all the usual conclusions of that model—if the agent is free to quit (in the parlance of the literature, if *interim participation constraints* must be met). Even if anti-slavery protections don't apply (*e.g.,* the agent is a firm), interim participation could still matter; for instance, in the last example, if γ is too big, then the agent may not have the financial resources to pay it (it would bankrupt him). Alternatively, in nations with an English law tradition, γ could be perceived as a penalty, and in many instances the courts will refuse to enforce contracts that call for one party to pay a penalty to another. In short, because *interim* participation constraints are often a feature of the real world, many situations that might seem to fit the hidden-knowledge model will ultimately prove to be screening-model problems instead.

[2] Observe we're assuming that $U_R(\cdot)$ doesn't change—in particular, doesn't diminish—after the agent signs the contract; that is, his outside option remains the same over time.

Multiple Agents and Implementation | 11

Our study of mechanism design has so far restricted attention to the case of one principal and one agent (this is true even of our analysis of second-degree price discrimination because, as seen in Section 5.2, the analysis can be done in terms of a single representative consumer). In this lecture note, I turn to situations with multiple agents. Moreover, the focus will be on situations—such as public goods problems—in which the agents *must* play the mechanism (*e.g.,* because they are citizens of the state). Such problems are known as *implementation* problems.

Public Choice Problems | 11.1

Societies frequently face the problem of making public choices. For instance, how many miles of highway to build, how many acres to set aside for park land, or how many tons of pollution to permit. Because everyone in the society is affected by these choices, they are known as *public-choice problems*.

As an ideal, suppose there were a benevolent social planner, a person or agency that sought to choose the best solution to the public-choice problem. Of course, "best solution" begs the question of what criterion the social planner should use to choose among possible allocations. Because the social planner is benevolent, we presume she seeks to maximize some social welfare measure. This measure, $\mathcal{W}$, is assumed to be a function of the public choice variable, $a \in \mathcal{A}$, that aggregates in some form the preferences of the N agents, where agent n's preferences are captured by his type, $\tau_n \in \mathcal{T}_n$.

Each agent has a utility $u_n(a, p_n|\tau_n)$, where $p_n \in \mathbb{R}$ is the amount of the private consumption good (effectively, money) transferred to the nth agent. Observe $p_n > 0$ means money is being transferred *to* agent n, while $p_n < 0$ means money is being transferred *from* agent n.

For example, each agent could have the utility function

$$u_n(a, p_n|\tau_n) = \tau_n a - \frac{a^2}{2} + p_n \qquad (11.1)$$

and the social planner could seek to maximize

$$\mathcal{W} = \sum_{n=1}^{N}\left(u_n(a,p_n|\tau_n) - p_n\right) = \sum_{n=1}^{N}\left(\tau_n a - \frac{a^2}{2}\right)\,; \tag{11.2}$$

that is, the social planner seeks to maximize the sum of the utilities of the agents less the social cost of paying them transfers (given quasi-linear utility for all citizens, this cost is just the value of the transfer).

The standard assumption is that $\{u_n(\cdot)\}_{n=1}^N$ is common knowledge, but that τ_n is the private information of agent n. The joint distribution function for $\boldsymbol{\tau} = (\tau_1,\ldots,\tau_N)$ is, however, common knowledge. In other words, while everyone knows the functional forms of the utility functions and the distribution of types, the realization of a given agent's type is known only to him. Let

$$\boldsymbol{\tau} = (\tau_1,\ldots,\tau_n) \in \mathcal{T} = \mathcal{T}_1 \times \cdots \times \mathcal{T}_n\,.$$

Mechanisms 11.2

At its most general, a *mechanism* can be defined as follows:

Definition 2 *A* mechanism *for N agents is*

- *a collection of N message spaces $\{\mathcal{M}_n\}_{n=1}^N$;*
- *rules concerning the order in which messages are sent and who hears what messages when (i.e., an extensive form); and*
- *a mapping $\sigma : \times_{n=1}^N \mathcal{M}_n \to \mathcal{A}\times\mathbb{R}^N$.*

For future reference, define $\mathcal{M} = \times_{n=1}^N \mathcal{M}_n$. Let $\mathbf{m} = (m_1,\ldots,m_N)$ denote an element of $\mathcal{M}$. Observe that σ can be decomposed into $N+1$ functions: $\sigma(\mathbf{m}) = \big(A(\mathbf{m}),P_1(\mathbf{m}),\ldots,P_N(\mathbf{m})\big)$, where $A:\mathcal{M}\to\mathcal{A}$ and $P_n:\mathcal{M}\to\mathbb{R}$.

Observe that a mechanism is a game to be played by the N agents. It defines actions, the $\mathcal{M}_n$; it defines an extensive form; and it defines results, which determine payoffs (*i.e.,* agent n's payoff if $\mathbf{m}$ is played is $u_n\big(A(\mathbf{m}),P_n(\mathbf{m})|\tau_n\big)$ if he's type τ_n). Typically, the extensive form is simultaneous announcements; that is, each agent sends his message without observing the messages sent by the other agents. We will, in fact, restrict attention to simultaneous-announcement mechanisms in what follows.

Assumption 4 *Mechanisms are simultaneous-announcement mechanisms.*

When, as will be the case shortly, attention is restricted to dominant-strategy mechanisms, this assumption will, ultimately, not matter; but assuming it now simplifies the discussion.

A *strategy* for an agent is a mapping from what he knows, which, in light of Assumption 4, can be summarized by his type, τ_n, into his message space, $\mathcal{M}_n$. That is, if $\mathbb{M}_n$ is the set of all functions from $\mathcal{T}_n$ into $\mathcal{M}_n$, then a strategy is the selection of an element, $m_n(\cdot)$, from $\mathbb{M}_n$.

A *solution concept* is the rule for solving the game defined by the mechanism. Common solution concepts are dominant strategy (to be explored in the next section) and Bayesian Nash (to be explored in Section 11.7). Observe that the choice of the solution concept is part of the design of the mechanism—a mechanism is designed to be solved according to a specified concept.

11.3 Dominant-Strategy Mechanisms

A strategy for agent n is a *dominant strategy* if agent n wishes to play it regardless of what messages the *other* agents will send; that is, $m_n^*(\cdot)$ is a dominant strategy for agent n if, for each $\tau_n \in \mathcal{T}_n$ and every $\mathbf{m}_{-n} \in \times_{j \neq n} \mathcal{M}_j$,

$$u_n\left(A\big[m_n^*(\tau_n), \mathbf{m}_{-n}\big], P_n\big[m_n^*(\tau_n), \mathbf{m}_{-n}\big] \big| \tau_n\right) \geq u_n\left(A\big[m_n, \mathbf{m}_{-n}\big], P_n\big[m_n, \mathbf{m}_{-n}\big] \big| \tau_n\right) \ \forall m_n \in \mathcal{M}_n\,. \tag{11.3}$$

If *all* agents have a dominant strategy, then the mechanism has a solution in dominant strategies:

Definition 3 *A mechanism has a* solution in dominant strategies *if each agent n has a dominant strategy given that mechanism; that is, for each n, there exists a mapping $m_n^*(\cdot) : \mathcal{T}_n \to \mathcal{M}_n$ such that* (11.3) *holds for all $\tau_n \in \mathcal{T}_n$ and all $\mathbf{m}_{-n} \in \times_{j \neq n} \mathcal{M}_j$. A mechanism with a solution in dominant strategies is a* dominant-strategy mechanism.

Observe that because agents are playing dominant strategies, it doesn't matter whether a given agent knows the announcements of the other agents when he chooses his announcement or not. That is, these dominant strategy mechanisms are robust to relaxing the assumption that announcements are simultaneous.

11.4 The Revelation Principle

In the single-principal-single-agent model, we saw that attention could be restricted, without loss of generality, to direct-revelation mechanisms (see Lecture 8, particularly Proposition 11). Hence, rather than trying to work with arbitrary classes of message spaces, attention could be limited to messages spaces that were the same as the type space (*i.e.,* the agent made announcements as to his type). The same result—the *revelation principle*—holds with many agents.

To be formal, recall that a mechanism is a *direct mechanism* if $\mathcal{M}_n = \mathcal{T}_n$ for all n; that is, a mechanism is direct if the space of possible messages is limited to announcements about type.

Define $\mathcal{E}_{\sigma,k} : \mathcal{T} \to \mathcal{M}$ to be the set of messages (the *message profile*) sent in the equilibrium of mechanism σ as a function of the realized type profile, where equilibrium is defined by the solution concept ("*k*oncept") k. So, for instance, if the solution concept is dominant strategy (*i.e.,* $k = \text{DS}$), then $\mathcal{E}_{\sigma,\text{DS}}(\boldsymbol{\tau})$ is the set of $\mathbf{m}^*$ that represent a dominant-strategy solution to the mechanism σ when the realized profile of types is $\boldsymbol{\tau}$.

Definition 4 *A direct mechanism σ is a* direct-revelation mechanism *under solution concept k if $\boldsymbol{\tau} \in \mathcal{E}_{\sigma,k}(\boldsymbol{\tau})$ for all $\boldsymbol{\tau} \in \Theta$. That is, if all agents telling the truth is always an equilibrium under solution concept k.*

Note, to be a solution under a solution concept, it cannot be that an agent does better by deviating from his part of that solution; that is, there cannot be profitable unilateral deviations. To avoid unduly complicating the notation, we will limit attention in what follows to solution concepts that have solutions in pure-strategies.[1] Because, under some solution concepts, each agent n needs to calculate the expected value of his utility over the other agents types, $\boldsymbol{\tau}_{-n}$, conditional on his own type, τ_n, define $U_n(a, p_n|\tau_n)$ to be that expected utility; that is,

$$U_n(a, p_n|\tau_n) \equiv \mathbb{E}_{\{\boldsymbol{\tau}_{-n} \in \mathcal{T}_{-n}\}}\big\{u_n(a, p_n|\tau_n)\big|\tau_n\big\}\,.$$

With this apparatus in hand, we can now state and prove the *revelation principle* for multiple-agent mechanisms.

Proposition 16 (The Revelation Principle) *Let σ be a mechanism under solution concept k and let $A_{\sigma,k}(\cdot)$ and $\mathbf{P}_{\sigma,k}(\cdot)$ be, respectively, the* equilibrium *allocation and transfer vector as a function of the agents' types. Then there exists a direct-revelation mechanism, ψ, with a solution under concept k such that $A_{\sigma,k}(\boldsymbol{\tau}) = A_{\psi,k}(\boldsymbol{\tau})$ and $\mathbf{P}_{\sigma,k}(\boldsymbol{\tau}) = \mathbf{P}_{\psi,k}(\boldsymbol{\tau})$ for all type profiles $\boldsymbol{\tau} \in \mathcal{T}$. That is, given that the class of mechanisms has been restricted to being solvable under concept k, there is no further loss of generality in restricting attention to direct-revelation mechanisms solvable under k.*

Proof: Let σ be the simultaneous-announcement mechanism $\mathcal{M}_1 \times \cdots \times \mathcal{M}_N \equiv \mathcal{M}$, $A : \mathcal{M} \to \mathcal{A}$, and $\mathbf{P} : \mathcal{M} \to \mathbb{R}^N$. Let $\mathbf{m}^*(\cdot)$ be a solution under k. Observe this implies

$$U_n\bigg(A\big(m_n^*(\tau_n), \mathbf{m}_{-n}^*(\boldsymbol{\tau}_{-n})\big), P_n\big(m_n^*(\tau_n), \mathbf{m}_{-n}^*(\boldsymbol{\tau}_{-n})\big)\bigg|\tau_n\bigg) \geq$$
$$U_n\bigg(A\big(m_n, \mathbf{m}_{-n}^*(\boldsymbol{\tau}_{-n})\big), P_n\big(m_n, \mathbf{m}_{-n}^*(\boldsymbol{\tau}_{-n})\big)\bigg|\tau_n\bigg)\ \forall m_n \in \mathcal{M}_n\,. \quad (11.4)$$

[1] Extending the analysis to permit mixed-strategies is straightforward. Note, however, that in practice most mechanisms are designed to be solvable in pure strategies.

Consider the direct mechanism ψ defined by $\tilde{\mathbf{P}} : \mathcal{T} \to \mathbb{R}^N$ and $\tilde{A} : \mathcal{T} \to \mathcal{A}$, where $\tilde{\mathbf{P}}(\boldsymbol{\tau}) \equiv \mathbf{P}(\mathbf{m}^*[\boldsymbol{\tau}])$ and $\tilde{A}(\boldsymbol{\tau}) \equiv A(\mathbf{m}^*[\boldsymbol{\tau}])$. Clearly, the equivalence of the allocation and payments under ψ and σ follows *if* ψ is a direct-revelation mechanism. So, to complete the proof, we need to establish that ψ is a direct-revelation mechanism with a solution under k. To this end, suppose ψ were *not* a direct-revelation mechanism with a solution under k. Then there must exist an agent n and a type τ_n for that agent such that there is a $\hat{\tau}_n \in \mathcal{T}_n$ for which the following is true:

$$U_n\left(\tilde{A}(\hat{\tau}_n, \boldsymbol{\tau}_{-n}), \tilde{P}_n(\hat{\tau}_n, \boldsymbol{\tau}_{-n}) \big| \tau_n\right) > U_n\left(\tilde{A}(\tau_n, \boldsymbol{\tau}_{-n}), \tilde{P}_n(\tau_n, \boldsymbol{\tau}_{-n}) \big| \tau_n\right). \quad (11.5)$$

But substituting for the tilde-functions in (11.5) yields

$$\begin{aligned} &U_n\left(A\big[m_n^*(\hat{\tau}_n), \mathbf{m}_{-n}^*(\boldsymbol{\tau}_{-n})\big], P_n\big[m_n^*(\hat{\tau}_n), \mathbf{m}_{-n}^*(\boldsymbol{\tau}_{-n})\big] \big| \tau_n\right) > \\ &\qquad U\left(A\big[m_n^*(\tau_n), \mathbf{m}_{-n}^*(\boldsymbol{\tau}_{-n})\big], P_n\big[m_n^*(\tau_n), \mathbf{m}_{-n}^*(\boldsymbol{\tau}_{-n})\big] \big| \tau_n\right). \quad (11.6) \end{aligned}$$

But expression (11.6) contradicts (11.4). In other words, expression (11.6) entails that a type τ_n agent n does better to send the message $m_n^*(\hat{\tau}_n)$ than he does by sending the message $m_n^*(\tau_n)$; which, in turn, contradicts the fact that $m_n^*(\cdot)$ is a solution under concept k. ■

Groves-Clarke Mechanisms 11.5

Recall the public-goods problem laid out on page 103; that is, where the agents' utilities and social welfare function are defined by (11.1) and (11.2), respectively. Assume $\mathcal{T}_n = (0, T)$, where $T > 0$. From (11.2), social welfare—given $\boldsymbol{\tau}$—is simply a function of a. Moreover, it is a globally concave function. This latter observation means that the first-order condition for maximizing $\mathcal{W}(a)$,

$$\mathcal{W}'(a) = \sum_{n=1}^{N}(\tau_n - a) = 0\,, \quad (11.7)$$

is sufficient, as well as necessary. Observe that the solution to (11.7) is

$$a^*(\boldsymbol{\tau}) = \frac{\sum_{n=1}^{N} \tau_n}{N}\,. \quad (11.8)$$

The social planner's objective is, thus, to devise a mechanism that implements $a^*(\cdot)$.

In this section, we restrict attention to dominant-strategy mechanisms. Moreover, the Revelation Principle (Proposition 16) allows us to restrict attention to direct-revelation mechanisms without any loss of generality. That is, in our direct-revelation mechanism, we would like $A(\boldsymbol{\tau}) = a^*(\boldsymbol{\tau})$ for all $\boldsymbol{\tau} \in \mathcal{T} \equiv (0, T)^N$.

To implement such an $A(\cdot)$, we need to find transfer functions (*i.e.,* $P_n(\cdot)$) such that truth-telling is a dominant strategy given that the allocation function is $A(\cdot)$. That is, $\mathbf{P} : (0,T)^N \to \mathbb{R}^N$ needs to satisfy

$$\tau_n \in \underset{\hat{\tau}_n}{\operatorname{argmax}} \ \tau_n A(\hat{\tau}_n, \boldsymbol{\tau}_{-n}) - \frac{1}{2}A(\hat{\tau}_n, \boldsymbol{\tau}_{-n})^2 + P_n(\hat{\tau}_n, \boldsymbol{\tau}_{-n}) \tag{11.9}$$

for all n and all $\boldsymbol{\tau} \in (0,1)^N$ (recall expression (11.3) above). That is, telling the truth, $\hat{\tau}_n = \tau_n$, has to be dominant strategy (*i.e.,* maximize utility) regardless of the *other* agents' announcements, $\boldsymbol{\tau}_{-n}$.

Let us hypothesize that there exists a differentiable $P_n(\cdot)$ function for each agent n that solves (11.9) and, moreover, makes the right-hand side of (11.9) globally concave (so that first-order conditions are sufficient as well as necessary). We will, of course, need to verify later that our solution makes (11.9) globally concave. Differentiating the right-hand side of (11.9) with respect to $\hat{\tau}_n$ and evaluating at $\hat{\tau}_n = \tau_n$, we must have the following first-order condition met for all n and all $\boldsymbol{\tau} \in (0,T)^N$:

$$\big(\tau_n - A(\tau_n, \boldsymbol{\tau}_{-n})\big)\frac{\partial A}{\partial \tau_n} + \frac{\partial P_n}{\partial \tau_n} \equiv 0\,. \tag{11.10}$$

Observe that (11.10) is an identity, which means that we can consider P_n defined by that differential equation. To get a closed-form solution for P_n requires solving the differential equation. As written, this would seem a difficult differential equation because of the first τ_n in (11.10); that is, for an arbitrary function $A(\cdot)$, we wouldn't know how to integrate $\tau_n \times \partial A/\partial \tau_n$.[2] To simplify this differential equation, observe that we can use the social planner's first-order condition, expression (11.7), to substitute out $\big(\tau_n - A(\tau_n, \boldsymbol{\tau}_{-n})\big)$. This yields:

$$\frac{\partial P_n}{\partial \tau_n} = \frac{\partial A}{\partial \tau_n}\sum_{j \neq n}\big(\tau_j - A(\tau_n, \boldsymbol{\tau}_{-n})\big)\,. \tag{11.11}$$

The differential equation (11.11) has the solution:

$$P_n(\boldsymbol{\tau}) = \sum_{j \neq n}\left(\tau_j A(\boldsymbol{\tau}) - \frac{A(\boldsymbol{\tau})^2}{2}\right) + h_n(\boldsymbol{\tau}_{-n})\,, \tag{11.12}$$

where $h_n(\boldsymbol{\tau}_{-n})$ is a constant of integration with respect to τ_n (*i.e.,* it can depend on the *other* agents' announcements, but not n's). Observe, from (11.12), that agent n's payment equals the sum of all the *other* agents' utilities (gross of monetary transfers). Looking back at (11.9), this means that *each* agent faces the same maximization program that the social planner would face if she knew all the agents' types. From above, we know this program is globally concave; hence, our use of (11.10) instead of (11.9) was valid. Moreover, having the

[2] In the specific example under consideration, we can, of course, do it because $\partial A/\partial \tau_n$ is a constant.

payment to each agent equal, gross of any constants of integration, the sum of the utilities of the *other* agents is a general feature of this class of mechanisms. This generality has earned this class of mechanisms a name: *Groves-Clarke mechanisms* (sometimes just Groves mechanisms).

More generally, we can prove the following result:

Proposition 17 (Groves-Clarke Mechanism) *Consider a public-goods problem with N agents. Let the preferences of agent n be described by the quasi-linear (additively separable) utility function $v_n(a|\tau_n) + p_n$, where a is the level of the public good ($a \in \mathbb{R}_+$), p_n is agent n's allocation (payment) of the transferable good (money), and $\tau_n \in \mathcal{T}_n$ is agent n's type. Assume, for all $\boldsymbol{\tau} \in \mathcal{T}$, there exists a finite value $a^*(\boldsymbol{\tau})$ that solves*

$$\max_{a \in \mathbb{R}_+} \sum_{n=1}^{N} v_n(a|\tau_n)\,. \tag{11.13}$$

Then this $a^*(\cdot)$ can be implemented by the following direct-revelation dominant-strategy mechanism:

$$A(\boldsymbol{\tau}) = a^*(\boldsymbol{\tau}) \text{ and } P_n(\boldsymbol{\tau}) = \sum_{j \neq n} v_j\big(A(\boldsymbol{\tau})|\tau_j\big) + h_n(\boldsymbol{\tau}_{-n})\,,$$

where the $h_n(\cdot)$ are arbitrary functions from $\mathcal{T}_{-n} \to \mathbb{R}$.

Proof: We need to establish that the mechanism implements $a^*(\cdot)$ via truth-telling as a dominant strategy. Because we've limited attention to dominant strategies, it must be that each agent n, regardless of the realized $\tau_n \in \mathcal{T}_n$, wants to tell the truth for any $\boldsymbol{\tau}_{-n} \in \mathcal{T}_{-n}$. Hence, each agent n must find that truth telling, announcing τ_n, yields at least as much utility as announcing $\hat{\tau}_n \neq \tau_n$, for any $\hat{\tau}_n \in \mathcal{T}_n$ and all $\boldsymbol{\tau}_{-n} \in \mathcal{T}_{-n}$. We'll establish this by contradiction. Suppose there were an n such that, for a realization $\tau_n \in \mathcal{T}_n$ and $\boldsymbol{\tau}_{-n} \in \mathcal{T}_{-n}$, there exists a lie $\hat{\tau}_n \in \mathcal{T}_n$ such that

$$v_n\big(A(\hat{\tau}_n, \boldsymbol{\tau}_{-n})|\tau_n\big) + P_n(\hat{\tau}_n, \boldsymbol{\tau}_{-n}) > v_n\big(A(\tau_n, \boldsymbol{\tau}_{-n})|\tau_n\big) + P_n(\tau_n, \boldsymbol{\tau}_{-n})\,.$$

Substituting for P_n, this entails

$$\sum_{j=1}^{N} v_n\big(A(\hat{\tau}_n, \boldsymbol{\tau}_{-n})|\tau_j\big) > \sum_{j=1}^{N} v_n\big(A(\tau_n, \boldsymbol{\tau}_{-n})|\tau_j\big)\,. \tag{11.14}$$

Expression (11.14) can hold only if $A(\hat{\tau}_n, \boldsymbol{\tau}_{-n}) \neq A(\tau_n, \boldsymbol{\tau}_{-n})$. Recall the latter term is $a^*(\tau_n, \boldsymbol{\tau}_{-n})$, while the former term is some other a; but, then, (11.14) yields

$$\sum_{j=1}^{N} v_n(a|\tau_j) > \sum_{j=1}^{N} v_n\big(a^*(\tau_n, \boldsymbol{\tau}_{-n})|\tau_j\big)\,. \tag{11.15}$$

Expression (11.15), however, contradicts that $a^*(\tau_n, \boldsymbol{\tau}_{-n})$ maximizes (11.13). So, by contradiction, we've established that truth-telling is a dominant strategy under this mechanism. ■

Budget Balancing 11.6

As seen, public-goods problems are readily solved by Groves-Clarke mechanisms. There is, however, a fly in the ointment. We might wish for the mechanism to exhibit *budget balancing*; specifically, we might wish for

$$\sum_{n=1}^{N} P_n(\boldsymbol{\tau}) = 0 \tag{11.16}$$

for *all* $\boldsymbol{\tau} \in \mathcal{T}$. The reason for such a wish is that we might worry about a mechanism that either allowed the social planner to pocket money (*i.e.,* one in which $\sum_{n=1}^{N} P_n(\boldsymbol{\tau}) < 0$ for some $\boldsymbol{\tau}$) or in which she had to contribute money (*i.e.,* one in which $\sum_{n=1}^{N} P_n(\boldsymbol{\tau}) > 0$ for some $\boldsymbol{\tau}$). In the first case, what does it mean for the social planner (government) to pocket money? Presumably, any extra funds will be used by the government for some purpose, which would, then, affect the agents' utilities and, thus, their incentives in the mechanism. Alternatively, the planner could simply burn any extra funds; but, in many contexts, it is difficult to imagine that the planner's commitment to burn funds is credible—the society in question would, presumably, insist on renegotiating and making use of the extra funds (including rebating them to the agents). But this, too, would distort incentives. Having the planner contribute money (*i.e.,* $\sum P_n > 0$) can also be unrealistic—from where does the social planner get these extra funds? If she gets them by reallocating funds from other government projects, that will affect citizens' utilities and, thus, incentives. If she levies a tax to generate the funds, then that will clearly affect incentives. Bottom line: there are contexts in which attention needs to be restricted to balanced-budget mechanisms.

Unfortunately, Groves-Clarke mechanisms are generically *un*balanced; that is, (11.16) cannot be made to hold for *all* realizations of $\boldsymbol{\tau}$. This can be established fairly generally—see Laffont and Maskin (1980)—but here it will suffice to consider an example. Recall the public-goods example with which we began the last section. Maintain all the assumptions, but fix $N = 2$. From the analysis in the previous section, the Groves-Clarke mechanism is

$$A(\boldsymbol{\tau}) = \frac{\tau_1 + \tau_2}{2} \text{ and } P_n(\boldsymbol{\tau}) = \tau_{-n}A(\boldsymbol{\tau}) - \frac{A(\boldsymbol{\tau})^2}{2} + h_n(\tau_{-n})$$
$$= \frac{3\tau_{-n}^2 + 2\tau_n\tau_{-n} - \tau_n^2}{8} + h_n(\tau_{-n})\,.$$

Adding P_1 and P_2 yields

$$P_1(\boldsymbol{\tau}) + P_2(\boldsymbol{\tau}) = \frac{\tau_1^2 + \tau_1\tau_2 + \tau_2^2}{4} + h_1(\tau_2) + h_2(\tau_1)\,.$$

In general, this sum is not zero; that is, the mechanism is unbalanced. To verify this—that is, to show $P_1(\boldsymbol{\tau}) + P_2(\boldsymbol{\tau}) \neq 0$ for some $\boldsymbol{\tau} \in \mathcal{T}$—suppose that it were balanced. Then $P_1(\boldsymbol{\tau}) + P_2(\boldsymbol{\tau}) \equiv 0$. Because this is an identity, the derivative of this sum with respect to τ_1 is always zero; that is,

$$\frac{2\tau_1 + \tau_2}{4} + h_2'(\tau_1) \equiv 0\,. \tag{11.17}$$

But (11.17) entails that $h_2(\cdot)$ is a function of τ_2, which it cannot be. So, by contradiction, it follows that $P_1(\boldsymbol{\tau}) + P_2(\boldsymbol{\tau}) \neq 0$ for some $\boldsymbol{\tau} \in \mathcal{T}$.

Bayesian Mechanisms | 11.7

In part because Groves-Clarke mechanisms are generically unbalanced, economists have sought to devise other mechanisms that are balanced. To do so, however, means we must use a less restrictive solution concept than solution in dominant strategies. Here we use the solution concept Bayesian Nash—the agents must play mutual best responses given their expectations of others' types and strategies.

To begin, recall our now familiar example in which

$$u_n(a, p_n|\tau_n) = \tau_n a - \frac{a^2}{2} + p_n \text{ and } \mathcal{W}(a) = \sum_{n=1}^{N}\left(\tau_n a - \frac{a^2}{2}\right).$$

Recall that $a^*(\boldsymbol{\tau}) = \frac{1}{N}\sum_{n=1}^{N}\tau_n$. Assume that $\mathcal{T}_n = (0, T)$ where $0 < T \leq \infty$. Assume, critically, that the τ_n are independently distributed on $(0, T)$. For convenience, assume that they have identical distributions as well. Consistent with our earlier analysis, the *realization* of τ_n is agent n's private information, while the *distribution* of τ_n is common knowledge.

Consider the following direct-revelation mechanism:[3] $A(\boldsymbol{\tau}) = a^*(\boldsymbol{\tau})$ and the payment (transfer) vector $\mathbf{P}(\boldsymbol{\tau}) \equiv \big(P_1(\boldsymbol{\tau}), \dots, P_N(\boldsymbol{\tau})\big)$, where the latter is defined by

$$P_n(\boldsymbol{\tau}) = \rho_n(\tau_n) - \frac{1}{N-1}\sum_{j\neq n}\rho_j(\tau_j)\,. \tag{11.18}$$

Summing the P_ns reveals that they sum to zero for all $\boldsymbol{\tau}$; that is, the mechanism is balanced by construction. To find the functions $\rho_n(\cdot)$, we use the fact that they must be incentive compatible with truth telling in a Bayesian equilibrium; that is,

$$\tau_n \in \underset{\hat{\tau}_n \in (0,T)}{\operatorname{argmax}}\, \mathbb{E}_{\boldsymbol{\tau}_{-n}}\left\{\tau_n A(\hat{\tau}_n, \boldsymbol{\tau}_{-n}) - \frac{A(\hat{\tau}_n, \boldsymbol{\tau}_{-n})^2}{2} + \rho_n(\hat{\tau}_n) - \frac{1}{N-1}\sum_{j\neq n}\rho_j(\tau_j)\right\}.$$

Observe that it is expected utility because each agent is playing a best response to the presumed truth telling of the other agents. In other words, we have a

[3] From Proposition 16 there is no loss of generality in restricting attention to direct-revelation mechanisms.

Bayesian Nash equilibrium if truth telling by each agent n is a best response to truth telling by the other agents.

Let's conjecture that the above maximization program is globally concave in $\hat{\tau}_n$, so that we can replace it with its corresponding first-order condition (we will verify global concavity at the end of this derivation):

$$\mathbb{E}_{\boldsymbol{\tau}_{-n}}\left\{\left(\tau_n - A(\hat{\tau}_n, \boldsymbol{\tau}_{-n})\right)\frac{\partial A}{\partial \hat{\tau}_n} + \rho_n'(\tau_n)\right\} = 0\,. \tag{11.19}$$

The function $\rho_n'(\tau_n)$ is independent of $\boldsymbol{\tau}_{-n}$ and, hence, can be passed through the expectations operator. In addition, from the social planner's first-order condition (11.7) on page 107,

$$\mathbb{E}_{\boldsymbol{\tau}_{-n}}\left\{\left(\tau_n - A(\tau_n, \boldsymbol{\tau}_{-n})\right)\frac{\partial A}{\partial \tau_n}\right\} = -\mathbb{E}_{\boldsymbol{\tau}_{-n}}\left\{\sum_{j\neq n}\left(\tau_j - A(\tau_n, \boldsymbol{\tau}_{-n})\right)\frac{\partial A}{\partial \tau_n}\right\}\,.$$

Using these two insights, we can rearrange (11.19) to become

$$\rho_n'(\tau_n) = \mathbb{E}_{\boldsymbol{\tau}_{-n}}\left\{\sum_{j\neq n}\left(\tau_j - A(\tau_n, \boldsymbol{\tau}_{-n})\right)\frac{\partial A}{\partial \tau_n}\right\}\,.$$

Solving this differential equation (*i.e.,* integrating through the expectations operator and reversing the chain rule) yields:

$$\rho_n(\tau_n) = \mathbb{E}_{\boldsymbol{\tau}_{-n}}\left\{\sum_{j\neq n}\left(\tau_j A(\boldsymbol{\tau}) - \frac{1}{2}A(\boldsymbol{\tau})^2\right)\right\} + h_n\,, \tag{11.20}$$

where h_n is an arbitrary constant of integration that does *not* depend on the agents' announcements. Observe, ignoring h_n, that $\rho_n(\tau_n)$ equals the *expectation* of the sum of the *other* agents' utilities (gross of transfers). This means that, in expectation, *each* agent faces the same maximization problem as the social planner would if she knew the agents' types. Because we know this problem is globally concave, we've, thus, verified that we were justified in replacing agent n's incentive compatibility constraint with the corresponding first-order condition. We also obtain a nice interpretation: Because the transfer function puts each agent on the socially desired margin with respect to his own maximization problem, each agent is induced to act in the public good.

We can generalize this conclusion:

Proposition 18 (Bayesian Mechanism) *Consider a public-goods problem with N agents. Let the preferences of agent n be described by the quasi-linear (additively separable) utility function $v_n(a|\tau_n) + p_n$, where a is the level of the public good ($a \in \mathbb{R}_+$), p_n is agent n's allocation (payment) of the transferable good (money), and $\tau_n \in \mathcal{T}_n$ is agent n's type. Assume that the types are distributed independently of each other (i.e., knowledge of τ_n yields agent n no* *additional information about* τ_j, $j \neq n$, *than he had prior to the realization of* τ_n*). Assume, for all* $\boldsymbol{\tau} \in \mathcal{T}$, *there exists a finite value* $a^*(\boldsymbol{\tau})$ *that solves*

$$\max_{a \in \mathbb{R}_+} \sum_{n=1}^{N} v_n(a|\tau_n) \,. \tag{11.21}$$

Then this $a^*(\cdot)$ *can be implemented by the following direct-revelation Bayesian mechanism:*

$$A(\boldsymbol{\tau}) = a^*(\boldsymbol{\tau}) \text{ and } P_n(\boldsymbol{\tau}) = \rho_n(\tau_n) - \frac{1}{N-1}\sum_{j \neq n} \rho_j(\tau_j) \,,$$

where

$$\rho_n(\tau_n) = \mathbb{E}_{\boldsymbol{\tau}_{-n}}\left\{\sum_{j \neq n} v_j\big(A(\boldsymbol{\tau})|\tau_j\big)\right\} + h_n \,,$$

where h_n *is an arbitrary constant.*

Proof: We need to establish that the mechanism implements $a^*(\cdot)$ via truth-telling as a Bayesian Nash equilibrium. This means each agent n must find that truth telling, announcing τ_n, yields at least as much utility as announcing $\hat{\tau}_n \neq \tau_n$, for any $\hat{\tau}_n \in \mathcal{T}_n$ given that he believes the other agents will tell the truth. We'll establish this by contradiction. Suppose there were an n such that, for a realization $\tau_n \in \mathcal{T}_n$, there exists a lie $\hat{\tau}_n \in \mathcal{T}_n$ such that

$$\mathbb{E}_{\boldsymbol{\tau}_{-n}}\big\{v_n\big(A(\hat{\tau}_n, \boldsymbol{\tau}_{-n})|\tau_n\big) + P_n(\hat{\tau}_n, \boldsymbol{\tau}_{-n})\big\} > \mathbb{E}_{\boldsymbol{\tau}_{-n}}\big\{v_n\big(A(\tau_n, \boldsymbol{\tau}_{-n})|\tau_n\big) + P_n(\tau_n, \boldsymbol{\tau}_{-n})\big\} \,.$$

Substituting for P_n and canceling like terms, this entails[4]

$$\mathbb{E}_{\boldsymbol{\tau}_{-n}}\left\{\sum_{j=1}^{N} v_n\big(A(\hat{\tau}_n, \boldsymbol{\tau}_{-n})|\tau_j\big)\right\} > \mathbb{E}_{\boldsymbol{\tau}_{-n}}\left\{\sum_{j=1}^{N} v_n\big(A(\tau_n, \boldsymbol{\tau}_{-n})|\tau_j\big)\right\} \,. \tag{11.22}$$

Expression (11.22) can hold only if $A(\hat{\tau}_n, \boldsymbol{\tau}_{-n}) \neq A(\tau_n, \boldsymbol{\tau}_{-n})$ for at least one realization of $(\hat{\tau}_n, \boldsymbol{\tau}_{-n})$ and $\boldsymbol{\tau}$. Recall the latter term is $a^*(\tau_n, \boldsymbol{\tau}_{-n})$, while the former term is some other a; but, then, (11.22) implies

$$\sum_{j=1}^{N} v_n(a|\tau_j) > \sum_{j=1}^{N} v_n\big(a^*(\tau_n, \boldsymbol{\tau}_{-n})|\tau_j\big) \tag{11.23}$$

for at least one realization of $\boldsymbol{\tau}$. Expression (11.23), however, contradicts that $a^*(\tau_n, \boldsymbol{\tau}_{-n})$ maximizes (11.21) for all $\boldsymbol{\tau}$. So, by contradiction, we've established that truth-telling is a best response to truth-telling by the other agents; that is, the Bayesian Nash equilibrium of this mechanism implements $a^*(\boldsymbol{\tau})$ for all $\boldsymbol{\tau} \in \mathcal{T}$. ∎

[4] Recall $\mathbb{E}\big\{\mathbb{E}\{X\}\big\} = \mathbb{E}\{X\}$.

Bibliographic Note

Bayesian mechanisms are due to d'Aspremont and Gérard-Varet (1979).

Hidden Action and Incentives

Purpose

A common economic occurrence is the following: Two parties, *principal* and *agent*, are in a situation—typically of their choosing—in which actions by the agent impose an externality on the principal. Not surprisingly, the principal will want to influence the agent's actions. This influence will often take the form of a contract that has the principal compensating the agent contingent on either his actions or the consequences of his actions. Table 2 lists some examples of situations like this. Note that, in many of these examples, the principal is buying a good or service from the agent. That is, many buyer-seller relationships naturally fit into the principal-agent framework. This part of the notes covers the basic tools and results of agency theory.

Table 2: Examples of Moral-Hazard Problems

Principal	Agent	Problem	Solution
Employer	Employee	Induce employee to take actions that increase employer's profits, but which he finds personally costly.	Base employee's compensation on employer's profits.
Plaintiff	Attorney	Induce attorney to expend costly effort to increase plaintiff's chances of prevailing at trial.	Make attorney's fee contingent on damages awarded plaintiff.
Homeowner	Contractor	Induce contractor to complete work (e.g., remodel kitchen) on time.	Give contractor bonus for completing job on time.
Landlord	Tenant	Induce tenant to make investments (e.g., in time or money) that preserve or enhance property's value to the landlord.	Pay the tenant a fraction of the increased value (e.g., share-cropping contract). Alternatively, make tenant post deposit to be forfeited if value declines too much.

To an extent, the principal-agent problem finds its root in the early literature on insurance. There, the concern was that someone who insures an asset might then fail to maintain the asset properly (*e.g.,* park his car in a bad neighborhood). Typically, such behavior was either unobservable by the insurance company or too difficult to contract against directly; hence, the insurance contract could not be directly contingent on such behavior. But because this behavior—known as *moral hazard*—imposes an externality on the insurance company (in this case, a negative one), insurance companies were eager to develop contracts that guarded against it. So, for example, many insurance contracts have *deductibles*—the first k dollars of damage must be paid by the insured rather than the insurance company. Because the insured now has $\$k$ at risk, he'll think twice about parking in a bad neighborhood. That is, the insurance contract is designed to mitigate the externality that the agent—the insured—imposes on the principal—the insurance company. Although principal-agent analysis is more general than this, the name "moral hazard" has stuck and, so, the types of problems considered here are often referred to as moral-hazard problems. A more descriptive name, which is also used in the literature, is *hidden-action problems*.

Bibliographic Note

This part of the lecture notes draws heavily from a set of notes that I co-authored with Bernard Caillaud.

12 The Moral-Hazard Setting

We begin with a general picture of the situation we wish to analyze.

1. Two players are in an economic relationship characterized by the following two features: First, the actions of one player, *the agent*, affect the well-being of the other player, *the principal*. Second, the players can agree *ex ante* to a reward schedule by which the principal pays the agent.[1] The reward schedule represents an *enforceable* contract (*i.e.,* if there is a dispute about whether a player has lived up to the terms of the contract, then a court or similar body can adjudicate the dispute).

2. The agent's action is *hidden*; that is, he knows what action he has taken but the principal does not directly observe his action. (Although we will consider, as a benchmark, the situation in which the action can be contracted on directly.) Moreover, the agent has complete discretion in choosing his action from some set of feasible actions.[2]

3. The actions determine, usually stochastically, some *performance measures*. In many models, these are identical to the benefits received by the principal, although in some contexts the two are distinct. The reward schedule is a function of (at least some) of these performance variables. In particular, the reward schedule can be a function of the *verifiable* performance measures.[3]

4. The *structure* of the situation is common knowledge between the players.

For example, consider a salesperson who has discretion over the amount of time or effort he expends promoting his company's products. Many of these actions are unobservable by his company. The company can, however, measure in a verifiable way the number of orders or revenue he generates. Because these measures are, presumably, correlated with his actions (*i.e.,* the harder he works, the more sales he generates *on average*), it may make sense for the company to

[1] Although we typically think in terms *positive* payments, in many applications payments could be *negative*; that is, the principal fines or otherwise punishes the agent.

[2] Typically, this set is assumed to be exogenous to the relationship. One could, however, imagine situations in which the principal has some control over this set *ex ante* (*e.g.,* she decides what tools the agent will have available).

[3] Information is *verifiable* if it can be observed perfectly (*i.e.,* without error) by third parties who might be called upon to adjudicate a dispute between principal and agent.

base his pay on his sales—put him on commission—to induce him to expend the appropriate level of effort.

Here, we will also be imposing some additional structure on the situation:

- The players are symmetrically informed at the time they agree to a reward schedule.
- Bargaining is take-it-or-leave-it (TIOLI): The principal proposes a contract (reward schedule), which the agent either accepts or rejects. If he rejects it, the game ends and the players receive their *reservation utilities* (their expected utilities from pursuing their next best alternatives). If he accepts, then both parties are bound by the contract.
- Contracts cannot be renegotiated.
- Once the contract has been agreed to, the only player to take further actions is the agent.
- The game is played once. In particular, there is only one period in which the agent takes actions and the agent completes his actions before any performance measures are realized.

All of these are common assumptions and, indeed, might be taken to constitute part of the "standard" principal-agent model.

The link between actions and performance can be seen as follows. Performance is a random variable and its probability distribution depends on the actions taken by the agent. So, for instance, a salesperson's efforts could increase his average (expected) sales, but he still faces upside risk (*e.g.,* an economic boom in his sales region) and downside risk (*e.g.,* introduction of a rival product). Because the performance measure is only stochastically related to the action, it is generally impossible to infer perfectly the action from the realization of the performance measure. That is, the performance measure does not, generally, reveal the agent's action—it remains "hidden" despite observing the performance measure.

The link between actions and performance can also be viewed in a indirect way in terms of a *state-space model*. Performance is a function of the agent's actions and of the *state of nature*; that is, a parameter (scalar or vector) that describe the economic environment (*e.g.,* the economic conditions in the salesperson's territory). In this view, the agent takes his action before knowing the state of nature. Typically, we assume that the state of nature is not observable to the principal. If she could observe it, then she could perfectly infer the agent's action by inverting from realized performance. In this model, it is not important whether the agent later observes the state of nature or not, given he could deduce it from his observation of his performance and his knowledge of his actions.

There is a strong assumption of physical causality in this setting, namely that actions by the agent determine performances. Moreover, the process is viewed as a static production process: There are neither dynamics nor feedback. In particular, the contract governs one period of production and the game between principal and agent encompasses only this period. In addition, when choosing his actions, the agent's information is identical to the principal's. Specifically, he can*not* adjust his actions as the performance measures are realized. The sequentiality between actions and performance is strict: First actions are completed and, only then, is performance realized.

Basic Two-Action Model | 13

We start with the simplest principal-agent model. Admittedly, it is so simple that a number of the issues one would like to understand about contracting under moral hazard disappear. On the other hand, many issues remain and, for pedagogical purposes at least, it is a good place to start.[1]

The Two-action Model | 13.1

Consider a salesperson, who will be the agent in this model and who works for a manufacturer, the principal. The manufacturer's problem is to design incentives for the salesperson to expend effort promoting the manufacturer's product to consumers.

Let x denote the level of sales that the salesperson reaches within the period under consideration. This level of sales depends upon many demand parameters that are beyond the salesperson's control; but, critically, they also depend upon the salesperson's effort—the more effort the salesperson expends, the more consumers will buy in expectation. Specifically, suppose that when the salesperson does not expend effort, sales are distributed according to distribution function $F_0(\cdot)$ on $\mathbb{R}_+$.[2] When he does expend effort, sales are distributed $F_1(\cdot)$. Observe effort, here, is a binary choice. Consistent with the story we've told so far, we want sales to be greater, in expectation, if the salesperson has expended effort; that is, assume[3]

$$\mathbb{E}_1\{x\} \equiv \int_0^{\infty} x dF_1(x) > \int_0^{+\infty} x dF_0(x)dx \equiv \mathbb{E}_0\{x\}\,. \qquad (13.1)$$

Having the salesperson expend effort is sales-enhancing, but it is also costly for the salesperson. Expending effort causes him disutility C compared to no

[1] But the pedagogical value of this section should not lead us to forget caution. And caution is indeed necessary as the model oversimplifies reality to a point that it delivers conclusions that have no match in a more general framework. One could say that the two-action model is tailored so as to fit with naïve intuition and to lead to the desired results without allowing us to see fully the (implicit) assumptions on which we are relying.

[2] Treating the set of "possible" sales as $[0, \infty)$ is without loss of generality, because the boundedness of sales can be captured by assuming $F(x \geq \bar{x}) = 0$ for some $\bar{x} < \infty$.

[3] Observe that this notation covers both the case in which $F_a(\cdot)$ is a differentiable distribution function or a discrete distribution function.

effort. The salesperson has discretion: He can incur a personal cost of C and boosts sales by choosing action $a = 1$ (expending effort), or he can expend no effort, action $a = 0$, which causes him no disutility but does not stimulate demand either. Like most individuals, the salesperson is sensitive to variations in his income; specifically, his preferences over income exhibit risk aversion. Assume, too, that his utility exhibits additive separability in money and action. Specifically, let his utility be

$$U(s,x,a) = u(s) - aC,$$

where s is a payment from the manufacturer and $u(\cdot)$ is strictly increasing and concave (the salesperson—agent—prefers more money to less and is risk averse). Of course, the salesperson could also simply choose not to work for the manufacturer. This would yield him an expected level of utility equal to U_R. The quantity U_R is the salesperson's *reservation utility*.

The manufacturer is a large risk-neutral company that cares about the sales realized on the local market net of the salesperson's remuneration or share of the sales. Hence, the manufacturer's preferences are captured by the utility (profit) function:

$$W(s,x,a) = x - s.$$

Assume the manufacturer's size yields it all the bargaining power in its negotiations with the salesperson.

Suppose, as a benchmark, that the manufacturer could observe and establish whether the salesperson had expended effort. We will refer to this benchmark as the *full or perfect information case*. Then the manufacturer could use a contract that is contingent on the salesperson's effort, a. Moreover, because the salesperson is risk averse, while the manufacturer is risk neutral, it is most efficient for the manufacturer to absorb all risk. Hence, in this benchmark case, the salesperson's compensation would *not* depend on the realization of sales, but only on the salesperson's effort. The contract would then be of the form:

$$s = \begin{cases} s_0 \text{ if } a = 0 \\ s_1 \text{ if } a = 1 \end{cases}.$$

If the manufacturer wants the salesperson to expend effort (*i.e.,* choose $a = 1$), then it must choose s_0 and s_1 to satisfy two conditions. First, conditional on accepting the contract, the salesperson must prefer to invest; that is,

$$u(s_1) - C \geq u(s_0)\ . \tag{IC}$$

A constraint like this is known as an *incentive compatibility* constraint (conventionally abbreviated IC): Taking the desired action must maximize the agent's expected utility. Second, conditional on the fact that he will be induced to expend effort, he must prefer to sign the contract than to forgo employment with the manufacturer; that is,

$$u(s_1) - C \geq U_R\,. \tag{IR}$$

A constraint like this is known as an *individual rationality* constraint (conventionally abbreviated IR). The IR constraint is also referred to as a *participation constraint*. Moreover, in selecting s_0 and s_1, the manufacturer wants to maximize its profits conditional on gaining acceptance of the contract and inducing $a = 1$. That is, it wishes to solve

$$\max_{s_0, s_1} \mathbb{E}_1\{x\} - s_1$$

subject to the constraints (IC) and (IR). If we postulate that

$$\begin{aligned} u(s_0) &< U_R \text{ and} \\ u(s_1) - C &= U_R \end{aligned} \tag{13.2}$$

both have solutions within the domain of $u(\cdot)$, then the solution to the manufacturer's problem is straightforward: s_1 solves (13.2) and s_0 is a solution to $u(s) < U_R$. It is readily seen that this solution satisfies the constraints. Moreover, because $u(\cdot)$ is strictly increasing, there is no smaller payment that the manufacturer could give the salesperson and still have him accept the contract. This contract is known as a *forcing contract*.[4] For future reference, let s_1^F be the solution to (13.2). Observe that $s_1^F = u^{-1}(U_R + C)$, where $u^{-1}(\cdot)$ is the inverse of the function $u(\cdot)$.

Another option for the manufacturer is, of course, not to bother inducing the salesperson to expend effort promoting the product. There are many contracts that would accomplish this goal, although the most "natural" is perhaps a *non-contingent* contract: $s_0 = s_1$. Given that the manufacturer doesn't seek to induce investment, there is no IC constraint—the salesperson inherently prefers not to invest—and the only constraint is the IR constraint:

$$u(s_0) \geq U_R.$$

The expected-profit-maximizing (cost-minimizing) payment is then the smallest payment satisfying this expression. Given that $u(\cdot)$ is increasing, this entails $s_0 = u^{-1}(U_R)$. We will refer to *this* value of s_0 as s_0^F.

[4] The solution to the manufacturer's maximization problem depends on the domain and range of the utility function $u(\cdot)$. Let $\mathcal{D}$, an interval in $\mathbb{R}$, be the domain and $\mathcal{R}$ be the range. Let $\underline{s}$ be $\inf \mathcal{D}$ (*i.e.,* the greatest lower bound of $\mathcal{D}$) and let $\bar{s}$ be $\sup \mathcal{D}$ (*i.e.,* the least upper bound of $\mathcal{D}$). As shorthand for $\lim_{s \downarrow \underline{s}} u(s)$ and $\lim_{s \uparrow \bar{s}} u(s)$, we'll write $u(\underline{s})$ and $u(\bar{s})$ respectively. If $u(s) - C < U_R$ for all $s \leq \bar{s}$, then *no* contract exists that satisfies (IR). In this case, the best the manufacturer could hope to do is implement $a = 0$. Similarly, if $u(\bar{s}) - C < u(\underline{s})$, then *no* contract exists that satisfies (IC). The manufacturer would have to be satisfied with implementing $a = 0$. Hence, $a = 1$ can be implemented if and only if $u(\bar{s}) - C > \max\{U_R, u(\underline{s})\}$. Assuming this condition is met, a solution is $s_0 \downarrow \underline{s}$ and s_1 solving

$$u(s) - C \geq \max\{U_R, u(\underline{s})\}.$$

Generally, conditions are imposed on $u(\cdot)$ such that a solution exists to $u(s) < U_R$ and (13.2). Henceforth, we will assume that these conditions have, indeed, been imposed. For an example of an analysis that considers bounds on $\mathcal{D}$ that are more binding, see Sappington (1983).

The manufacturer's expected profit conditional on inducing a under the *optimal* contract for inducing a is $\mathbb{E}_a[x] - s_a^F$. The manufacturer will, thus, prefer to induce $a = 1$ if

$$\mathbb{E}_1[x] - s_1^F > \mathbb{E}_0[x] - s_0^F .$$

In what follows, we will assume that this condition is met: That is, in our benchmark case of verifiable action, the manufacturer prefers to induce effort than not to.

Observe the steps taken in solving this benchmark case: First, for each possible action we solved for the optimal contract that induces that action. Then we calculated the expected profits for each possible action assuming the optimal contract. The action that is induced is, then, the one that yields the largest expected profit. This two-step process for solving for the optimal contract is frequently used in contract theory, as we will see.

13.2 The Optimal Incentive Contract

Now, we return to the case of interest: The salesperson's (agent's) action is hidden. Consequently, the manufacturer cannot make its payment contingent on whether the salesperson expends effort. The only verifiable variable is performance, as reflected by realized sales, x. A contract, then, is a function mapping sales into compensation for the salesperson: $s = S(x)$. Facing such a contract, the salesperson then freely chooses the action that maximizes his expected utility. Consequently, the salesperson chooses action $a = 1$ if and only if:[5]

$$\mathbb{E}_1[u(S(x))] - C \geq \mathbb{E}_0[u(S(x))] . \qquad (\text{IC}')$$

If this inequality is violated, the salesperson will simply not expend effort. Observe that this is the incentive compatibility (IC) constraint in this case.

The game we analyze is in fact a simple Stackelberg game, where the manufacturer is the first mover—it chooses the payment schedule—to which it is committed; and the salesperson is the second mover—choosing his action in response to the payment schedule. The solution to

$$\max_a \mathbb{E}_a[u(S(x))] - aC$$

[5] We assume, when indifferent among a group of actions, that the agent chooses from that group the action that the principal prefers. This assumption, although often troubling to those new to agency theory, is not truly a problem. Recall that the agency problem is a game. Consistent with game theory, we're looking for an equilibrium of this game; *i.e.,* a situation in which players are playing mutual best responses and in which they correctly anticipate the best responses of their opponents. Were the agent to behave differently when indifferent, then we wouldn't have an equilibrium because the principal would vary her strategy—offer a different contract—so as to break this indifference. Moreover, it can be shown that in many models the *only* equilibrium has the property that the agent chooses among his best responses (the actions among which he is indifferent given the contract) the one most preferred by the principal.

(with ties going to the manufacturer) gives the salesperson's equilibrium choice of action by the agent as a *function* of the payment function $S(\cdot)$. Solving this contracting problem then requires us to understand what kind of contract the manufacturer could and will offer.

Observe first that if she were to offer the fixed-payment contract $S(x) = s_0^F$ for all x, then, as above, the agent would accept the contract and not bother to expend effort. Among all contracts that induce the agent to choose action $a = 0$ in equilibrium, this is clearly the cheapest one for the manufacturer. The fixed-payment contract set at s_1^F will, however, no longer work given the hidden-action problem: Since the salesperson gains s_1^F whatever his efforts, he will choose the action that has lesser cost for him, $a = 0$. It is in fact immediate that any fixed-payment contract, which would be optimal if the only concern were efficient risk-sharing, will induce an agent to choose his least costly action. Given that it is desirable, at least in the benchmark full-information case, for the salesperson to expend effort selling the product, it seems plausible that the manufacturer will try to induce effort even though—as we've just seen—that must entail the *in*efficient (relative to the first best) allocation of risk to the salesperson.

We now face two separate questions. First, conditional on the manufacturer's wanting the salesperson to expend effort, what is the optimal—least-expected-cost—contract for the manufacturer to offer? Second, are the manufacturer's expected profits greater doing this than not inducing the salesperson to expend effort (*i.e.,* greater than the expected profits from offering the fixed-payment contract $S(x) = s_0^F$)?

As in the benchmark case, not only must the contract give the salesperson an incentive to acquire information (*i.e.,* meet the IC constraint), it must also be individually rational:

$$\mathbb{E}_1 \left[u(S(x))\right] - C \geq U_R. \qquad (\text{IR}')$$

The optimal contract is then the solution to the following program:

$$\max_{S(\cdot)} \mathbb{E}_1 \left[x - S(x)\right] \qquad (13.3)$$
$$\text{subject to } (\text{IC}') \text{ and } (\text{IR}').$$

The next few sections will consider the solution to (13.3) under a number of different assumptions about the distribution functions $F_a(\cdot)$.

Two assumptions on $u(\cdot)$, in addition to those already given, that will be common to these analyses are:

1. The domain of $u(\cdot)$ is $(\underline{s}, \infty)$, $\underline{s} \geq -\infty$.[6]

2. $\lim_{s \downarrow \underline{s}} u(s) = -\infty$ and $\lim_{s \uparrow \infty} u(s) = \infty$.

[6] Since its domain is an open interval and it is concave, $u(\cdot)$ is continuous everywhere on its domain (see van Tiel, 1984, p. 5).

An example of a function satisfying *all* the assumptions on $u(\cdot)$ is $\log(\cdot)$ with $\underline{s} = 0$. For the most part, these assumptions are for convenience and are more restrictive than we need (for instance, many of our results will also hold if $u(s) = \sqrt{x}$, although this fails the second assumption). Some consequences of these and our earlier assumptions are

- $u(\cdot)$ is invertible (a consequence of its strict monotonicity). Let $u^{-1}(\cdot)$ denote its inverse.
- The domain of $u^{-1}(\cdot)$ is $\mathbb{R}$ (a consequence of the last two assumptions).
- $u^{-1}(\cdot)$ is continuous, strictly increasing, and convex (a consequence of the continuity of $u(\cdot)$, its concavity, and that it is strictly increasing).

Two-outcome Model | 13.3

Imagine that there are only two possible realizations of sales, high and low, denoted respectively by x_H and x_L, with $x_H > x_L$. Assume that

$$F_a(x_L) = 1 - qa,$$

where $q \in (0,1)$is a known constant.

A contract is $s_H = S(x_H)$ and $s_L = S(x_L)$. We can, thus, write program (13.3) as

$$\max_{s_H, s_L} q(x_H - s_H) + (1-q)(x_L - s_L)$$

subject to

$$qu(s_H) + (1-q)u(s_L) - C \geq 0 \cdot u(s_H) + 1 \cdot u(s_L) \qquad \text{(IC)}$$

and

$$qu(s_H) + (1-q)u(s_L) - C \geq U_R \qquad \text{(IR)}$$

We could solve this problem mechanically using the usual techniques for maximizing a function subject to constraints, but it is far easier, here, to use a little intuition. To begin, we need to determine which constraints are binding. Is IC binding? Well, suppose it were not. Then the problem would simply be one of optimal risk sharing, because, by supposition, the incentive problem no longer binds. But we know optimal risk sharing entails $s_H = s_L$; that is, a fixed-payment contract.[7] As we saw above, however, a fixed-payment contract

[7] The proof is straightforward if $u(\cdot)$ is differentiable: Let λ be the Lagrange multiplier on (IR). The first-order conditions with respect to s_L and s_H are

$$1 - q - \lambda(1-q)u'(s_L) = 0$$

and

$$q - \lambda q u'(s_H) = 0,$$

respectively. Solving, it is clear that $s_L = s_H$. The proof when $u(\cdot)$ is not (everywhere) differentiable is only slightly harder and is left to the reader.

can*not* satisfy IC:

$$u(s) - C < u(s).$$

Hence, IC must be binding.

What about IR? Is it binding? Suppose it were not (*i.e.,* it were a strict inequality) and let s_L^* and s_H^* be the optimal contract. Then there must exist an $\varepsilon > 0$ such that

$$q - \big(u(s_H^*) - \varepsilon\big) + (1-q)\big(u(s_L^*) - \varepsilon\big) - C \geq U_R.$$

Let $\tilde{s}_n = u^{-1}\big(u(s_n^*) - \varepsilon\big)$, $n \in \{L, H\}$. Clearly, $\tilde{s}_n < s_n^*$ for both n, so that the $\{\tilde{s}_n\}$ contract costs the manufacturer less than the $\{s_n^*\}$ contract; or, equivalently, the $\{\tilde{s}_n\}$ contract yields the manufacturer greater expected profits than the $\{s_n^*\}$ contract. Moreover, the $\{\tilde{s}_n\}$ contract satisfies IC:

$$\begin{aligned} qu(\tilde{s}_H) + (1-q)u(\tilde{s}_L) - C &= q\big(u(s_H^*) - \varepsilon\big) + (1-q)\big(u(s_L^*) - \varepsilon\big) - C \\ &= qu(s_H^*) + (1-q)u(s_L^*) - C - \varepsilon \\ &\geq 0 \cdot u(s_H^*) + 1 \cdot u(s_L^*) - \varepsilon \quad (\{s_n^*\} \text{ satisfies IC}) \\ &= u(\tilde{s}_L). \end{aligned}$$

But this means that $\{\tilde{s}_n\}$ satisfies *both* constraints *and* yields greater expected profits, which contradicts the optimality of $\{s_n^*\}$. Therefore, by contradiction, we may conclude that IR is also binding under the optimal contract for inducing $a = 1$.

We're now in a situation where the two constraints must bind at the optimal contract. But, given we have only two unknown variables, s_H and s_L, this means we can solve for the optimal contract merely by solving the constraints. Doing so yields

$$\hat{s}_H = u^{-1}\left(U_R + \frac{1}{q}C\right) \text{ and } \hat{s}_L = u^{-1}(U_R). \tag{13.4}$$

Observe that the payments *vary* with the state (as we knew they must because fixed payments fail the IC constraint).

Recall that *were* the saleperson's action verifiable, the contract would be $S(x) = s_1^F = u^{-1}(U_R + C)$. Rewriting (13.4) we see that

$$\hat{s}_H = u^{-1}\left(u\left(s_1^F\right) + \frac{1-q}{q}C\right) \text{ and } \hat{s}_L = u^{-1}\big(u\left(s_1^F\right) - C\big);$$

that is, one payment is *above* the payment under full information, while the other is *below* the payment under full information. Moreover, the *expected* payment to the salesperson is greater than s_1^F:

$$\begin{aligned} q\hat{s}_H + (1-q)\hat{s}_L &= qu^{-1}\left(u\left(s_1^F\right) + \tfrac{1-q}{q}C\right) + (1-q)u^{-1}\big(u\left(s_1^F\right) - C\big) \\ &\geq u^{-1}\Big(q\left(u\left(s_1^F\right) + \tfrac{1-q}{q}C\right) + (1-q)\left(u\left(s_1^F\right) - C\right)\Big) \qquad (13.5) \\ &= u^{-1}\big(u\left(s_1^F\right)\big) = s_1^F; \end{aligned}$$

where the inequality follows from Jensen's inequality.[8] Provided the agent is strictly risk averse, the above inequality is strict: Inducing the agent to choose $a = 1$ costs strictly more in expectation when the principal cannot verify the agent's action.

Before proceeding, it is worth considering why the manufacturer (principal) suffers from its inability to verify the salesperson's (agent's) action (*i.e.,* from the existence of a hidden-action problem). *Ceteris paribus*, the salesperson prefers $a = 0$ to $a = 1$ because expending effort is personally costly to him. Hence, when the manufacturer wishes to induce $a = 1$, its interests and the salesperson's are not aligned. To align their interests, the manufacturer must offer the salesperson incentives to choose $a = 1$. The problem is that the manufacturer cannot directly tie these incentives to the variable in which it is interested, namely the action itself. Rather, it must tie these incentives to sales, which are imperfectly correlated with action. These incentives, therefore, expose the agent to risk. We know, relative to the first best, that this is inefficient. Someone must bear the cost of this inefficiency. Because the bargaining game always yields the salesperson the same expected utility (*i.e.,* IR is always binding), the cost of this inefficiency must, thus, be borne by the manufacturer.

Another way to view this last point is that because the agent is exposed to risk, which he dislikes, he must be compensated. This compensation takes the form of a higher *expected* payment.

To begin to appreciate the importance of the hidden-action problem, observe that

$$\begin{aligned}\lim_{q\uparrow 1} q\hat{s}_H + (1-q)\,\hat{s}_L &= \lim_{q\uparrow 1} \hat{s}_H \\ &= u^{-1}\big(u\left(s_1^F\right)\big) \\ &= s_1^F.\end{aligned}$$

Hence, when $q = 1$, there is effectively no hidden-action problem: Low sales, x_L, constitute proof that the salesperson failed to invest, because $\Pr\{x = x_L | a = 1\} = 0$ in that case. The manufacturer is, thus, free to punish the salesperson for low sales in whatever manner it sees fit; thereby deterring $a = 0$. But because there is no risk when $a = 1$, the manufacturer does not have to compensate the salesperson for bearing risk and can, thus, satisfy the IR constraint paying the same compensation as under full information. When $q = 1$, we have what is known as a *shifting support*.[9] We will consider shifting supports in greater

[8] Jensen's inequality for convex functions states that if $g(\cdot)$ is convex function, then $\mathbb{E}\{g(X)\} \geq g(\mathbb{E}X)$, where X is a random variable whose support is an interval of $\mathbb{R}$ and $\mathbb{E}$ is the expectations operator with respect to X (see, *e.g.,* van Tiel, 1984, p. 11, for a proof). If $g(\cdot)$ is strictly convex and the distribution of X is not degenerate (*i.e.,* does not concentrate all mass on one point), then the inequality is strict. For *concave* functions, the inequalities are reversed.

[9] The *support* of distribution G over random variable X, sometimes denoted $\text{supp}\{X\}$, is the set of x's such that for all $\varepsilon > 0$,

$$G(x) - G(x - \varepsilon) > 0.$$

depth later.

To see the importance of the salesperson's risk aversion, note that *were* the salesperson risk neutral, then the inequality in (13.5) would, instead, be an equality and the expected wage paid the salesperson would equal the wage paid under full information. Given that the manufacturer is risk *neutral* by assumption, it would be indifferent between an expected wage of s_1^F and paying s_1^F with certainty: There would be no loss, relative to full information, of overcoming the hidden-action problem by basing compensation on sales. It is important to note, however, that assuming a risk-neutral agent does *not* obviate the need to pay contingent compensation (*i.e.,* we still need $s_H > s_L$)—as can be seen by checking the IC constraint; agent risk neutrality only means that the principal suffers no loss from the fact that the agent's action is hidden.

We can also analyze this version of the model graphically. A graphical treatment is facilitated by switching from compensation space to utility space; that is, rather than put s_L and s_H on the axes, we put $u_L \equiv u(s_L)$ and $u_H \equiv u(s_H)$ on the axes. With this change of variables, program (13.3) becomes:

$$\max_{u_L, u_H} q\left(x_H - u^{-1}(u_H)\right) + (1-q)\left(x_L - u^{-1}(u_L)\right)$$

subject to

$$qu_H + (1-q)u_L - C \geq u_L \text{ and} \qquad (\text{IC}'')$$

$$qu_H + (1-q)u_L - C \geq U_R \qquad (\text{IR}'')$$

Observe that, in this space, the salesperson's indifference curves are straight lines, with lines farther from the origin corresponding to greater expected utility. The manufacturer's iso-expected-profit curves are concave relative to the origin, with curves closer to the origin corresponding to greater expected profit. Figure 13.1 illustrates. Note that in Figure 13.1 the salesperson's indifference curves and the manufacturer's iso-expected-profit curves are tangent only at the 45° line, a well-known result from the insurance literature.[10] This shows, graphically, why efficiency (in a first-best sense) requires that the agent not bear risk.

We can re-express (IC″) as

$$u_H \geq u_L + \frac{C}{q}. \qquad (13.6)$$

Loosely speaking, it is the set of x's that have positive probability of occurring.

[10] **Proof:** Let $\phi(\cdot) = u^{-1}(\cdot)$. Then the MRS for the manufacturer is

$$-\frac{(1-q)\phi'(u_L)}{q\phi'(u_H)};$$

whereas the MRS for the salesperson is

$$-\frac{(1-q)}{q}.$$

Since $\phi(\cdot)$ is strictly convex, $\phi'(u) = \phi'(v)$ only if $u = v$; that is, the MRS's of the appropriate iso-curves can be tangent only on the 45° line.

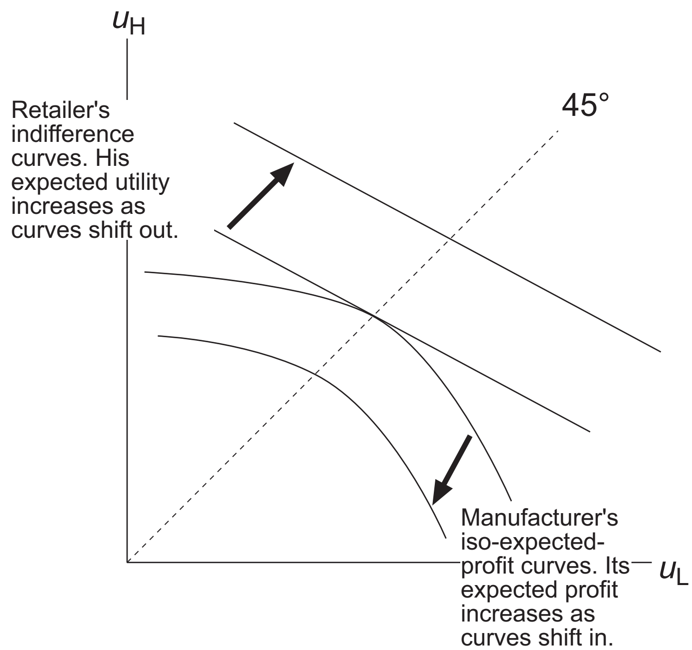

Figure 13.1: Indifference curves in utility space for the manufacturer (principal) and salesperson (agent).

Hence, the set of contracts that are incentive compatible lie on or above a line above, but parallel, to the 45° line. Graphically, we now see that an incentive-compatible contract requires that we abandon non-contingent contracts. Figure 13.2 shows the space of incentive-compatible contracts.

The set of individually rational contracts are those that lie on or above the line defined by (IR$''$). This is also illustrated in Figure 13.2. The intersection of these two regions then constitutes the set of feasible contracts for inducing the salesperson to choose $a = 1$. Observe that the lowest iso-expected-profit curve—corresponding to the largest expected profit—that intersects this set is the one that passes through the "corner" of the set—consistent with our earlier conclusion that *both* constraints are binding at the optimal contract.

Lastly, let's consider the variable q. We can interpret q as representing the correlation—or, more accurately, the informativeness—of sales to the action taken. At first glance, it might seem odd to be worried about the informativeness of sales since, in equilibrium, the principal can accurately predict the agent's choice of action from the structure of the game and her knowledge of the contract. But that's not the point: The principal is forced to design a contract that pays the agent based on performance measures that are informative about

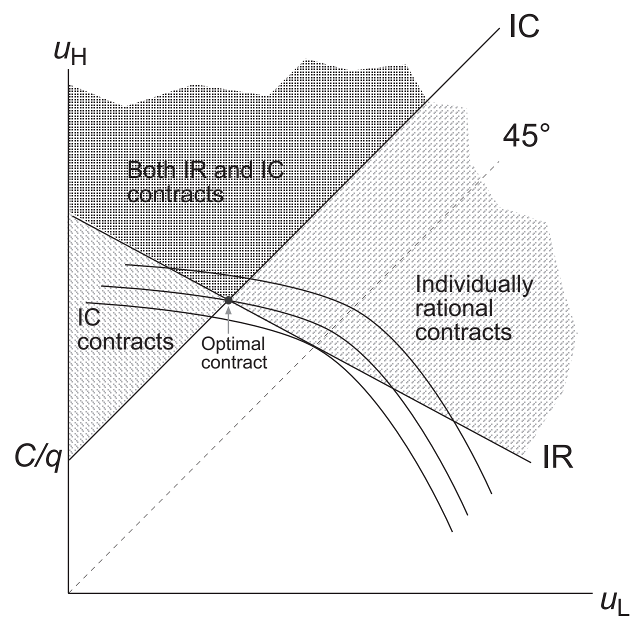

Figure 13.2: The set of feasible contracts.

the variable upon which she would truly like to contract, namely his action. The more informative these performance measures are—loosely, the more correlated they are with action—the closer the principal is getting to the ideal of contracting on the agent's action.

In light of this discussion it wouldn't be surprising if the manufacturer's expected profit under the optimal contract for inducing $a = 1$ increases as q increases. Clearly its expected *revenue*,

$$qx_H + (1 - q)\, x_L,$$

is increasing in q. Hence, it is sufficient to show merely that its expected *cost*,

$$q\hat{s}_H + (1 - q)\, \hat{s}_L,$$

is non-increasing in q. To do so, it is convenient to work in terms of utility (*i.e.*, u_H and u_L) rather than directly with compensation. Let q_1 and q_2 be two distinct values of q, with $q_1 < q_2$. Let $\left\{u_n^1\right\}$ be the optimal contract (expressed in utility terms) when $q = q_1$.

Define $a = q_1/q_2$ and define

$$\tilde{u}_H = a u_H^1 + (1 - a)\, u_L^1 \text{ and}$$
$$\tilde{u}_L = u_L^1.$$

Let us now see that $\{\tilde{u}_n\}$ satisfies both the IR and IC constraints when $q = q_2$. Observe first that

$$\begin{aligned} q_2\tilde{u}_H + (1-q_2)\,\tilde{u}_L &= aq_2u_H^1 + \big((1-a)q_2 + 1 - q_2\big)u_L^1 \\ &= q_1u_H^1 + (1-q_1)\,u_L^1\,. \end{aligned}$$

Given that $\left\{u_n^1\right\}$ satisfies IR and IC when $q = q_1$, it follows, by transitivity, that $\{\tilde{u}_n\}$ solves IR and IC when $q = q_2$. We need, now, simply show that

$$q_2u^{-1}\left(\tilde{u}_H\right) + (1-q_2)\,u^{-1}\left(\tilde{u}_L\right) \le q_1u^{-1}\left(u_H^1\right) + (1-q_1)\,u^{-1}\left(u_L^1\right).$$

To do this, observe that

$$\begin{aligned} &u^{-1}\left(\tilde{u}_H\right) \le au^{-1}\left(u_H^1\right) + (1-a)\,u^{-1}\left(u_L^1\right) \text{ and} \\ &u^{-1}\left(\tilde{u}_L\right) \le u^{-1}\left(u_L^1\right), \end{aligned}$$

by Jensen's inequality (recall $u^{-1}(\cdot)$ is convex). Hence we have

$$\begin{aligned} q_2u^{-1}\left(\tilde{u}_H\right) + (1-q_2)\,u^{-1}\left(\tilde{u}_L\right) &\le q_2\Big(au^{-1}\left(u_H^1\right) + (1-a)\,u^{-1}\left(u_L^1\right)\Big) \\ &\qquad + (1-q_2)\,u^{-1}\left(u_L^1\right) \\ &= q_1u^{-1}\left(u_H^1\right) + (1-q_1)\,u^{-1}\left(u_L^1\right)\,. \end{aligned}$$

In other words, the manufacturer's expected cost is no greater when $q = q_2$ as when $q = q_1$—as was to be shown.

Summary: Although we've considered the simplest of agency models in this section, there are, nevertheless, some general lessons that come from this. First, the optimal contract for inducing an action other than the action that the agent finds least costly requires a contract that is fully contingent on the performance measure. This is a consequence of the action being unobservable to the principal, not the agent's risk aversion. When, however, the agent is risk averse, then the principal's expected cost of solving the hidden-action problem is greater than it would be in the benchmark full-information case: Exposing the agent to risk is inefficient (relative to the first best) and the cost of this inefficiency is borne by the principal. The size of this cost depends on how good an approximation the performance measure is for the variable upon which the principal really desires to contract, the agent's action. The better an approximation (statistic) it is, the lower is the principal's expected cost. If, as here, that shift also raises expected revenue, then a more accurate approximation means greater expected profits. It is also worth pointing out that one result, which might seem as though it should be general, is not: Namely, the result that compensation is increasing with performance (*e.g.*, $\hat{s}_H > \hat{s}_L$). Although this is true when there are only two possible realizations of the performance measure (as we've proved), this result does not hold generally when there are more than two possible realizations.

Multiple-outcomes Model | 13.4

Now we assume that there are multiple possible outcomes, including, possibly, an infinite number. Without loss of generality, we may assume the set of possible sales levels is $(0, \infty)$, given that impossible levels in this range can be assigned zero probability. We will also assume, henceforth, that $u(\cdot)$ exhibits strict risk aversion (*i.e.,* is strictly concave).

Recall that our problem is to solve program (13.3), on page 127. In this context, we can rewrite the problem as

$$\max_{S(\cdot)} \int_0^\infty (x - S(x))\, dF_1(x)$$

subject to

$$\int_0^\infty u[S(x)]\, dF_1(x) - C \geq U_R \text{ and} \tag{13.7}$$

$$\int_0^\infty u[S(x)]\, dF_1(x) - C \geq \int_0^\infty u[S(x)]\, dF_0(x), \tag{13.8}$$

which are the IR and IC constraints, respectively. In what follows, we assume that there is a well-defined density function, $f_a(\cdot)$, associated with $F_a(\cdot)$ for both a. For instance, if $F_a(\cdot)$ is differentiable everywhere, then $f_a(\cdot) = F_a'(\cdot)$ and the $\int dF_a(x)$ notation could be replaced with $\int f_a(x)\, dx$. Alternatively, the possible outcomes could be discrete, $x = x_1, \ldots, x_N$, in which case

$$f_a(\hat{x}) = F_a(\hat{x}) - \lim_{x \uparrow \hat{x}} F_a(x)$$

and the $\int dF_a(x)$ notation could be replaced with $\sum_{n=1}^N f_a(x_n)$.

We solve the above program using standard Kuhn-Tucker techniques. Let μ be the (non-negative) Lagrange multiplier on the incentive constraint and let λ be the (non-negative) Lagrange multiplier on the individual rationality constraint. The Lagrangian of the problem is, thus,

$$\begin{aligned}\mathcal{L}(S(\cdot), \lambda, \mu) = & \int_0^{+\infty} [x - S(x)]\, dF_1(x) + \lambda \left(\int_0^{+\infty} u(S(x)) dF_1(x) - C \right) \\ & + \mu \left(\int_0^{+\infty} u(S(x)) dF_1(x) - \int_0^{+\infty} u(S(x)) dF_0(x) - C \right).\end{aligned}$$

The necessary first-order conditions are $\lambda \geq 0$, $\mu \geq 0$, (13.7), (13.8),

$$u'[S(x)] \left(\lambda + \mu \left[1 - \frac{f_0(x)}{f_1(x)} \right] \right) - 1 = 0, \tag{13.9}$$

$$\lambda > 0 \Rightarrow \int_0^{+\infty} u(S(x)) dF_1(x) = C + U_R, \text{ and}$$

$$\mu > 0 \Rightarrow \int_0^{+\infty} u(S(x))dF_1(x) - \int_0^{+\infty} u(S(x))dF_0(x) = C.$$

From our previous reasoning, we already know that the IC constraint is binding. To see this again, observe that if it were not (*i.e.,* $\mu = 0$), then (13.9) would reduce to $u'[S(x)] = 1/\lambda$ for all x; that is, a fixed payment.[11] But we know a fixed-payment contract is *not* incentive compatible. It is also immediate that the participation constraint must be satisfied as an equality; otherwise, the manufacturer could reduce the payment schedule, thereby increasing her profits, in a manner that preserved the incentive constraint (*i.e.,* replace $S^*(x)$ with $\tilde{S}(x)$, where $\tilde{S}(x) = u^{-1}(u[S^*(x)] - \varepsilon)$).

Note that the necessary conditions above are also sufficient given the assumed concavity of $u(\cdot)$. At every point where it is maximized with respect to s,

$$\left(\lambda + \mu\left[1 - \frac{f_0(x)}{f_1(x)}\right]\right)$$

must be positive—observe it equals

$$\frac{1}{u'(s)} > 0$$

—so the second derivative with respect to s must, therefore, be negative.

The least-cost contract inducing $a = 1$ therefore corresponds to a payment schedule $S(\cdot)$ that varies with the level of sales in a non-trivial way given by (13.9). That expression might look complicated, but its interpretation is central to the model and easy to follow. Observe, in particular, that because $u'(\cdot)$ is a decreasing function ($u(\cdot)$, recall, is strictly concave), $S(x)$ is positively correlated with

$$\lambda + \mu\left[1 - \frac{f_0(x)}{f_1(x)}\right];$$

that is, the larger (smaller) is this term, the larger (smaller) is $S(x)$.

The reward for a given level of sales x depends upon the *likelihood ratio*

$$r(x) \equiv \frac{f_0(x)}{f_1(x)}$$

of the probability that sales are x when action $a = 0$ is taken relative to that probability when action $a = 1$ is taken.[12] This ratio has a clear statistical meaning: It measures how more likely it is that the distribution from which sales have been determined is $F_0(\cdot)$ rather than $F_1(\cdot)$. When $r(x)$ is high, observing sales equal to x allows the manufacturer to draw a statistical inference that it

[11] Note that we've again established the result that, *absent* an incentive problem, a risk-neutral player should absorb all the risk when trading with a risk-averse player.

[12] Technically, if $F_a(\cdot)$ is differentiable (*i.e.,* $f_a(\cdot)$ is a probability density function), then the likelihood ratio can be interpreted as the probability that sales lie in $(x, x + dx)$ when $a = 0$ divided by the probability of sales lying in that interval when $a = 1$.

is much more likely that the distribution of sales was actually $F_0(.)$; that is, the salesperson did not expend effort promoting the product. In this case,

$$\lambda + \mu \left[1 - \frac{f_0(x)}{f_1(x)} \right]$$

is small (but necessarily positive) and $S(x)$ must also be small as well. When $r(x)$ is small, the manufacturer should feel rather confident that the salesperson expended effort and it should, then, optimally reward him highly. That is, sales levels that are relatively more likely when the agent has behaved in the desired manner result in larger payments to the agent than sales levels that are relatively rare when the agent has behaved in the desired manner.

The minimum-cost incentive contract that induces the costly action $a = 1$ in essence commits the principal (manufacturer) to behave like a Bayesian statistician who holds some diffuse prior over which action the agent has taken:[13] She should use the observation of sales to revise her beliefs about what action the agent took and she should reward the agent more for outcomes that cause her to revise *upward* her beliefs that he took the desired action and she should reward him less (punish him) for outcomes that cause a *downward* revision in her beliefs.[14] As a consequence, the payment schedule is connected to sales only through their statistical content (the *relative* differences in the densities), *not* through their accounting properties. In particular, there is now no reason to believe that higher sales (larger x) should be rewarded more than lower sales.

As an example of non-monotonic compensation, suppose that there are three possible sales levels: low, medium, and high (x_L, x_M, and x_H, respectively). Suppose, in addition, that

$$f_a(x) = \begin{cases} \frac{1}{3}, \text{ if } x = x_L \\ \frac{2-a}{6}, \text{ if } x = x_M \\ \frac{2+a}{6}, \text{ if } x = x_H \end{cases} .$$

Then

$$\lambda + \mu \left[1 - \frac{f_0(x)}{f_1(x)} \right] = \begin{cases} \lambda, \text{ if } x = x_L \\ \lambda - \mu, \text{ if } x = x_M \\ \lambda + \frac{\mu}{3}, \text{ if } x = x_H \end{cases} .$$

Hence, low sales are rewarded more than medium sales—low sales are uninformative about the salesperson's action, whereas medium sales suggest that the salesperson has *not* expended effort. Admittedly, non-monotonic compensation is rarely, if ever, observed in real life. We will see below what additional properties are required, in this model, to ensure monotonic compensation.

---

[13] A diffuse prior is one that assigns positive probability to each possible action.

[14] Of course, as a rational player of the game, the principal can infer that, if the contract is incentive compatible, the agent will have taken the desired action. Thus, there is not, in some sense, a real inference problem. Rather the issue is that, to be incentive compatible, the principal must commit to act *as if* there were an inference problem.

Note, somewhat implicit in our analysis to this point, is an assumption that $f_1(x) > 0$ except, possibly, on a subset of x that are impossible (have zero measure). Without this assumption, (13.9) would entail division by zero, which is, of course, not permitted. If, however, we let $f_1(\cdot)$ go to zero on some subset of x that had positive measure under $F_0(\cdot)$, then we see that μ must also tend to zero because

$$\lambda + \mu \left[1 - \frac{f_0(x)}{f_1(x)}\right]$$

must be positive. In essence, then, the shadow price (cost) of the incentive constraint vanishes as $f_1(\cdot)$ goes to zero. This makes perfect sense: Were $f_1(\cdot)$ zero on some subset of x that could occur (had positive measure) under $F_0(\cdot)$, then the occurrence of any x in this subset, $\mathcal{X}_0$, would be proof that the agent had failed to take the desired action. We can use this, then, to design a contract that induces $a = 1$, but which costs the principal no more than the optimal full-information fixed-payment contract $S(x) = s_1^F$. That is, the incentive problem ceases to be costly; so, not surprisingly, its shadow cost is zero.

To see how we can construct such a contract when $f_1(x) = 0$ for all $x \in \mathcal{X}_0$, let

$$S(x) = \begin{cases} \underline{s} + \varepsilon, \text{ if } x \in \mathcal{X}_0 \\ s_1^F, \text{ if } x \notin \mathcal{X}_0 \end{cases},$$

where $\varepsilon > 0$ is arbitrarily small ($\underline{s}$, recall, is the greatest lower bound of the domain of $u(\cdot)$). Then

$$\int_0^{+\infty} u(S(x))dF_1(x) = u\left(s_1^F\right) \text{ and}$$
$$\int_0^{+\infty} u(S(x))dF_0(x) = \int_{\mathcal{X}_0} u\left(\underline{s} + \varepsilon\right) dF_0(x) + \int_{\mathbb{R}_+ \setminus \mathcal{X}_0} u\left(s_1^F\right) dF_0(x)$$
$$= u\left(\underline{s} + \varepsilon\right) F_0(\mathcal{X}_0) + u\left(s_1^F\right)\left(1 - F_0(\mathcal{X}_0)\right).$$

From the last expression, it's clear that $\int_0^{+\infty} u(S(x))dF_0(x) \to -\infty$ as $\varepsilon \to 0$; hence, the IC constraint is met trivially. By the definition of s_1^F, IR is also met. That is, this contract implements $a = 1$ at full-information cost. Again, as we saw in the two-outcome model, having a *shifting support* (*i.e.,* the property that $F_0(\mathcal{X}_0) > 0 = F_1(\mathcal{X}_0)$) allows us to implement the desired action at full-information cost.

To conclude this section, we need to answer one final question. Does the manufacturer prefer to induce $a = 1$ or $a = 0$ given the "high" cost of the former? The manufacturer's choice can be viewed as follows: Either it imposes the fixed-payment contract s_0^F, which induces action $a = 0$, or it imposes the contract $S(\cdot)$ that has been derived above, which induces action $a = 1$. The expected-profit-maximizing choice results from the simple comparison of these two contracts; that is, the manufacturer imposes the incentive contract $S(\cdot)$ if and only if:

$$\mathbb{E}_0[x] - s_0^F < \mathbb{E}_1[x] - \mathbb{E}_1[S(x)] \tag{13.10}$$

The right-hand side of this inequality corresponds to the value of the maximization program (13.3). Given that the incentive constraint is binding in this program, this value is strictly smaller than the value of the same program without the incentive constraint; hence, just as we saw in the two-outcome case, the value is smaller than full-information profits, $E_1[x] - s_1^F$. Observe, therefore, that it is possible that

$$\mathbb{E}_1[x] - s_1^F > \mathbb{E}_0[x] - s_0^F > \mathbb{E}_1[x] - \mathbb{E}_1[S(x)]:$$

Under full information the principal would induce $a = 1$, but not if there's a hidden-action problem. In other words, imperfect observability of the agent's action imposes a cost on the principal that may induce her to distort the action that she induces the agent to take.

Monotonicity of the Optimal Contract 13.5

Let us suppose that (13.10) is, indeed, satisfied so that the contract $S(\cdot)$ derived above is the optimal contract. Can we exhibit additional and meaningful assumptions that would imply interesting properties of the optimal contract?

We begin with monotonicity, the idea that greater sales should mean greater compensation for the salesperson. As we saw above (page 137), there is no guarantee in the multiple-outcome model that this property should hold everywhere. From (13.9), it does hold if and only if the likelihood ratio, $r(x)$, is not decreasing and increasing at least somewhere. As this is an important property, it has a name:

Definition 5 *The likelihood ratio $r(x) = f_0(x)/f_1(x)$ satisfies the monotone likelihood ratio property(* MLRP*) if $r(\cdot)$ is non-increasing almost everywhere and strictly decreasing on at least some set of x's that occur with positive probability given action $a = 1$.*

The MLRP states that the greater is the outcome (*i.e.,* x), the greater the *relative* probability of x given $a = 1$ than given $a = 0$.[15] In other words, under MLRP, better outcomes are more likely when the salesperson expends effort than when he doesn't. To summarize:

Proposition 19 *In the model of this section, if the likelihood ratio, $r(\cdot)$, satisfies the monotone likelihood ratio property, then the salesperson's compensation, $S(\cdot)$, under the optimal incentive contract for inducing him to expend effort (i.e., to choose $a = 1$) is non-decreasing everywhere.*

---

[15] Technically, if $f_a(x) = F'_a(x)$, then we should say "the greater the relative probability of an $\hat{x} \in (x, x + dx)$ given $a = 1$ than given $a = 0$."

In fact, because we know that $S(\cdot)$ can't be constant—a fixed-payment contract is not incentive compatible for inducing $a = 1$—we can conclude that $S(\cdot)$ must, therefore, be increasing over some set of x.[16]

Is MLRP a reasonable assumption? To some extent is simply a strengthening of our assumption that $\mathbb{E}_1[x] > \mathbb{E}_0[x]$, given it can readily be shown that MLRP implies $\mathbb{E}_1[x] > \mathbb{E}_0[x]$.[17] Moreover, many standard distributions satisfy MLRP. But it quite easy to exhibit meaningful distributions that do not. For instance, consider our example above (page 137). We could model these distributions as the consequence of a two-stage stochastic phenomenon: With probability 1/3, a second new product is successfully introduced that eliminates the demand for the manufacturer's product (*i.e.,* $x_L = 0$). With probability 2/3, this second product is not successfully introduced and it is "business as usual," with sales being more likely to be x_M if the salesperson doesn't expend effort and more likely to be x_H if he does. These "compound" distributions do not satisfy MLRP. In such a situation, MLRP is not acceptable and the optimal reward schedule is not monotonic, as we saw.

Although there has been a lot of discussion in the literature on the monotonicity issue, it may be overemphasized. If we return to the economic reality that the mathematics seeks to capture, the discussion relies on the assumption that a payment schedule with the feature that the salesperson is penalized for increasing sales in some range does not actual induce a new agency problem. For instance, if the good is perishable or costly to ship, it might be possible for the salesperson to pretend, when sales are x_M, that they are x_L (if the allegedly

---

[16] Even if MLRP does *not* hold, $r(\cdot)$ must be decreasing over some measurable range. To see this, suppose it were not true; that is, suppose that $r(\cdot)$ is almost everywhere non-decreasing under distribution $F_1(\cdot)$. Note this entails that x and $r(x)$ are *non-negatively* correlated under $F_1(\cdot)$. To make the exposition easier, suppose for the purpose of this aside that $f_a(\cdot) = F'_a(\cdot)$. Then

$$\mathbb{E}_1\left[\frac{f_0(x)}{f_1(x)}\right] = \int_0^\infty \left(\frac{f_0(x)}{f_1(x)}\right) f_1(x)dx = \int_0^\infty f_0(x)dx = 1$$

Because x and $r(x)$ are non-negatively correlated, we thus have

$$\int_0^\infty x[r(x) - \mathbb{E}_1\{r(x)\}]f_1(x)dx \geq 0.$$

Substituting, this implies

$$0 \leq \int_0^\infty x[\frac{f_0(x)}{f_1(x)} - 1]f_1(x)dx = \int_0^\infty xf_0(x)dx - \int_0^\infty xf_1(x)dx = \mathbb{E}_0[x] - \mathbb{E}_1[x].$$

But this contradicts our assumption that investing ($a = 1$) yields *greater* expected revenues than does not investing ($a = 0$). Hence, by contradiction it must be that $r(\cdot)$ is decreasing over some measurable range. But then this means that $S(\cdot)$ is increasing over some measurable range. However, without MLRP, we can't conclude that it's not also decreasing over some other measurable range.

Conclusion: *If* $\mathbb{E}_0[x] < \mathbb{E}_1[x]$*, then* $S(\cdot)$ is *increasing over some set of* x *that has positive probability of occurring given action* $a = 1$ *even if* MLRP *does* not *hold.*

[17] This can most readily be seen from the previous footnote: Simply assume that $r(\cdot)$ satisfies MLRP, which implies x and $r(x)$ are *negatively* correlated. Then, following the remaining steps, it quickly falls out that $\mathbb{E}_1[x] > \mathbb{E}_0[x]$.

unsold amount of the good cannot be returned). That is, a non-monotonic incentive scheme could introduce a new agency problem of making sure the salesperson reports his sales honestly. Of course, if the manufacturer can verify the salesperson's inventory, a non-monotonic scheme might be possible. But think of another situation where the problem is to provide a worker (the agent) incentives to produces units of output; isn't it natural, then, to think that the worker could very well stop his production at x_L or destroy his extra production $x_M - x_L$? Think of yet another situation where the problem is to provide incentives to a manager; aren't there many ways to spend money in a hardly detectable way so as to make profits look smaller than what they actually are? In short, the point is that if the agent can freely and secretly diminish his performance, then it makes no sense for the principal to have a reward schedule that is decreasing with performance over some range. In other words, there is often an *economic* justification for monotonicity even when MLRP doesn't hold.

13.6 Informativeness of the Performance Measure

Now, we again explore the question of the informativeness of the performance measures used by the principal. To understand the issue, suppose that the manufacturer in our example can also observe the sales of another of its products sold by the salesperson. Let y denote these sales of the other good. These sales are, in part, also random, affected by forces outside the parties' control; but also, possibly, determined by how much effort the salesperson expends promoting the first product. For example, consumers could consider the two goods complements (or substitutes). Sales y will then co-vary positively (or negatively) with sales x. Alternatively, both goods could be normal goods, so that sales of y could then convey information about general market conditions (the incomes of the customers). Of course, it could also be that the demand for the second product is wholly unrelated to the demand for the first; in which case sales y would be insensitive to the salesperson's action.

Let $f_0(x,y)$ and $f_1(x,y)$ denote the joint probability densities of sales x and y for action $a = 0$ and $a = 1$. An incentive contract can now be a function of both performance variables; that is, $s = S(x,y)$. It is immediate that the same approach as before carries through and yields the following optimality condition:[18]

$$u'\left[S(x,y)\right]\left(\lambda + \mu\left[1 - \frac{f_0(x,y)}{f_1(x,y)}\right]\right) - 1 = 0. \qquad (13.11)$$

When is it optimal to make compensation a function of y as well as of x?

[18] Although we use the same letters for the Lagrange multipliers, it should be clear that their values at the optimum are not related to their values in the previous, one-performance-measure, contracting problem.

The answer is straightforward: When the likelihood ratio,

$$r(x,y) = \frac{f_0(x,y)}{f_1(x,y)},$$

actually depends upon y. Conversely, when the likelihood ratio is independent of y, then there is no gain from contracting on y to induce $a = 1$; indeed, it would be sub-optimal in this case because such a compensation scheme would fail to satisfy (13.11). The likelihood ratio is independent of y if and only if the following holds: There exist three functions $h(\cdot,\cdot)$, $g_0(\cdot)$, and $g_1(\cdot)$ such that, for all (x,y),

$$f_a(x,y) = h(x,y)g_a(x). \tag{13.12}$$

Sufficiency is obvious (divide $f_0(x,y)$ by $f_1(x,y)$ and observe the ratio, $g_0(x)/g_1(x)$, is independent of y). Necessity is also straightforward: Set $h(x,y) = f_1(x,y)$, $g_1(x) = 1$, and $g_0(x) = r(x)$. This condition of multiplicative separability, (13.12), has a well-established meaning in statistics: If (13.12) holds, then x is a *sufficient statistic* for the action a given data (x,y). In words, were we trying to infer a, our inference would be just as good if we observed only x as it would be if we observed the pair (x,y). That is, conditional on knowing x, y is *un*informative about a.

This conclusion is, therefore, quite intuitive, once we recall that the value of performance measures to our contracting property rests solely on their statistical properties. The optimal contract should be based on all performance measures that convey information about the agent's decision; but it is not desirable to include performance measures that are statistically redundant with other measures. As a corollary, there is no gain from considering *ex post* random contracts (*e.g.,* a contract that based rewards on $x+\eta$, where η is some random variable—noise—distributed independently of a that is added to x).[19] As a second corollary, if the principal could freely eliminate noise in the performance measure—that is, switch from observing $x+\eta$ to observing x—she would do better (at least weakly).

13.7 Conclusions from the Two-action Model

It may be worth summing up all the conclusions we have reached within the two-action model in a proposition:

Proposition 20 *If the agent is strictly risk averse, there is no shifting support, and the principal seeks to implement the costly action (i.e., $a = 1$), then the principal's expected profits are smaller than under full (perfect) information. In some instances, this reduction in expected profits may lead the principal to implement the "free" action (i.e., $a = 0$).*

[19] *Ex ante* random *contracts* may, however, be of some value, as explained later.

- *When (13.10) holds, the reward schedule imposes some risk on the risk-averse agent: Performances that are more likely when the agent takes the correct action* $a = 1$ *are rewarded more than performances that are more likely under* $a = 0$*.*
- *Under* MLRP *(or when the agent can destroy output), the optimal reward schedule is non-decreasing in performance.*
- *The optimal reward schedule depends only upon performance measures that are sufficient statistics for the agent's action.*

To conclude, allow me to stress the two major themes that I would like the reader to remember from this section. First, imperfect information implies that the contractual reward designed by the principal should perform two tasks: Share the risks involved in the relationship and provide incentives to induce the agent to undertake the desired action. Except in trivial cases (*e.g.,* a risk-neutral agent or a shifting support), these two goals are in conflict. Consequently, the optimal contract may induce an inefficient action *and* a Pareto suboptimal sharing of risk.[20] Second, the optimal reward schedule establishes a link between rewards and performances that depends upon the statistical properties of the performance measures with respect to the agent's action.

Bibliographic Notes

The analysis presented so far is fairly standard. The two-step approach—first determine, separately, the optimal contracts for implementing $a = 0$ and $a = 1$, then choose which yields greater profits—is due to Grossman and Hart (1983). The analysis, in the two-outcome case, when q varies is also based on their work. They also consider the monotonicity of the optimal contract, although our analysis here draws more from Holmström (1979). Holmström is also the source for the sufficient-statistic result. Finally, the expression

$$\frac{1}{u'\left[S\left(x\right)\right]} = \lambda + \mu\left[1 - \frac{f_0(x)}{f_1(x)}\right],$$

which played such an important part in our analysis, is frequently referred to as the *modified* Borch sharing rule, in honor of Borch (1968), who worked out the rules for optimal risk sharing *absent* a moral-hazard problem (hence, the adjective "modified").

[20] Think of this as yet another example of the two-instruments principle.

General Framework 14

As we've just seen, the two-action model yields strong results. But the model incorporates a lot of structure and it relies on strong assumptions. Consequently, it's hard to understand which findings are robust and which are merely artifacts of an overly simple formalization. The basic ideas behind the incentive model are quite deep and it is worthwhile, therefore, to consider whether and how they generalize in less constrained situations.

Our approach is to propose a very general framework that captures the situation described in the opening section. Such generality comes at the cost of tractability, so we will again find ourselves making specific assumptions. But doing so, we will try to motivate the assumptions we have to make and discuss their relevance or underline how strong they are.

The situation of incentive contracting under hidden action or imperfect monitoring involves:

- a principal;
- an agent;
- a set of possible actions, $\mathcal{A}$, from which the agent chooses (we take $\mathcal{A}$ to be exogenously determined here);
- a set of verifiable signals or performance measures, $\mathcal{X}$;
- a set of benefits, $\mathcal{B}$, for the principal that are affected by the agent's action (possibly stochastically);
- rules (functions, distributions, or some combination) that relate elements of $\mathcal{A}$, $\mathcal{X}$, and $\mathcal{B}$;
- preferences for the principal and agent; and
- a bargaining game that establishes the contract between principal and agent (here, recall, we've fixed the bargaining game as the principal makes a take-it-or-leave-it offer, so that the only element of the bargaining game of interest here is the agent's reservation utility, U_R).[1]

[1] We could also worry about whether the principal wants to participate—even make a take-it-or-leave-it offer—but because our focus is on the contract design and its execution, stages of the game not reached if she doesn't wish to participate, we will not explicitly consider this issue here.

In many settings, including the one explored above, the principal's benefit is the same as the verifiable performance measure (*i.e.,* $b = x$). But this need not be the case. We could, for instance, imagine that there is a *function* mapping the elements of $\mathcal{A}$ onto $\mathcal{B}$. For example, the agent's action could be fixing the "true" quality of a product produced for the principal. This quality is also, then, the principal's benefit (*i.e.,* $b = a$). The only *verifiable* measure of quality, however, is some noisy (*i.e.,* stochastic) measure of true quality (*e.g.,* $x = a + \eta$, where η is some randomly determined distortion). As yet another possibility, the benchmark case of full information entails $\mathcal{X} = \mathcal{X}' \times \mathcal{A}$, where $\mathcal{X}'$ is some set of performance measures other than the action.

We need to impose some structure on $\mathcal{X}$ and $\mathcal{B}$ and their relationship to $\mathcal{A}$: We take $\mathcal{X}$ to be a Euclidean vector space and we let $dF(\cdot|a)$ denote the probability measure over $\mathcal{X}$ conditional on a. Similarly, we take $\mathcal{B}$ to be a Euclidean vector space and we let $dG(\cdot,\cdot|a)$ denote the joint probability measure over $\mathcal{B}$ and $\mathcal{X}$ conditional on a (when $b \equiv x$, we will write $dF(\cdot|a)$ instead of $dG(\cdot,\cdot|a)$). This structure is rich enough to encompass the possibilities enumerated in the previous paragraph (and more).

Although we could capture the preferences of the principal and agent with*out* assuming the validity of the expected-utility approach to decision-making under uncertainty (we could, for instance, take as primitives the indifference curves shown in Figures 13.1 and 13.2), this approach has not been taken in the literature.[2] Instead, the expected-utility approach is assumed to be valid and we let $W(s, x, b)$ and $U(s, x, a)$ denote the respective von Neumann-Morgenstern utilities of the principal and of the agent, where s denotes the transfer from the principal to the agent (to principal from agent if $s < 0$).

In this situation, the obvious contract is a function that maps $\mathcal{X}$ into $\mathbb{R}$. We define such a contract as

Definition 6 *A* simple *incentive contract is a reward schedule $S : \mathcal{X} \to \mathbb{R}$ that determines the level of reward $s = S(x)$ to be decided as a function of the realized performance level x.*

There is admittedly no other verifiable variable that can be used to write more elaborate contracts. There is, however, the possibility of creating verifiable variables, by having one or the other or both players take *verifiable* actions from some specified action spaces. Consistent with the mechanism-design approach, the most natural interpretation of these new variables are that they are public announcements made by the players; but nothing that follows requires this interpretation. For example, suppose both parties have to report to the third party charged with enforcing the contract their observation of x, or the agent

[2]Given some well-documented deficiencies in expected-utility theory (see, *e.g.,* Epstein, 1992; Rabin, 1997), this might, at first, seem somewhat surprising. However, as Epstein, §2.5, notes many of the predictions of expected-utility theory are robust to relaxing some of the more stringent assumptions that support it (*e.g.,* such as the independence axiom). Given the tractability of the expected-utility theory combined with the general empirical support for the predictions of agency theory, the gain from sticking with expected-utility theory would seem to outweigh the losses, if any, associated with that theory.

must report which action he has chosen. We could even let the principal make a "good faith" report of what action she believes the agent took, although this creates its own moral-hazard problem because, in most circumstances, the principal could gain *ex post* by claiming she believes the agent's action was unacceptable. It turns out, as we will show momentarily, that there is nothing to be gained by considering such elaborate contracts; that is, there is no such contract that can improve over the optimal simple contract.

To see this, let us suppose that a contract determines a normal-form game to be played by both players after the agent has taken his action.[3,4] In particular, suppose the agent takes an action $h \in \mathcal{H}$ after choosing his action, but prior to the realization of x; that he takes an action $m \in \mathcal{M}$ after the realization of x; and that the principal also takes an action $n \in \mathcal{N}$ after x has been realized. One or more of these sets could, but need not, contain a single element, a "null" action. We assume that the actions in these sets are costless—if we show that costless elaboration does no better than simple contracts, then costly elaboration also cannot do better than simple contracts. Finally, let the agent's compensation under this elaborate contract be: $s = \tilde{S}(x, h, m, n)$. We can now establish the following:

Proposition 21 (Simple contracts are sufficient) *For any general contract* $\left\langle \mathcal{H}, \mathcal{M}, \mathcal{N}, \tilde{S}(\cdot) \right\rangle$ *and associated (perfect Bayesian) equilibrium, there exists a simple contract* $S(\cdot)$ *that yields the same equilibrium outcome.*

Proof: Consider a (perfect Bayesian) equilibrium of the original contract, involving strategies $(a^*, h^*(\cdot), m^*(\cdot, \cdot))$ for the agent, where $h^*(a)$ and $m^*(a, x)$ describe the agent's choice within $\mathcal{H}$ and $\mathcal{M}$ after he's taken action a and performance x has been observed. Similarly, $n^*(x)$ gives the principal's choice of action as a function of the observed performance. Let us now consider the simple contract defined as follows: For all $x \in \mathcal{X}$,

$$S(x) \equiv \tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)).$$

Suppose that, facing this contract, the agent chooses an action a different from a^*. This implies that:

$$\int_{\mathcal{X}} U(S(x), x, a) dF(x|a) > \int_{\mathcal{X}} U(S(x), x, a^*) dF(x|a^*),$$

[3]Note this may require that there be some way that the parties can verify that the agent has taken an action. This may simply be the passage of time: The agent must take his action before a certain date. Alternatively, there could be a verifiable signal that the agent has acted (but which does not reveal *how* he's acted).

[4]Considering a extensive-form game with the various steps just considered would not alter the reasoning that follows; so, we avoid these unnecessary details by restricting attention to a normal-form game.

or, using the definition of $S(\cdot)$,

$$\int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)), x, a\right] dF(x|a) >$$
$$\int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)), x, a^*\right] dF(x|a^*).$$

Because, *in the equilibrium of the normal-form game that commences* after *the agent chooses his action*, $h^*(\cdot)$ and $m^*(\cdot, \cdot)$ must satisfy the following inequality:

$$\int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a), m^*(x, a), n^*(x)), x, a\right] dF(x|a) \geq$$
$$\int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)), x, a\right] dF(x|a),$$

it follows that

$$\int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a), m^*(x, a), n^*(x)), x, a\right] dF(x|a) >$$
$$\int_{\mathcal{X}} U\left[\tilde{S}(x, h^*(a^*), m^*(x, a^*), n^*(x)), x, a^*\right] dF(x|a^*).$$

This contradicts the fact the a^* is an equilibrium action in the game defined by the original contract. Hence, the simple contract $S(\cdot)$ gives rise to the same action choice, and therefore the same distribution of outcomes than the more complicated contract. ■

As a consequence, there is no need to consider sophisticated announcement mechanisms in this setting, at least in the simple situation we have described. The style in which we proved Proposition 21 should be familiar. In particular, it was how we proved the *Revelation Principle* (Proposition 11 and Theorem 16).

The contracting problem under imperfect information can now easily be stated. The principal, having the bargaining power in the negotiation process, simply has to choose a (simple) contract, $S(\cdot)$, so as to maximize her expected utility from the relationship given two constraints. First, the contract $S(\cdot)$ induces the agent to choose an action that maximizes his expected utility (*i.e.,* the IC constraint must be met). Second, given the contract and the action it will induce, the agent must receive an expected utility at least as great as his reservation utility (*i.e.,* the IR constraint must be met). In this general setting, the IC constraint can be stated as the action induced, a, must satisfy

$$a \in \underset{a'}{\operatorname{argmax}} \int_{\mathcal{X}} U(S(x), x, a') dF(x|a'). \tag{14.1}$$

Observe that choosing $S(\cdot)$ amounts to choosing a as well, at least when there exists a unique optimal choice for the agent. To take care of the possibility of multiple optima for the agent, one can simply imagine that the principal chooses a pair $(S(\cdot), a)$ subject to the incentive constraint (14.1). The IR constraint takes the simple form:

$$\max_{a'} \int_{\mathcal{X}} U(S(x), x, a')dF(x|a') \geq U_R. \tag{14.2}$$

The principal's problem is, thus,

$$\max_{(S(\cdot),a)} \int_{\mathcal{X}} W(S(x), x, b)dG(b, x|a) \tag{14.3}$$
$$\text{s.t. (14.1) and (14.2).}$$

Observe, as we did in Lecture Note 13, it is perfectly permissible to solve this maximization program in two steps. First, for each action a, find the expected-profit-maximizing contract that implements action a subject to the IC and IR constraints; this amounts to solving a similar program, taking action a as fixed:

$$\max_{S(.)} \int_{\mathcal{X}} W(S(x), x, b)dG(b, x|a) \tag{14.4}$$
$$\text{s.t. (14.1) and (14.2).}$$

Second, optimize the principal's objectives with respect to the action to implement; if we let $S_a(\cdot)$ denote the expected-profit-maximizing contract for implementing a, this second step consists of:

$$\max_{a \in \mathcal{A}} \int_{\mathcal{X}} W(S_a(x), x, b)dG(b, x|a).$$

In this more general framework, it's worth revisiting the full-information benchmark. Before doing that, however, it is worth assuming that the domain of $U(\cdot, x, a)$ is sufficiently broad:

- EXISTENCE OF A PUNISHMENT: There exists some s_P in the domain of $U(\cdot, x, a)$ such that, for all $a \in \mathcal{A}$,

$$\int_{\mathcal{X}} U(s_P, x, a)dF(x|a) < U_R\,.$$

- EXISTENCE OF A SUFFICIENT REWARD: There exists some s_R in the domain of $U(\cdot, x, a)$ such that, for all $a \in \mathcal{A}$,

$$\int_{\mathcal{X}} U(s_R, x, a)dF(x|a) \geq U_R\,.$$

In light of the second assumption, we can always satisfy (14.2) for any action a (there is no guarantee, however, that we can also satisfy (14.1)).

With these two assumptions in hand, suppose that we're in the full-information case; that is, $\mathcal{X} = \mathcal{X}' \times \mathcal{A}$ (note $\mathcal{X}'$ could be a single-element space, so that we're also allowing for the possibility that, effectively, the only performance measure is the action itself). In the full-information case, the principal can rely on *forcing contracts*; that is, contracts that effectively leave the agent with no choice over the action he chooses. Hence, writing (x', a) for an element of $\mathcal{X}$, a forcing contract for implementing $\hat{a}$ is

$$\begin{aligned} S(x', a) &= s_P \text{ if } a \neq \hat{a} \\ &= S^F(x') \text{ if } a = \hat{a}, \end{aligned}$$

where $S^F(\cdot)$ satisfies (14.2). Given that $S^F(x') = s_R$ satisfies (14.2) by assumption, we know that we can find *an* $S^F(\cdot)$ function that satisfies (14.2). In equilibrium, the agent will choose to sign the contract—the IR constraint is met—and he will take action $\hat{a}$ since this is his only possibility for getting at least his reservation utility. Forcing contracts are very powerful because they transform the contracting problem into a simple *ex ante* Pareto computation program:

$$\begin{aligned} &\max_{(S(\cdot),a)} \int_{\mathcal{X}} W(S(x), x, b) dG(b, x|a) \\ &\text{s.t. } (14.2), \end{aligned} \tag{14.5}$$

where only the agent's participation constraint matters. This *ex ante* Pareto program determines the efficient risk-sharing arrangement for the full-information optimal action, as well as the full-information optimal action itself. Its solution characterizes the optimal contract under perfect information.

At this point, we've gone about as far as we can go without imposing more structure on the problem. The next couple of Lecture Notes consider more structured variations of the problem.

The Finite Model | 15

In this Lecture Note, we will assume $\mathcal{A}$, the set of possible actions, is finite with J elements. Likewise, the set of possible verifiable performance measures, $\mathcal{X}$, is also taken to be finite with N elements, indexed by n (although, at the end of this section, we'll discuss the case where $\mathcal{X} = \mathbb{R}$). Given the world is, in reality, finite, this is the most general version of the principal-agent model (although not necessarily the most analytically tractable).[1]

We will assume that the agent's utility is additively separable between payments and action. Moreover, it is not, directly, dependent on performance. Hence,

$$U(s, x, a) = u(s) - c(a);$$

where $u : \mathcal{S} \to \mathbb{R}$ maps some subset $\mathcal{S}$ of $\mathbb{R}$ into $\mathbb{R}$ and $c : \mathcal{A} \to \mathbb{R}$ maps the action space into $\mathbb{R}$. As before, we assume that $\mathcal{S} = (\underline{s}, \infty)$, where $u(s) \to -\infty$ as $s \downarrow \underline{s}$. Observe that this assumption entails the existence of a punishment, s_P, as described in the previous section. We further assume that $u(\cdot)$ is strictly monotonic and concave (at least weakly). Typically, we will assume that $u(\cdot)$ is, in fact, strictly concave, implying the agent is risk averse. Note, that the monotonicity of $u(\cdot)$ implies that the inverse function $u^{-1}(\cdot)$ exists and, since $u(\mathcal{S}) = \mathbb{R}$, is defined for all $u \in \mathbb{R}$.

We assume, now, that $\mathcal{B} \subset \mathbb{R}$ and that the principal's utility is a function only of the difference between her benefit, b, and her payment to the agent; that is,

$$W(s, x, b) = w(b - s),$$

where $w(\cdot)$ is assumed to be strictly increasing and concave. In fact, in most applications—particularly most applications of interest in the study of strategy and organization—it is reasonable to assume that the principal is risk neutral. We will maintain that assumption here (the reader interested in the case of a risk-averse principal should consult Holmström, 1979, among other work). In what follows, let $B(a) = \mathbb{E}\{b|a\}$.

In addition to being discrete, we assume that there exists some partial order on $\mathcal{X}$ (*i.e.,* to give meaning to the idea of "better" or "worse" performance) and

[1] Limitations on measurement—of both inputs and outputs—make the world discrete. For instance, effort, measured as time on the job, can take only discrete values because only discrete intervals of time can be measured. Similarly, output, measured as volume, can take only discrete values because only discrete intervals of volume can be measured. Because the world is also bounded—each of us is allotted only so much time and there are only so many atoms in the universe (under current cosmological theories)—it follows that the world is finite.

that, with respect to this partial order, $\mathcal{X}$ is a chain (*i.e.,* if $\preceq$ is the partial order on $\mathcal{X}$, then $x \preceq x'$ or $x' \preceq x$ for any two elements, x and x' in $\mathcal{X}$). Because identical signals are irrelevant, we may also suppose that no two elements of $\mathcal{X}$ are the same (*i.e.,* $x \prec x'$ or $x' \prec x$ for any two elements in $\mathcal{X}$). The most natural interpretation is that $\mathcal{X}$ is a subset of distinct real numbers—different "performance scores"—with $\leq$ as the partial order. Given these assumptions, we can write $\mathcal{X} = \{x_1, \ldots, x_N\}$, where $x_m \prec x_n$ if $m < n$. Likewise the distribution function $F(x|a)$ gives, for each x, the probability that an $x' \preceq x$ is realized conditional on action a. The corresponding density function is then defined by

$$f(x_n|a) = \begin{cases} F(x_1|a) & \text{if } n = 1 \\ F(x_n|a) - F(x_{n-1}|a) & \text{if } n > 1 \end{cases} .$$

In much of what follows, it will be convenient to write $f_n(a)$ for $f(x_n|a)$. It will also be convenient to write the density as a vector:

$$\mathbf{f}(a) = (f_1(a), \ldots, f_N(a)) .$$

The "Two-step" Approach 15.1

As we have done already in Lecture Note 13, we will pursue a two-step approach to solving the principal-agent problem:

Step 1: For each $a \in \mathcal{A}$, the principal determines whether it can be implemented. Let $\mathcal{A}^I$ denote the set of implementable actions. For each $a \in \mathcal{A}^I$, the principal determines the least-cost contract for implementing a subject to the IC and IR constraints. Let $C(a)$ denote the principal's expected cost (expected payment) of implementing a under this least-cost contract.

Step 2: The principal then determines the solution to the maximization problem

$$\max_{a \in \mathcal{A}^I} B(a) - C(a) .$$

If a^* is the solution to this maximization problem, the principal offers the least-cost contract for implementing a^*.

Note that this two-step process is analogous to a standard production problem, in which a firm, first, solves its cost-minimization problems to determine the least-cost way of producing any given amount of output (*i.e.,* derives its cost function); and, then, it produces the amount of output that maximizes the difference between revenues (benefits) and cost. As with production problems, the first step is generally the harder step.

The full-information benchmark

As before, we consider as a benchmark the case where the principal can observe and verify the agent's action. Consequently, as we discussed at the end of

Lecture Note 14, the principal can implement any action $\hat{a}$ that she wants using a forcing contract: The contract punishes the agent sufficiently for choosing actions $a \neq \hat{a}$ that he would never choose any action other than $\hat{a}$; and the contract rewards the agent sufficiently for choosing $\hat{a}$ that he is just willing to sign the principal's contract. This last condition can be stated formally as

$$u(\hat{s}) - c(\hat{a}) = U_R, \qquad (\text{IR}^F)$$

where $\hat{s}$ is what the agent is paid if he chooses action $\hat{a}$. Solving this last expression for $\hat{s}$ yields

$$\hat{s} = u^{-1}\left[U_R + c(\hat{a})\right] \equiv C^F(\hat{a}).$$

The function $C^F(\cdot)$ gives the cost, under full information, of implementing actions.

The hidden-action problem

Now, and henceforth, we assume that a hidden-action problem exists. Consequently, the only feasible contracts are those that make the agent's compensation contingent on the verifiable performance measure. Let $s(x)$ denote the payment made to the agent under such a contract if x is realized. It will prove convenient to write s_n for $s(x_n)$ and to consider the compensation *vector* $\mathbf{s} = (s_1, \ldots, s_N)$. The optimal—expected-cost-minimizing—contract for implementing $\hat{a}$ (assuming it can be implemented) is the contract that solves the following program:[2]

$$\min_{\mathbf{s}} \mathbf{f}(\hat{a}) \cdot \mathbf{s}$$

subject to

$$\sum_{n=1}^{N} f_n(\hat{a})\, u(s_n) - c(\hat{a}) \geq U_R$$

(the IR constraint) and

$$\hat{a} \in \max_a \sum_{n=1}^{N} f_n(a)\, u(s_n) - c(a)$$

(the IC constraint—see (14.1)). Observe that an equivalent statement of the IC constraint is

$$\sum_{n=1}^{N} f_n(\hat{a})\, u(s_n) - c(\hat{a}) \geq \sum_{n=1}^{N} f_n(a)\, u(s_n) - c(a) \quad \forall a \in \mathcal{A}.$$

As we've seen above, it is often easier to work in terms of utility payments than in terms of monetary payments. Specifically, because $u(\cdot)$ is invertible,

[2] Observe, given the separability between the principal's benefit and cost, minimizing her expected wage payment is equivalent to maximizing her expected profit.

we can express a contract as an N-dimensional vector of contingent utilities, $\mathbf{u} = (u_1, \ldots, u_N)$, where $u_n = u(s_n)$. Using this "trick," the principal's program becomes

$$\min_{\mathbf{u}} \sum_{n=1}^{N} f_n(\hat{a})\, u^{-1}(u_n) \tag{15.1}$$

subject to

$$\mathbf{f}(\hat{a}) \cdot \mathbf{u} - c(\hat{a}) \geq U_R \tag{IR}$$

and

$$\mathbf{f}(\hat{a}) \cdot \mathbf{u} - c(\hat{a}) \geq \mathbf{f}(a) \cdot \mathbf{u} - c(a) \quad \forall a \in \mathcal{A}. \tag{IC}$$

Definition 7 *An action $\hat{a}$ is* implementable *if there exists at least one contract solving (*IR*) and (*IC*).*

A key result is the following:

Proposition 22 *If $\hat{a}$ is implementable, then there exists a contract that implements $\hat{a}$ and satisfies (*IR*) as an equality. Moreover, (*IR*) is met as an equality (i.e., is binding) under the optimal contract for implementing $\hat{a}$.*

Proof: Suppose not: Let $\mathbf{u}$ be a contract that implements $\hat{a}$ and suppose that

$$\mathbf{f}(\hat{a}) \cdot \mathbf{u} - c(\hat{a}) > U_R.$$

Define

$$\varepsilon = \mathbf{f}(\hat{a}) \cdot \mathbf{u} - c(\hat{a}) - U_R.$$

By assumption, $\varepsilon > 0$. Consider a new contract, $\widetilde{\mathbf{u}}$, where $\tilde{u}_n = u_n - \varepsilon$. By construction, this new contract satisfies (IR). Moreover, because

$$\mathbf{f}(a) \cdot \widetilde{\mathbf{u}} = \mathbf{f}(a) \cdot \mathbf{u} - \varepsilon,$$

for all $a \in \mathcal{A}$, this new contract also satisfies (IC). Observe, too, that this new contract is superior to $\mathbf{u}$: It satisfies the contracts, while costing the principal less. Hence, a contract cannot be optimal unless (IR) is an equality under it. ■

In light of this proposition, it follows that an action $\hat{a}$ can be implemented if there is a contract $\mathbf{u}$ that solves the following system:

$$\mathbf{f}(\hat{a}) \cdot \mathbf{u} - c(\hat{a}) = U_R \tag{15.2}$$

and

$$\mathbf{f}(a) \cdot \mathbf{u} - c(a) \leq U_R \quad \forall a \in \mathcal{A} \backslash \{\hat{a}\} \tag{15.3}$$

(where (15.3) follows from (IC) and (15.2)). We are now in position to establish the following proposition:

Proposition 23 *Action $\hat{a}$ is implementable if and only if there is no strategy for the agent that induces the same density over signals as $\hat{a}$ and which costs the agent less, in terms of expected disutility, than $\hat{a}$ (where "strategy" refers to mixed, as well as, pure strategies).*

Proof: Let $j = 1, \ldots, J-1$ index the elements in $\mathcal{A}$ *other* than $\hat{a}$. Then the system (15.2) and (15.3) can be written as $J+1$ inequalities:

$$\begin{gathered}
\mathbf{f}(\hat{a}) \cdot \mathbf{u} \leq U_R + c(\hat{a}) \\
[-\mathbf{f}(\hat{a})] \cdot \mathbf{u} \leq -U_R - c(\hat{a}) \\
\mathbf{f}(a_1) \cdot \mathbf{u} \leq U_R + c(a_1) \\
\vdots \\
\mathbf{f}(a_{J-1}) \cdot \mathbf{u} \leq U_R + c(a_{J-1})
\end{gathered}$$

By a well-known result in convex analysis (see, *e.g.,* Rockafellar, 1970, page 198), there is a $\mathbf{u}$ that solves this system if and only if there is *no* vector

$$\boldsymbol{\mu} = (\hat{\mu}_+, \hat{\mu}_-, \mu_1, \ldots, \mu_{J-1}) \geq \mathbf{0}_{J+1}$$

(where $\mathbf{0}_K$ is a K-dimensional vector of zeros) such that

$$\hat{\mu}_+ \mathbf{f}(\hat{a}) + \hat{\mu}_- [-\mathbf{f}(\hat{a})] + \sum_{j=1}^{J-1} \mu_j \mathbf{f}(a_j) = \mathbf{0}_N \tag{15.4}$$

and

$$\hat{\mu}_+ [U_R + c(\hat{a})] + \hat{\mu}_- [-U_R - c(\hat{a})] + \sum_{j=1}^{J-1} \mu_j [U_R + c(a_j)] < 0. \tag{15.5}$$

Observe that if such a μ exists, then (15.5) entails that not all elements can be zero. Define $\mu_* = \hat{\mu}_+ - \hat{\mu}_-$. By post-multiplying (15.4) by $\mathbf{1}_N$ (an N-dimensional vector of ones), we see that

$$\mu_* + \sum_{j=1}^{J-1} \mu_j = 0. \tag{15.6}$$

Equation (15.6) implies that $\mu_* < 0$. Define $\sigma_j = \mu_j / (-\mu_*)$. By construction each $\sigma_j \geq 0$ (with at least some being strictly greater than 0) and, from (15.6), $\sum_{j=1}^{J-1} \sigma_j = 1$. Hence, we can interpret these σ_j as probabilities and, thus, as a mixed strategy over the elements of $\mathcal{A} \setminus \{\hat{a}\}$. Finally, observe that if we divide both sides of (15.4) and (15.5) by $-\mu_*$ and rearrange, we can see that (15.4) and (15.5) are equivalent to

$$\mathbf{f}(\hat{a}) = \sum_{j=1}^{J-1} \sigma_j \mathbf{f}(a_j) \tag{15.7}$$

and

$$c(\hat{a}) > \sum_{j=1}^{J-1} \sigma_j c(a_j); \tag{15.8}$$

that is, there is a contract $\mathbf{u}$ that solves the above system of inequalities if and only if there is no (mixed) strategy that induces the same density over the performance measures as $\hat{a}$ (*i.e.,* (15.7)) and that has lower expected cost (*i.e.,* (15.8)). ■

The truth of the necessity condition (only if part) of Proposition 23 is straightforward: Were there such a strategy—one that always produced the same expected utility over money as $\hat{a}$, but which cost the agent less than $\hat{a}$—then it would clearly be impossible to implement $\hat{a}$ as a pure strategy. What is less obvious is the sufficiency (if part) of the proposition. Intuitively, if the density over the performance measure induced by $\hat{a}$ is distinct from the density induced by any other strategy, then the performance measure is informative with respect to determining whether $\hat{a}$ was the agent's strategy or whether he played a different strategy. Because the range of $u(\cdot)$ is unbounded, even a small amount of information can be exploited to implement $\hat{a}$ by rewarding the agent for performance that is relatively more likely to occur when he plays the strategy $\hat{a}$, or by punishing him for performance that is relatively unlikely to occur when he plays the strategy $\hat{a}$, or both .[3] Of course, even if there are other strategies that induce the same density as $\hat{a}$, $\hat{a}$ is still implementable if the agent finds these other strategies more costly than $\hat{a}$.

Before solving the principal's problem (Step 1, page 152), it's worth considering, and then dismissing, two "pathological" cases. The first is the ability to implement *least-cost actions* at their full-information cost. The second is the ability to implement any action at its full-information cost when there is a *shifting support* (of the right kind).

Definition 8 *An action $\tilde{a}$ is a* least-cost action *if $\tilde{a} \in \arg\min_{a\in\mathcal{A}} c(a)$. That is, $\tilde{a}$ is a least-cost action if the agent's disutility from choosing any other action is at least as great as his disutility from choosing $\tilde{a}$.*

Proposition 24 *If $\tilde{a}$ is a least-cost action, then it is implementable at its full-information cost.*

---

[3] We can formalize this notion of informationally distinct as follows: The condition that no strategy duplicate the density over performance measures induced by $\hat{a}$ is equivalent to saying that there is *no* density (strategy) $(\sigma_1, \dots, \sigma_{J-1})$ over the *other* $J-1$ elements of $\mathcal{A}$ such that

$$\mathbf{f}(\hat{a}) = \sum_{j=1}^{J-1} \sigma_j \mathbf{f}(a_j).$$

Mathematically, that's equivalent to saying that $\mathbf{f}(\hat{a})$ is *not* a convex combination of $\{\mathbf{f}(a)\}_{a\in\mathcal{A}\setminus\{\hat{a}\}}$; or, equivalently that $\mathbf{f}(\hat{a})$ is *not* in the *convex hull* of $\{\mathbf{f}(a)|a \neq \hat{a}\}$. See Hermalin and Katz (1991) for more on this "convex-hull" condition and its interpretation. Finally, from Proposition 23, the condition that $\mathbf{f}(\hat{a})$ *not* be in the convex hull of $\{\mathbf{f}(a)|a \neq \hat{a}\}$ is *sufficient* for $\hat{a}$ to be implementable.

Proof: Consider the fixed-payment contract that pays the agent $u_n = U_R + c(\tilde{a})$ for all n. This contract clearly satisfies (IR) and, because $c(\tilde{a}) \leq c(a)$ for all $a \in \mathcal{A}$, it also satisfies (IC). The cost of this contract to the principal is $u^{-1}[U_R + c(\tilde{a})] = C^F(\tilde{a})$, the full-information cost. ■

Of course, there is nothing surprising to Proposition 24: When the principal wishes to implement a least-cost action, her interests and the agent's are perfectly aligned; that is, there is no agency problem. Consequently, it is not surprising that the full-information outcome obtains.

Definition 9 *There is a* meaningful shifting support *associated with action $\tilde{a}$ if there exists a subset of $\mathcal{X}$, $\mathcal{X}_0$, such that $F(\mathcal{X}_0|a) > 0 = F(\mathcal{X}_0|\tilde{a})$ for* all *actions a such that $c(a) < c(\tilde{a})$.*

Proposition 25 *Let there be a meaningful shifting support associated with action $\tilde{a}$. Then action $\tilde{a}$ is implementable at its full-information cost.*

Proof: Fix some arbitrarily small $\varepsilon > 0$ and define $u_P = u(\underline{s} + \varepsilon)$. Consider the contract $\mathbf{u}$ that sets $u_m = u_P$ if $x_m \in \mathcal{X}_0$ (where $\mathcal{X}_0$ is defined above) and that sets $u_n = U_R + c(\tilde{a})$ if $x_n \notin \mathcal{X}_0$. It follows that $\mathbf{f}(\tilde{a}) \cdot \mathbf{u} = U_R + c(\tilde{a})$; that $\mathbf{f}(a) \cdot \mathbf{u} \to -\infty$ as $\varepsilon \downarrow 0$ for a such that $c(a) < c(\tilde{a})$; and that

$$\mathbf{f}(a) \cdot \mathbf{u} - c(a) \leq U_R + c(\tilde{a}) - c(a) \leq \mathbf{f}(\tilde{a}) \cdot \mathbf{u} - c(\tilde{a})$$

for a such that $c(a) \geq c(\tilde{a})$. Consequently, this contract satisfies (IR) and (IC). Moreover, the *equilibrium* cost of this contract to the principal is $u^{-1}[U_R + c(\tilde{a})]$, the full-information cost. ■

Intuitively, when there is a meaningful shifting support, observing an $x \in \mathcal{X}_0$ is proof that the agent took an action other than $\tilde{a}$. Because the principal has this proof, she can punish the agent as severely as she wishes when such an x appears (in particular, she doesn't have to worry about how this punishment changes the risk faced by the agent, given the agent is never in jeopardy of suffering this punishment if he takes the desired action, $\tilde{a}$).[4] Moreover, such a draconian punishment will deter the agent from taking an action that induces a positive probability of suffering the punishment. In effect, such actions have been dropped from the original game, leaving $\tilde{a}$ as a least-cost action of the new

[4] It is worth noting that this argument relies on being able to punish the agent sufficiently in the case of an $x \in \mathcal{X}_0$. Whether the use of such punishments is really feasible could, in some contexts, rely on assumptions that are overly strong. First, that the agent hasn't (or can contractually waive) protections against severe punishments. For example, in the English common-law tradition, this is generally not true; moreover, courts in these countries are generally loath to enforce contractual clauses that are deemed to be penalties. Second, that the agent has faith in his understanding of the distributions (*i.e.,* he is sure that taking action $\tilde{a}$ guarantees that an $x \in \mathcal{X}_0$ won't occur). Third, that the agent has faith in his own rationality; that is, in particular, he is sufficiently confident that won't make a mistake (*i.e.,* choose an a such that $F(\mathcal{X}_0|a) > 0$).

game. It follows, then, from Proposition 24, that $\tilde{a}$ can be implemented at its full-information cost.

It is worth noting that the full-information benchmark is just a special case of Proposition 25, in which the support of $\mathbf{f}(a)$ lies on a separate plane, $\mathcal{X}' \times \{a\}$, for each action a.

Typically, it is assumed that the action the principal wishes to implement is neither a least-cost action, nor has a meaningful shifting support associated with it. Henceforth, we will assume that the action that principal wishes to implement, $\hat{a}$, is not a least-cost action (*i.e.,* $\exists a \in \mathcal{A}$ such that $c(a) < c(\hat{a})$). Moreover, we will rule out all shifting supports by assuming that $f_n(a) > 0$ for all n and all a.

We now consider the whether there is a solution to Step 1 when there is no shifting support and the action to be implemented is not a least-cost action. That is, we ask the question: If $\hat{a}$ is implementable is there an optimal contract for implementing it? We divide the analysis into two cases: $u(\cdot)$ affine (risk neutral) and $u(\cdot)$ strictly concave (risk averse).

Proposition 26 *Assume $u(\cdot)$ is affine and that $\hat{a}$ is implementable. Then $\hat{a}$ is implementable at its full-information cost.*

Proof: Let $\mathbf{u}$ solve (IR) and (IC). From Proposition 22, we may assume that (IR) is binding. Then, because $u(\cdot)$ and, thus, $u^{-1}(\cdot)$ are affine:

$$\sum_{n=1}^{N} f_n(\hat{a})\, u^{-1}(u_n) = u^{-1}\left(\sum_{n=1}^{N} f_n(\hat{a})\, u_n\right) = u^{-1}\left[U_R + c(\hat{a})\right],$$

where the last inequality follows from the fact that (IR) is binding. ■

Note, given that we can't do better than implement an action at full-information cost, this proposition also tells us that, with a risk-neutral agent, an optimal contract exists for inducing any implementable action. The hidden-action problem (the lack of full information) is potentially costly to the principal for two reasons. First, it may mean a desired action is not implementable. Second, even if it is implementable, it may be implementable at a higher cost. Proposition 26 tells us that this second source of cost must be due solely to the agent's risk aversion; an insight consistent with those derived earlier.

In fact, if we're willing to assume that the principal's benefit is alienable—that is, she can sell the rights to receive it to the agent—and that the agent is risk neutral, then we can implement the optimal full-information action, a^* (*i.e.,* the solution to Step *2* under full information) at full-information cost. In other words, we can achieve the complete full-information solution in this case:

Proposition 27 (Selling the store) *Assume that $u(\cdot)$ is affine and that the principal's benefit is alienable. Then the principal can achieve the same expected utility with a hidden-action problem as she could under full information.*

Proof: Under full information, the principal would induce a^* where

$$a^* \in \arg\max_{a \in \mathcal{A}} B(a) - C^F(a).$$

Define

$$t^* = B(a^*) - C^F(a^*).$$

Suppose the principal offers to sell the right to her benefit to the agent for t^*. If the agent accepts, then the principal will enjoy the same utility she would have enjoyed under full information. Will the agent accept? Note that because $u(\cdot)$ is affine, there is no loss of generality in assuming it is the identity function. If he accepts, he faces the problem

$$\max_{a \in \mathcal{A}} \int_{\mathcal{B}} (b - t^*)\, dG(b|a) - c(a).$$

This is equivalent to

$$\max_{a \in \mathcal{A}} B(a) - c(a) - B(a^*) + c(a^*) + U_R; \text{ or to}$$
$$\max_{a \in \mathcal{A}} B(a) - [c(a) + U_R] - B(a^*) + c(a^*) + 2U_R.$$

Because $B(a) - [c(a) + U_R] = B(a) - C^F(a)$, rational play by the agent conditional on accepting means his utility will be U_R; which also means he'll accept. ■

People often dismiss the case where the agent is risk neutral by claiming that there is no agency problem because the principal could "sell the store (productive asset)" to the agent. As this last proposition makes clear, such a conclusion relies critically on the ability to literally sell the asset; that is, if the principal's benefit is not alienable, then this conclusion might not hold.[5] In other words, it is not solely the agent's risk aversion that causes problems with a hidden action.

Corollary 5 *Assume that $u(\cdot)$ is affine and that the principal's benefit equals the performance measure (i.e., $\mathcal{B} = \mathcal{X}$ and $G(\cdot|a) = F(\cdot|a)$). Then the principal can achieve the same expected utility with a hidden-action problem as she could under full information.*

Proof: Left to the reader. (*Hint:* Let $s(x) = x - t$, where t is a constant.) ■

Now we turn our attention to the case where $u(\cdot)$ is strictly concave (the agent is risk averse). Observe (i) this entails that $u^{-1}(\cdot)$ is strictly convex; (ii), because $\mathcal{S}$ is an open interval, that $u(\cdot)$ is continuous; and (iii) that $u^{-1}(\cdot)$ is continuous.

[5] To see this, suppose the benefit is *un*alienable. Assume, too, that $\mathcal{A} = \{1/4, 1/2, 3/4\}$, $\mathcal{X} = \{1, 2\}$, $c(a) = \sqrt{a}$, $f_2(a) = a$, $U_R = 0$, and $B(a) = 4 - 4\left(a - \frac{1}{2}\right)^2$. Then it is readily seen that $a^* = 1/2$. However, from Proposition 23, a^* is *not* implementable, so the full-information outcome is unobtainable when the action is hidden (even though the agent is risk neutral).

Proposition 28 *Assume that $u(\cdot)$ is strictly concave. If $\hat{a}$ is implementable, then there exists a unique contract that implements $\hat{a}$ at minimum expected cost.*

Proof: Existence.[6] Define

$$\Omega(\mathbf{u}) = \sum_{n=1}^{N} f_n(\hat{a})\, u^{-1}(u_n). \tag{15.9}$$

The strict convexity and continuity of $u^{-1}(\cdot)$ implies that $\Omega(\cdot)$ is also a strictly convex and continuous function. Observe that the principal's problem is to choose $\mathbf{u}$ to minimize $\Omega(\mathbf{u})$ subject to (IR) and (IC). Let $\mathcal{U}$ be the set of contracts that satisfy (IR) and (IC) (by assumption, $\mathcal{U}$ is not empty). Were $\mathcal{U}$ closed and bounded, then a solution to the principal's problem would certainly exist because $\Omega(\cdot)$ is a continuous real-valued function.[7] Unfortunately, $\mathcal{U}$ is not bounded (although it is closed given that all the inequalities in (IR) and (IC) are weak inequalities). Fortunately, we can artificially bound $\mathcal{U}$ by showing that any solution outside some bound is inferior to a solution inside the bound. Consider any contract $\mathbf{u}^0 \in \mathcal{U}$ and consider the contract $\mathbf{u}^*$, where $u_n^* = U_R + c(\hat{a})$. Let $\mathcal{U}^{IR}$ be the set of contracts that satisfy (IR). Note that $\mathcal{U} \subset \mathcal{U}^{IR}$. Note, too, that both $\mathcal{U}$ and $\mathcal{U}^{IR}$ are convex sets. Because $\Omega(\cdot)$ has a minimum on $\mathcal{U}^{IR}$, namely $\mathbf{u}^*$, the set

$$\mathcal{V} \equiv \left\{\mathbf{u} \in \mathcal{U}^{IR} | \Omega(\mathbf{u}) \le \Omega\left(\mathbf{u}^0\right)\right\}$$

is closed, bounded, and convex.[8] By construction, $\mathcal{U} \cap \mathcal{V}$ is non-empty; moreover, for any $\mathbf{u}^1 \in \mathcal{U} \cap \mathcal{V}$ and any $\mathbf{u}^2 \in \mathcal{U} \backslash \mathcal{V}$, $\Omega\left(\mathbf{u}^2\right) > \Omega\left(\mathbf{u}^1\right)$. Consequently, nothing is lost be limiting the search for an optimal contract to $\mathcal{U} \cap \mathcal{V}$. The set $\mathcal{U} \cap \mathcal{V}$ is closed and bounded and $\Omega(\cdot)$ is continuous, hence it follows that an optimal contract must exist.

Uniqueness. Suppose the optimal contract, $\mathbf{u}$, were not unique. That is, there exists another contract $\tilde{\mathbf{u}}$ such that $\Omega(\mathbf{u}) = \Omega(\tilde{\mathbf{u}})$ (where $\Omega(\cdot)$ is defined by (15.9)). It is readily seen that if these two contracts each satisfy both the (IR) and (IC) constraints, then any convex combination of them must as well (*i.e.,* both are elements of $\mathcal{U}$, which is convex). That is, the contract

$$\mathbf{u}_\lambda \equiv \lambda \mathbf{u} + (1-\lambda)\tilde{\mathbf{u}},$$

$\lambda \in (0,1)$ must be feasible (*i.e.,* satisfy (IR) and (IC)). Because $\Omega(\cdot)$ is strictly convex, Jensen's inequality implies

$$\Omega(\mathbf{u}_\lambda) < \lambda\Omega(\mathbf{u}) + (1-\lambda)\Omega(\tilde{\mathbf{u}}) = \Omega(\mathbf{u}).$$

---

[6] The existence portion of this proof is somewhat involved mathematically and can be omitted without affecting later comprehension of the material.

[7] This is a well-known result from analysis (see, *e.g.,* Fleming, 1977, page 49).

[8] The convexity of $\mathcal{V}$ follows because $\Omega(\cdot)$ is a convex function and $\mathcal{U}^{IR}$ is a convex set. That $\mathcal{V}$ is closed follows given $\mathcal{U}^{IR}$ is also closed. To see that $\mathcal{V}$ is bounded, recognize that, as one "moves away" from $\mathbf{u}^*$—while staying in $\mathcal{U}^{IR}$—$\Omega(\mathbf{u})$ increases. Because $\Omega(\cdot)$ is convex, any such movement away from $\mathbf{u}^*$ must eventually (*i.e.,* for finite $\mathbf{u}$) lead to a $\Omega(\mathbf{u}) > \Omega\left(\mathbf{u}^0\right)$ (convex functions are unbounded above). Hence $\mathcal{V}$ is bounded.

But this contradicts the optimality of $\mathbf{u}$. By contradiction, uniqueness is established. ■

Having concluded that a solution to Step 1 exists, we can—at last—calculate what it is. From Proposition 26, the problem is trivial if $u(\cdot)$ is affine, so we will consider only the case in which $u(\cdot)$ is strictly concave. The principal's problem is a standard nonlinear programming problem: Minimize a convex function (*i.e.,* $\sum_{n=1}^{N} f_n(\hat{a})\, u^{-1}(u_n)$) subject to J constraints (one individual rationality constraint and $J-1$ incentive compatibility constraints, one for each action other than $\hat{a}$). If we further assume, as we do henceforth, that $u(\cdot)$ is differentiable, then the standard Lagrange-multiplier techniques can be employed. Specifically, let λ be the Lagrange multiplier on the IR constraint and let μ_j be the Lagrange multiplier on the IC constraint between $\hat{a}$ and a_j, where $j = 1, \ldots, J-1$ indexes the elements of $\mathcal{A}$ other than $\hat{a}$. It is readily seen that the first-order condition with respect to the contract are

$$\frac{f_n(\hat{a})}{u'\left[u^{-1}(u_n)\right]} - \lambda f_n(\hat{a}) - \sum_{j=1}^{J-1} \mu_j \left[f_n(\hat{a}) - f_n(a_j)\right] = 0;\ n = 1, \ldots, N.$$

We've already seen (Proposition 22) that the IR constraint binds, hence $\lambda > 0$. Because $\hat{a}$ is not a least-cost action and there is no shifting support, it is readily shown that at least one IC constraint binds (*i.e.,* $\exists j$ such that $\mu_j > 0$). It's convenient to rewrite the first-order condition as

$$\frac{1}{u'\left[u^{-1}(u_n)\right]} = \lambda + \sum_{j=1}^{J-1} \mu_j \left(1 - \frac{f_n(a_j)}{f_n(\hat{a})}\right);\ n = 1, \ldots, N. \tag{15.10}$$

Note the resemblance between (15.10) and (13.9) in Section 13.4. The difference is that, now, we have more than one Lagrange multiplier on the actions (as we now have more than two actions). In particular, we can give a similar interpretation to the likelihood ratios, $f_n(a_j)/f_n(\hat{a})$, that we had in that earlier section; with the caveat that we now must consider more than one action.

15.2 Properties of the Optimal Contract

Having solved for the optimal contract, we can now examine its properties. In particular, we will consider three questions:

1. Under what conditions does the expected cost of implementing an action under the optimal contract for the hidden-action problem exceed the full-information cost of implementing that action?
2. Recall the performance measures, x, constitute distinct elements of a chain. Under what conditions is the agent's compensation increasing with the value of the signal (*i.e.,* when does $x \prec x'$ imply $s(x) \leq s(x')$)?

3. Consider two principal-agent models that are identical except that the information structure (*i.e.*, $\{\mathbf{f}(a)\,|a \in \mathcal{A}\}$) in one is more informative than the information structure in the other. How do the costs of implementing actions vary between these two models.

The answer to the first question is given by

Proposition 29 *Consider a hidden-information problem. Assume there is no shifting support (i.e., $f_n(a) > 0$ for all n and all a). Assume, too, that $u(\cdot)$ is strictly concave. If $\hat{a}$ is not a least-cost action, then it cannot be implemented at its full-information cost.*

Proof: If $\hat{a}$ is not implementable, then the result is obvious; hence, we'll assume $\hat{a}$ is implementable. Define $\mathcal{U}^{IR}$ to be the set of all contracts that satisfy the IR constraint for $\hat{a}$. Let $\mathbf{u}^*$ be the contract in which $u_n^* = U_R + c(\hat{a})$ for all n. Note $\mathbf{u}^* \in \mathcal{U}^{IR}$. Finally define,

$$\Omega(\mathbf{u}) = \sum_{n=1}^{N} f_n(\hat{a})\, u^{-1}(u_n)\,.$$

Because $u(\cdot)$ is strictly concave, the principal's expected cost if the agent chooses $\hat{a}$ under contract $\mathbf{u}$, $\Omega(\mathbf{u})$, is a strictly convex function of $\mathbf{u}$. By Jensen's inequality and the fact that there is no shifting support, $\Omega(\cdot)$, therefore, has a *unique* minimum in $\mathcal{U}^{IR}$, namely $\mathbf{u}^*$. Clearly, $\Omega(\mathbf{u}^*) = C^F(\hat{a})$. The result, then, follows if we can show that $\mathbf{u}^*$ is not incentive compatible. Given that $\hat{a}$ is not a least-cost action, there exists an a such that $c(\hat{a}) > c(a)$. But

$$\mathbf{f}(a) \cdot \mathbf{u}^* - c(a) = U_R + c(\hat{a}) - c(a) > U_R = \mathbf{f}(\hat{a}) \cdot \mathbf{u}^* - c(\hat{a})\,;$$

that is, $\mathbf{u}^*$ is not incentive compatible. ■

Note the elements that go into this proposition *if $\hat{a}$ is implementable:* There must be an agency problem—mis-alignment of interests (*i.e.,* $\hat{a}$ is not least cost); there must, in fact, be a *significant* hidden-action problem (*i.e.,* no shifting support); and the agent must be risk averse. We saw earlier that without any one of these elements, an implementable action is implementable at full-information cost (Propositions 24–26); that is, each element is individually necessary for cost to increase when we go from full information to a hidden action. This last proposition shows, *inter alia*, that they are collectively sufficient for the cost to increase.

Next we turn to the second question. We already know from our analysis of the two-action model that the assumptions we have so far made are *in*sufficient for us to conclude that compensation will be monotonic. From our analysis of that model, we might expect that we need some monotone likelihood ratio property. In particular, we assume

MLRP: Assume there is no shifting support. Then the monotone likelihood ratio property is said to hold if, for any a and $\hat{a}$ in $\mathcal{A}$, $c(a) \leq c(\hat{a})$ implies that $f_n(a)/f_n(\hat{a})$ is nonincreasing in n.

Intuitively, MLRP is the condition that actions that the agent finds more costly be more likely to produce better performance.

Unlike the two-action case, however, MLRP is not sufficient for us to obtain monotone compensation (see Grossman and Hart, 1983, for an example in which MLRP is satisfied but compensation is non-monotone). We need an additional assumption:

CDFP: The agency problem satisfies the *concavity of distribution function property* if, for any a, a', and $\hat{a}$ in $\mathcal{A}$,

$$c(\hat{a}) = \lambda c(a) + (1-\lambda) c(a') \quad \exists \lambda \in (0,1)$$

implies that $F(\cdot|\hat{a})$ first-order stochastically dominates $\lambda F(\cdot|a) + (1-\lambda) F(\cdot|a')$.[9]

Another way to state the CDFP is that the distribution over performance is better—more likely to produce high signals—if the agent plays a pure strategy than it is if he plays any mixed strategy over two actions when that mixed strategy has the same expected disutility as the pure strategy.

We can now answer the second question:

Proposition 30 *Assume there is no shifting support, that $u(\cdot)$ is strictly concave and differentiable, and that* MLRP *and* CDFP *are met. Then the optimal contract given the hidden-action problem satisfies $s_1 \leq \cdots \leq s_N$.*

Proof: Let $\hat{a}$ be the action the principal wishes to implement. If $\hat{a}$ is a least-cost action, then the result follows from Proposition 24; hence assume that $\hat{a}$ is not a least-cost action. Let $\mathcal{A}' = \{a | c(a) \leq c(\hat{a})\}$; that is, $\mathcal{A}'$ is the set of actions that cause the agent less disutility than $\hat{a}$. Consider the principal's problem of implementing $\hat{a}$ under the assumption that the space of contracts is $\mathcal{A}'$. By MLRP, $f_n(a)/f_n(\hat{a})$ is nonincreasing in n for all $a \in \mathcal{A}'$, so it follows from (15.10) that $s_1 \leq \cdots \leq s_N$ under the optimal contract for this *restricted* problem. The result then follows if we can show that this contract remains optimal when we expand $\mathcal{A}'$ to $\mathcal{A}$—adding actions cannot reduce the cost of implementing $\hat{a}$, hence we are done if we can show that the optimal contract for the restricted problem is incentive compatible in the *un*restricted problem. That is, if there is *no* a, $c(a) > c(\hat{a})$, such that

$$\mathbf{f}(a) \cdot \mathbf{u} - c(a) > \mathbf{f}(\hat{a}) \cdot \mathbf{u} - c(\hat{a}), \tag{15.11}$$

where $\mathbf{u} = (u(s_1), \ldots, u(s_N))$. As demonstrated in the proof of Proposition 29, the incentive compatibility constraint between $\hat{a}$ and at least one $a' \in \mathcal{A}'$, $c(a') < c(\hat{a})$, is binding; *i.e.,*

$$\mathbf{f}(a') \cdot \mathbf{u} - c(a') = \mathbf{f}(\hat{a}) \cdot \mathbf{u} - c(\hat{a}). \tag{15.12}$$

[9] Recall that distribution $G(\cdot)$ *first-order stochastically dominates* distribution $H(\cdot)$ if $G(z) \leq H(z)$ for all z and strictly less than for some z.

Because $c(\hat{a}) \in (c(a'), c(a))$, there exists a $\lambda \in (0,1)$ such that $c(\hat{a}) = (1-\lambda)\,c(a') + \lambda c(a)$. Using CDFP and the fact that $u(s_1) \leq \cdots \leq u(s_N)$, we have

$$\begin{aligned} \mathbf{f}(\hat{a}) \cdot \mathbf{u} - c(\hat{a}) &\geq (1-\lambda)\,\mathbf{f}(a') \cdot \mathbf{u} + \lambda \mathbf{f}(a) \cdot \mathbf{u} - c(\hat{a}) \\ &= (1-\lambda)\left[\mathbf{f}(a') \cdot \mathbf{u} - c(a')\right] + \lambda\left[\mathbf{f}(a) \cdot \mathbf{u} - c(a)\right]. \end{aligned}$$

But this and (15.12) are inconsistent with (15.11); that is, (15.11) cannot hold, as was required. ■

Lastly, we come to question 3. An *information structure* for a principal-agent problem is $\mathbf{F} \equiv \{\mathbf{f}(a) \,|\, a \in \mathcal{A}\}$. A principal-agent problem can, then, be summarized as $\mathfrak{P} = \langle \mathcal{A}, \mathcal{X}, \mathbf{F}, B(\cdot), c(\cdot), u(\cdot), U_R \rangle$.

Proposition 31 *Consider two principal-agent problems that are identical except for their information structures (i.e., consider*

$$\mathfrak{P}^1 = \left\langle \mathcal{A}, \mathcal{X}, \mathbf{F}^1, B(\cdot), c(\cdot), u(\cdot), U_R \right\rangle$$

and

$$\mathfrak{P}^2 = \left\langle \mathcal{A}, \mathcal{X}, \mathbf{F}^2, B(\cdot), c(\cdot), u(\cdot), U_R \right\rangle\text{)}.$$

Suppose there exists a stochastic transformation matrix $\mathbf{Q}$ *(i.e., a* garbling*)*[10] *such that* $\mathbf{f}^2(a) = \mathbf{Q}\mathbf{f}^1(a)$ *for all* $a \in \mathcal{A}$*, where* $\mathbf{f}^i(a)$ *denotes an element of* $\mathbf{F}^i$*. Then, for all* $a \in \mathcal{A}$*, the principal's expected cost of optimally implementing action* a *in the first principal-agent problem,* $\mathfrak{P}^1$*, is* not *greater than her expected cost of optimally implementing* a *in the second principal-agent problem,* $\mathfrak{P}^2$*.*

Proof: Fix a. If a is not implementable in $\mathfrak{P}^2$, then the result follows immediately. Suppose, then, that a is implementable in $\mathfrak{P}^2$ and let $\mathbf{u}^2$ be the optimal contract for implementing a in that problem. Consider the contract $\mathbf{u}^1 = \mathbf{Q}^\top \mathbf{u}^2$.[11] We will show that $\mathbf{u}^1$ implements a in $\mathfrak{P}^1$. Because

$$\mathbf{f}^1(a)^\top \mathbf{u}^1 = \mathbf{f}^1(a)^\top \mathbf{Q}^\top \mathbf{u}^2 = \mathbf{f}^2(a)^\top \mathbf{u}^2,$$

the fact that $\mathbf{u}^2$ satisfies IR and IC in $\mathfrak{P}^2$ can readily be shown to imply that $\mathbf{u}^1$ satisfies IR and IC in $\mathfrak{P}^1$. The principal's cost of *optimally* implementing a in $\mathfrak{P}^1$ is no greater than her cost of implementing a in $\mathfrak{P}^1$ using $\mathbf{u}^1$. By construction, $u_n^1 = \mathbf{q}_{\cdot n}^\top \mathbf{u}^2$, where $\mathbf{q}_{\cdot n}$ is the nth column of $\mathbf{Q}$. Because $s_n^i = u^{-1}\left(u_n^i\right)$ and $u^{-1}(\cdot)$ is convex, it follows from Jensen's Inequality that

$$s_n^1 \leq \sum_{m=1}^{N} q_{mn} s_m^2$$

[10] A stochastic transformation matrix, sometimes referred to as a garbling, is a matrix in which each column is a probability density (*i.e.,* has non-negative elements that sum to one).

[11] For this proof it is necessary to distinguish between row vectors and column vectors, as well as transposes of matrices. All vectors should be assumed to be column vectors. To make a vector $\mathbf{x}$ a row vector, we write $\mathbf{x}^\top$. Observe that $\mathbf{H}^\top$, where $\mathbf{H}$ is a matrix, is the transpose of $\mathbf{H}$. Observe that $\mathbf{x}^\top \mathbf{y}$ is the dot-product of $\mathbf{x}$ and $\mathbf{y}$ (what we've been writing as $\mathbf{x} \cdot \mathbf{y}$).

(recall $\mathbf{q}_{\cdot n}$ is a probability vector). Consequently,

$$\sum_{n=1}^{N} f_n^1(a)\, s_n^1 \leq \sum_{n=1}^{N} f_n^1(a) \sum_{m=1}^{N} q_{mn} s_m^2 = \sum_{m=1}^{N} f_n^2(a)\, s_n^2.$$

The result follows. ■

Proposition 31 states that if two principal-agent problems are the same, except that they have different information structures, where the information structure of the first problem is more informative than the information structure of the second problem (in the sense of Blackwell's Theorem), then the principal's expected cost of optimally implementing any action is no greater in the first problem than in the second problem. By strengthening the assumptions slightly, we can, in fact, conclude that the principal's expected cost is strictly less in the first problem. In other words, making the signal more informative about the agent's action makes the principal better off. This is consistent with our earlier findings that (i) the value of the performance measures is solely their statistical properties as correlates of the agent's action; and (ii) the better correlates—technically, the more informative—they are, the lower the cost of the hidden-action problem.

It is worth observing that Proposition 31 implies that the optimal incentive scheme never entails paying the agent with lotteries over money (*i.e.,* randomly mapping the realized performance levels via weights $\mathbf{Q}$ into payments).

15.3 A Continuous Performance Measure

Suppose that $\mathcal{X}$ were a real interval—which, without loss of generality, we can take to be $\mathbb{R}$—rather than a discrete space and suppose, too, that $F(x|a)$ were a continuous and differentiable function with corresponding probability density function $f(x|a)$. How would this change our analysis? By one measure, the answer is not much. Only three of our proofs rely on the assumption that $\mathcal{X}$ is finite; namely the proofs of Propositions 23, 28 (and, there, only the existence part),[12] and 31. Moreover, the last of the three can fairly readily be extended to the continuous case. Admittedly, it is troubling not to have *general* conditions for implementability and existence of an optimal contract, but in many specific situations we can, nevertheless, determine the optimal contract.[13]

---

[12] Where our existence proof "falls down" when $\mathcal{X}$ is continuous is that our proof relies on the fact that a continuous function from $\mathbb{R}^N \to \mathbb{R}$ has a minimum on a closed and bounded set. But, here, the contract space is no longer a subset of $\mathbb{R}^N$, but rather the space of all *functions* from $\mathcal{X} \to \mathbb{R}$; and there is no general result guaranteeing the existence of a minimum in this case.

[13] Page (1987) considers conditions for existence in this case (actually he also allows for $\mathcal{A}$ to be a continuous space). Most of the assumptions are technical, but not likely to be considered controversial. Arguably a problematic assumption in Page is that the space of

With $\mathcal{X} = \mathbb{R}$, the principal's problem—the equivalent of (15.1)—becomes

$$\min_{u(x)} \int_{-\infty}^{\infty} u^{-1}\left[u\left(x\right)\right] f\left(x|\hat{a}\right) dx$$

subject to

$$\int_{-\infty}^{\infty} u\left(x\right) f\left(x|\hat{a}\right) dx - c\left(\hat{a}\right) \geq U_R; \text{ and}$$
$$\int_{-\infty}^{\infty} u\left(x\right) f\left(x|\hat{a}\right) dx - c\left(\hat{a}\right) \geq \int_{-\infty}^{\infty} u\left(x\right) f\left(x|a\right) dx - c\left(a\right) \quad \forall a \in \mathcal{A}.$$

We know the problem is trivial if there is a shifting support, so assume the support of x, $\operatorname{supp}\{x\}$, is invariant with respect to a.[14] Assuming an optimal contract exists to implement $\hat{a}$, that contract must satisfy the modified Borch sharing rule:

$$\frac{1}{u'\left[u^{-1}\left(u\left(x\right)\right)\right]} = \lambda + \sum_{j=1}^{J-1} \mu_j \left[1 - \frac{f\left(x|a_j\right)}{f\left(x|\hat{a}\right)}\right] \text{ for almost every } x \in \operatorname{supp}\{x\}.$$

Observe that this is just a variation on (13.9) or (15.10).

Bibliographic Note

Much of the analysis in this section has been drawn from Grossman and Hart (1983). In particular, they deserve credit for Propositions 22, 26, and 28–31 (although, here and there, we've made slight modifications to the statements or proofs). Proposition 23 is based on Hermalin and Katz (1991). The rest of the analysis represent well-known results.

possible contracts is constrained; that is, assumptions are imposed on an *endogenous* feature of the model, the contracts. In particular, if $\mathfrak{S}$ is the space of permitted contracts, then there exist L and $M \in \mathbb{R}$ such that $L \leq s\left(x\right) \leq M$ for all $s\left(\cdot\right) \in \mathfrak{S}$ and all $x \in \mathcal{X}$. Moreover, $\mathfrak{S}$ is closed under the topology of pointwise convergence. On the other hand, it could be argued that range of real-life contracts must be bounded: Legal and other constraints on what payments the parties can make effectively limit the space of contracts to some set of bounded functions.

[14] That is,

$$\{x|f\left(x|a\right) > 0\} = \left\{x|f\left(x|a'\right) > 0\right\}$$

for all a and a' in $\mathcal{A}$.

Continuous Action Space | 16

So far, we've limited attention to finite action spaces. Realistic though this may be, it can serve to limit the tractability of many models, particularly when we need to assume the action space is large. A large action space can be problematic for two, related, reasons. First, under the two-step approach, we are obligated to solve for the optimal contract for *each* $a \in \mathcal{A}$ (or at least each $a \in \mathcal{A}^I$) then, letting $C(a)$ be the expected cost of inducing action a under its corresponding optimal contact, we next maximize $B(a) - C(a)$—expected benefit net expected cost. If $\mathcal{A}$ is large, then this is clearly a time-consuming and potentially impractical method for solving the principal-agent problem. The second reason a large action space can be impractical is because it can mean many constraints in the optimization program involved with finding the optimal contract for a given action (recall, *e.g.,* that we had $J-1$ constraints—one for each action other than the given action). Again, this raises issues about the practicality of solving the problem.

These problems suggest that we would like a technique that allows us to solve program (14.3) on page 149,

$$\max_{(S(\cdot),a)} \int_{\mathcal{X}} W(S(x),x,b)dG(b,x|a) \tag{16.1}$$

subject to

$$a \in \arg\max_{a'} \int_{\mathcal{X}} U(S(x),x,a')dF(x|a') \tag{16.2}$$

and

$$\max_{a'} \int_{\mathcal{X}} U(S(x),x,a')dF(x|a') \geq U_R,$$

directly, in a one-step procedure. Generally, to make such a maximization program tractable, we would take $\mathcal{A}$ to be a compact and continuous space (*e.g.,* a closed interval on $\mathbb{R}$), and employ standard programming techniques. A number of complications arise, however, if we take such an approach.

Most of these complications have to do with how we treat the IC constraint, expression (16.2). To make life simpler, suppose that $\mathcal{A} = [\underline{a}, \bar{a}] \subset \mathbb{R}$, $\mathcal{X} = \mathbb{R}$, that $F(\cdot|a)$ is differentiable and, moreover, that the expression in (16.2) is itself differentiable for all $a \in \mathcal{A}$. Then, a natural approach would be to observe that *if* $a \in (\underline{a}, \bar{a})$ maximizes that expression, it must necessarily be the solution to the first-order condition to (16.2):

$$\int_{\mathcal{X}} \left(U_a\left[S(x),x,a\right]f(x|a) + U\left[S(x),x,a\right]f_a(x|a)\right)dx = 0 \tag{16.3}$$

(where subscripts denote partial derivatives). Conversely, if we knew that the *second*-order condition was also met, (16.3) would be equivalent to (16.2) and we could use it instead of (16.2)—*at least locally.* Unhappily, we don't, in general, know (i) that the second-order conditions is met and (ii) that, even if it is, the a solving (16.3) is a global rather than merely local maximum. For many modeling problems in economics, we would avoid these headaches by simply assuming that (16.2) is globally strictly concave in a, which would ensure both the second-order condition and the fact that an a solving (16.3) is a global maximum. We can't, however, do that here: The concavity of (16.2) will, in general, depend on $S(x)$; but since $S(\cdot)$ is *endogenous*, we can't make assumptions about it. If, then, we want to substitute (16.3) for (16.2), we need to look for other ways to ensure that (16.3) describes a global maximum.

An additional complication arises with whether (16.1) also satisfies the properties that would allow us to conclude from first-order conditions that a global maximum has, indeed, been reached. Fortunately, in many problems, this issue is less severe because we typically impose the functional form

$$W(S(x), x, b) = x - S(x),$$

which gives the problem sufficient structure to allow us to validate a "first-order approach."

In the rest of this section, we develop a simple model in which a first-order approach is valid.

16.1 The First-order Approach with a Spanning Condition

Assume, henceforth, that $\mathcal{A} = [\underline{a}, \bar{a}] \subset \mathbb{R}$, $\mathcal{X} = [\underline{x}, \bar{x}] \subset \mathbb{R}$, and that $F(\cdot|a)$ is differentiable. Let $f(\cdot|a)$ be the associated probability density function for each $a \in \mathcal{A}$. We further assume that

1. $U(S(x), x, a) = u[S(x)] - a$;
2. $u(\cdot)$ is strictly increasing and strictly concave;
3. the domain of $u(\cdot)$ is $(\underline{s}, \infty)$, $\lim_{s\downarrow\underline{s}} u(s) = -\infty$, and $\lim_{s\uparrow\infty} u(s) = \infty$;
4. $f(x|a) > 0$ for all $x \in \mathcal{X}$ and for all $a \in \mathcal{A}$ (*i.e.,* there is no shifting support);
5. $F(x|a) = \gamma(a) F_H(x) + (1-\gamma(a)) F_L(x)$ and $f(x|a) = \gamma(a) f_H(x) + (1-\gamma(a)) f_L(x)$ for all x and a, where $\gamma : \mathcal{A} \to [0,1]$ and $F_H(\cdot)$ and $F_L(\cdot)$ are distribution functions on $\mathcal{X}$;
6. $\gamma(\cdot)$ is strictly increasing, strictly concave, and twice differentiable; and
7. $f_L(x)/f_H(x)$ satisfies the MLRP (*i.e.,* $f_L(x)/f_H(x)$ is non-increasing in x and there exist x' and x'' in $\mathcal{X}$, $x' < x''$, such that $f_L(x')/f_H(x') > f_L(x'')/f_H(x'')$).

Observe Assumptions 5–7 allow us, *inter alia*, to assume that $c(a) = a$ without loss of generality. Assumption 5 is known as a *spanning condition*.

In what follows, the following result will be critical:

Lemma 9 *$F_H(\cdot)$ dominates $F_L(\cdot)$ in the sense of first-order stochastic dominance.*

Proof: We need to show $F_H(x) \leq F_L(x)$ for all $x \in \mathcal{X}$ and strictly less for some x. To this end, define

$$\Delta(z) = \int_{\underline{x}}^{z} [f_H(x) - f_L(x)]\, dx.$$

We wish to show that $\Delta(z) \leq 0$ for all $z \in \mathcal{X}$ and strictly less for some z. Observe that

$$\Delta(z) = \int_{\underline{x}}^{z} \left[1 - \frac{f_L(x)}{f_H(x)}\right] f_H(x)\, dx.$$

Let

$$\delta(x) = 1 - \frac{f_L(x)}{f_H(x)}.$$

By MLRP, $\delta(\cdot)$ is non-decreasing everywhere and increases at one x at least. Because $\Delta(\bar{x}) = 0$, $f_H(\cdot) > 0$, $\bar{x} - \underline{x} > 0$, and $\delta(\cdot)$ is not constant, it follows that $\delta(\cdot)$ must be negative on some sub-interval of $\mathcal{X}$ and positive on some other. Because $\delta(\cdot)$ is non-decreasing, there must, therefore, exist an $\hat{x} \in (\underline{x}, \bar{x})$ such that

$\delta(x) \leq 0$ for all $x < \hat{x}$ (and strictly less for $\underline{x} < x < x' \leq \hat{x}$, for some x'); and
$\delta(x) \geq 0$ for all $x > \hat{x}$ (and strictly greater for $\bar{x} > x > x'' \geq \hat{x}$, for some x'').

For $z \leq \hat{x}$, this implies that $\Delta(z) < 0$—it is the integral of a quantity that is negative over some range of the integral and never positive anywhere on that range. Finally, consider, $z \in (\hat{x}, \bar{x})$. $\Delta'(z) = \delta(z) f_H(z) \geq 0$ for all z in that interval. Hence, because $\Delta(\hat{x}) < 0$ and $\Delta(\bar{x}) = 0$, it must be that $\Delta(z) \leq 0$ for all $z \in (\hat{x}, \bar{x})$. We've just shown that $\Delta(z) \leq 0$ for all $z \in \mathcal{X}$ and strictly less for some z, which yields the result. ■

A consequence of this Lemma is that if $\phi(\cdot)$ is an increasing function, then

$$\int_{\underline{x}}^{\bar{x}} \phi(x) [f_H(x) - f_L(x)]\, dx > 0. \tag{16.4}$$

It follows, then, that if $S(\cdot)$ (and, thus, $u[S(x)]$) is increasing, then (16.2) is globally concave in a. To see this, observe

$$\int_{\mathcal{X}} U(S(x), x, a) dF(x|a) = \int_{\underline{x}}^{\bar{x}} u[S(x)] [\gamma(a) f_H(x) + (1 - \gamma(a)) f_L(x)]\, dx - a;$$

so we have

$$\frac{d}{da}\int_{\mathcal{X}} U(S(x),x,a)dF(x|a) = \int_{\underline{x}}^{\bar{x}} u\left[S\left(x\right)\right]\left[f_H\left(x\right)-f_L\left(x\right)\right]\gamma'\left(a\right)dx > 0$$

by (16.4) and the assumption that $\gamma'(\cdot) > 0$. Moreover,

$$\frac{d^2}{da^2}\int_{\mathcal{X}} U(S(x),x,a)dF(x|a) = \int_{\underline{x}}^{\bar{x}} u\left[S\left(x\right)\right]\left[f_H\left(x\right)-f_L\left(x\right)\right]\gamma''\left(a\right)dx < 0$$

by (16.4) and the assumption that $\gamma''(\cdot) < 0$. To summarize:

Corollary 6 *If $S(\cdot)$ is increasing, then the agent's choice-of-action problem is globally concave. That is, we're free to substitute*

$$\int_{\underline{x}}^{\bar{x}} u\left[S\left(x\right)\right]\left[f_H\left(x\right)-f_L\left(x\right)\right]\gamma'\left(a\right)dx = 0 \tag{16.5}$$

for (16.2).

We'll now proceed as follows. We'll suppose that $S(\cdot)$ is increasing and we'll solve the principal's problem. Of course, when we're done, we'll have to double check that our solution indeed yields an increasing $S(\cdot)$. It will, but if it didn't, then our approach would be invalid. The principal's problem is

$$\max_{S(\cdot),a}\int_{\underline{x}}^{\bar{x}}\left[x-S\left(x\right)\right]f\left(x|a\right)dx$$

subject to (16.5) and the IR constraint,

$$\int_{\underline{x}}^{\bar{x}} u\left[S\left(x\right)\right]f\left(x|a\right)dx - a \geq U_R.$$

As we've shown many times now, this last constraint must be binding; so we have a classic constrained optimization program. Letting λ be the Lagrange multiplier on the IR constraint and letting μ be the Lagrange multiplier on (16.5), we obtain the first-order conditions:

$$-f\left(x|a\right)+\mu u'\left[S\left(x\right)\right]\left[f_H\left(x\right)-f_L\left(x\right)\right]\gamma'\left(a\right)+\lambda u'\left[S\left(x\right)\right]f\left(x|a\right)=0$$

differentiating by $S(x)$; and

$$\begin{aligned}&\int_{\underline{x}}^{\bar{x}}\left[x-S\left(x\right)\right]\left[f_H\left(x\right)-f_L\left(x\right)\right]\gamma'\left(a\right)dx\\&+\mu\int_{\underline{x}}^{\bar{x}} u\left[S\left(x\right)\right]\left[f_H\left(x\right)-f_L\left(x\right)\right]\gamma''\left(a\right)dx\\=\ &0\end{aligned} \tag{16.6}$$

differentiating by a (there's no λ expression in the second condition because, by (16.5), the derivative of the IR constraint with respect to a is zero). We can rearrange the first condition into our familiar modified Borch sharing rule:

$$\begin{aligned}\frac{1}{u'[S(x)]} &= \lambda + \mu\frac{[f_H(x) - f_L(x)]\,\gamma'(a)}{\gamma(a)\,f_H(x) + (1-\gamma(a))\,f_L(x)} \\ &= \lambda + \mu\frac{[1 - r(x)]\,\gamma'(a)}{\gamma(a)\,[1 - r(x)] + r(x)},\end{aligned}$$

where $r(x) = f_L(x)/f_H(x)$. Recall that $1/u'[\cdot]$ is an increasing function; hence, to test whether $S(\cdot)$ is indeed increasing, we need to see whether the right-hand side is *decreasing* in $r(x)$ since $r(\cdot)$ is decreasing. Straightforward calculations reveal that the derivative of the right-hand side with respect to $r(x)$ is

$$\frac{-\gamma'(a)}{(r(x) + (1 - r(x))\,\gamma(a))^2} < 0.$$

We've therefore shown that $S(\cdot)$ is indeed increasing as required; that is, our use of (16.5) for (16.2) was valid.

Observe, from (16.6), that, because the *agent's* second-order condition is met, the first line in (16.6) must be positive; that is,

$$\int_{\underline{x}}^{\bar{x}} [x - S(x)]\,[f_H(x) - f_L(x)]\,dx > 0.$$

But this implies that, for this $S(\cdot)$, the *principal's* problem is globally concave in a:

$$\frac{d^2}{da^2}\int_{\underline{x}}^{\bar{x}} [x - S(x)]\,f(x|a)\,dx = \int_{\underline{x}}^{\bar{x}} [x - S(x)]\,[f_H(x) - f_L(x)]\,\gamma''(a)\,dx < 0.$$

Moreover, for any $S(\cdot)$, the principal's problem is (trivially) concave in $S(\cdot)$. Hence, we may conclude that the first-order approach is, indeed, valid for this problem.

Admittedly, the spanning condition is a fairly stringent condition; although it does have an economic interpretation. Suppose there are two distributions from which the performance measure could be drawn, "favorable" (*i.e.,* $F_H(\cdot)$) and "unfavorable" (*i.e.,* $F_L(\cdot)$). The harder—higher a—the agent chooses, the greater the probability, $\gamma(a)$, that the performance measure will be drawn from the favorable distribution. For instance, suppose there are two types of potential customers, those who tend to buy a lot—the H type—and those who tend not to buy much—the L type. By investing more effort, a, in learning his territory, the salesperson (agent) increases the probability that he will sell to H types rather than L types.

Bibliographic Note

The first papers to use the first-order approach were Holmström (1979) and Shavell (1979). Grossman and Hart (1983) was, in large part, a response to the potential invalidity of the first-order approach. Our analysis under the spanning condition draws, in part, from Hart and Holmström (1987).

Bibliography

Borch, Karl H., *The Economics of Uncertainty*, Princeton, NJ: Princeton University Press, 1968.

Caillaud, Bernard and Benjamin E. Hermalin, "The Use of an Agent in a Signalling Model," *Journal of Economic Theory*, June 1993, *60* (1), 83–113.

_ , Roger Guesnerie, Patrick Rey, and Jean Tirole, "Government Intervention in Production and Incentives Theory: A Review of Recent Contributions," *RAND Journal of Economics*, Spring 1988, *19* (1), 1–26.

d'Aspremont, Claude and Louis-André Gérard-Varet, "Incentives and Incomplete Information," *Journal of Public Economics*, February 1979, *11* (1), 25–45.

Edlin, Aaron S. and Benjamin E. Hermalin, "Contract Renegotiation and Options in Agency," *Journal of Law, Economics, & Organization*, 2000, *16*, 395–423.

Epstein, Larry G., "Behavior Under Risk: Recent Developments in Theory and Applications," in Jean-Jacques Laffont, ed., *Advances in Economic Theory: Sixth World Congress*, Vol. 2, Cambridge, England: Cambridge University Press, 1992.

Fleming, Wendell, *Functions of Several Variables*, Berlin: Springer-Verlag, 1977.

Gibbard, Allan, "Manipulation of Voting Schemes," *Econometrica*, July 1973, *41* (4), 587–601.

Gonik, Jacob, "Tie Salesmen's Bonuses to Their Forecasts," *Harvard Business Review*, May-June 1978, pp. 116–122.

Green, Jerry and Jean-Jacques Laffont, "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, March 1977, *45* (2), 427–438.

Grossman, Sanford J. and Oliver D. Hart, "An Analysis of the Principal-Agent Problem," *Econometrica*, January 1983, *51* (1), 7–46.

Guesnerie, Roger and Jean-Jacques Laffont, "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-managed Firm," *Journal of Public Economics*, December 1984, *25* (3), 329–369.

Hart, Oliver D. and Bengt Holmström, "Theory of Contracts," in Truman Bewley, ed., *Advances in Economic Theory: Fifth World Congress*, Cambridge, England: Cambridge University Press, 1987.

Hermalin, Benjamin E. and Michael L. Katz, "Moral Hazard and Verifiability: The Effects of Renegotiation in Agency," *Econometrica*, November 1991, *59* (6), 1735–1753.

Holmström, Bengt, "Moral Hazard and Observability," *Bell Journal of Economics*, Spring 1979, *10* (1), 74–91.

Jullien, Bruno, "L'impact des options exterieures sur les echanges en information asymetrique (The Impact of Outside Options on Exchange under Asymmetric Information)," *Revue Economique*, mai 1996, *47* (3), 437–446.

Laffont, Jean-Jacques and Eric Maskin, "A Differential Approach to Dominant Strategy Mechanisms," *Econometrica*, September 1980, *48* (6), 1507–1520.

_ and Jean Tirole, *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press, 1993.

Lewis, Tracy R. and David E.M. Sappington, "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, December 1989, *49* (2), 294–313.

Maggi, Giovanni and Andres Rodriguez-Clare, "On the Countervailing Incentives," *Journal of Economic Theory*, June 1995, *66* (1), 238–263.

Mas-Colell, Andreu, Michael Whinston, and Jerry Green, *Microeconomic Theory*, Oxford, England: Oxford University Press, 1995.

Maskin, Eric, "Randomization in Incentive Schemes," 1981. Mimeo, Harvard University.

Milgrom, Paul and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, January 1994, *62* (1), 157–180.

Myerson, Roger, "Incentive Compability and the Bargaining Problem," *Econometrica*, 1979, *47*, 61–73.

Page, Frank H., "The Existence of Optimal Contracts in the Principal-Agent Model," *Journal of Mathematical Economics*, 1987, *16* (2), 157–167.

Rabin, Matthew, "Risk Aversion, Diminishing Marginal Utility, and Expected-Utility Theory: A Calibration Theorem," 1997. mimeo, Department of Economics, University of California at Berkeley.

Rochet, Jean-Charles and Philippe Choné, "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, July 1998, *66* (4), 783–826.

Rockafellar, R. Tyrrell, *Convex Analysis*, Princeton, NJ: Princeton University Press, 1970.

Salanié, Bernard, *The Economics of Contracts: A Primer*, Cambridge, MA: MIT Press, 1997.

Sappington, David E. M., "Limited Liability Contracts between Principal and Agent," *Journal of Economic Theory*, February 1983, *29* (1), 1–21.

Shavell, Steven, "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, Spring 1979, *10* (1), 55–73.

Steele, J. Michael, *The Cauchy-Schwarz Master Class*, Cambridge, England: Cambridge University Press, 2004.

Uhlig, Harald, "A Law of Large Numbers for Large Economies," *Economic Theory*, January 1996, *8*, 41–50.

van Tiel, Jan, *Convex Analysis*, New York: John Wiley and Sons, 1984.

Varian, Hal R., "Price Discrimination," in Richard Schmalensee and Robert Willig, eds., *Handbook of Industrial Organization*, Vol. 1, Amsterdam: North-Holland, 1989.

_ , *Microeconomic Analysis*, 3rd ed., New York: W.W. Norton, 1992.

Willig, Robert D., "Consumer's Surplus Without Apology," *American Economic Review*, September 1976, *66* (4), 589–597.

Wilson, Robert B., *Nonlinear Pricing*, Oxford, England: Oxford University Press, 1993.

Index

☡, xiii

additive separability, 124
agent, 63, 117
allocation
by a mechanism, 75
arbitrage, 28, 35

Bayesian Nash
definition of, 111
Blackwell's Theorem, 165
Borch, Karl H., 143
sharing rule, modified, 143, 166
budget balancing, 110

Caillaud, Bernard, 61, 83*n*, 92*n*, 118
capacity constraints, 35
CDFP, *see* concavity of distribution function property
Choné, Philippe, 82
concavity
log, 14
of distribution function property, 163
consumer surplus, 6
aggregate, 9
contract
enforceable, 119
forcing, 125
in a screening model, 76
least-cost, 136
CS, *see* consumer surplus

d'Aspremont, Claude, 114
deadweight loss, 22
triangular shape of, 22
demand
aggregate, 9
inverse, 5*n*
L-shaped, 10
march down, 25
direct mechanism, 76, 106
direct revelation, 73
direct-revelation mechanism, 77, 106
distortion at the bottom, 46, 49
distortion at the top, no, 46, 50, 94

Edlin, Aaron S., 82
elastic, 20
elasticity
of demand, 11
unit, 20
entry fee, 26
ϵ, *see* elasticity
Epstein, Larry G., 146
ex post efficiency, 66
externality
positive, 25

first-order stochastic dominance, 163*n*
Fleming, Wendell, 160*n*
full information, 124
function
density, 11
linear, definition of, 3*n*
survival, 10

Gérard-Varet, Louis-André, 114
Gibbard, Allan, 77*n*
Giffen good, 5*n*
Gonik, Jacob, 74*n*
Green, Jerry R., 5, 77*n*
Grossman, Sanford J., 143, 163, 166, 171

Groves-Clarke mechanism, 109
Guesnerie, Roger, 61, 92*n*

Hart, Oliver D., 143, 163, 166, 171
hazard rate, 11, 56
Hermalin, Benjamin E., 82, 83*n*, 156*n*, 166
hidden action, 118
hidden information, 61
hidden-knowledge model, 99
Holmstrom, Bengt R., 143, 151, 171

IC, *see* incentive-compatibility
implementation, 103
incentive-compatibility constraint, 41, 68
individual-rationality constraint, 26, 68, 125
inelastic, 20
information rent, 43, 61, 67
informed player, 63
IR, *see* individual-rationality constraint

Jensen's Inequality, 130*n*
Jullien, Bruno, 84, 92

Katz, Michael L., 156*n*, 166

Laffont, Jean-Jacques, 61, 77*n*, 92*n*, 110
least-cost action, 156
Lerner markup rule, 18, 21
Lerner, Abba, 21*n*
Lewis, Tracy R., 84
likelihood ratio, 136
log concave, 14

Maggi, Giovanni, 84
marginal cost, 17
marginal revenue, 17
Mas-Colell, Andreu, 5
Maskin, Eric, 82, 110
MC, *see* marginal cost
mechanism, 104
    Bayesian, 112
    direct, *see* direct mechanism
    direct-revelation, *see* direct-revelation mechanism
    dominant-strategy, 105
    Groves-Clarke, 109
    simultaneous-announcement, 104
mechanism design, 26
message profile, 106
MHRP, *see* monotone hazard rate property
Milgrom, Paul, 87
Mills ratio, 56, 95*n*
MLRP, *see* monotone likelihood ratio property
model
    state-space, 120
    two-type, 65
monopsony, 73*n*
monotone hazard rate property, 15
monotone likelihood ratio property, 139, 162
moral hazard, 118
MR, *see* marginal revenue
MRS(marginal rate of substitution, 30
Myerson, Roger, 77*n*, 88

outcome
    of a mechanism, 75

packaging, 28, 35*n*
Page, Frank H., 165*n*
participation constraint, 26, 125
    interim, 101
performance measure, 119
Pólya, George, 14
price
    elasticity of demand, *see* elasticity
price discrimination
    first degree, 25
    perfect, *see* first degree
    quality distortions and, 39
    quantity discounts and, 39
    second degree, 32, 39
    third degree, 32
pricing
    linear, 3
    nonlinear, 3

principal, 61, 63, 117
public-choice problem, 103
punishment
 existence of, 78

Rabin, Matthew, 146
reservation utility, 76, 120, 124
 type-dependent, 84
 type-independent, 84
revelation, 39
revelation constraint, 41, *see* incentive compatibility
Revelation Principle, 77, 105, 106
revenue, 3
Rey, Patrick, 61, 92*n*
Rochet, Jean-Charles, 82
Rodriquez-Clare, Andres, 84

Salanié, Bernard, 61
Sappington, David E.M., 84, 125*n*
screening, 41, 63
self-selection constraint, *see* incentive compatibility
separation, 65
Shannon, Chris, 87
Shavell, Steven, 171
shifting support, 130, 138
 meaningful, 157
signaling, 63
simple monopoly pricing, 17
single-crossing condition, 41, 53*n*, 86
solution
 concept, 105
 in dominant strategies, 105
Spence-Mirrlees condition, 43, 53*n*, 86
standard framework, 81
Steele, J. Michael, 14
strategy
 dominant, 105
 in implementation game, 105
sufficient statistic, 142
support
 definition of, 130*n*
surplus
 full-information, 81

take-it-or-leave-it offer, 24
tariff
 multi-part, 39
 two-part, 26
Taxation Principle, 77, 78
TIOLI, *see* take-it-or-leave-it offer
Tirole, Jean, 61, 92*n*
truth-telling constraint, *see* incentive compatibility
two-instruments principle, 28, 143*n*
type, 31, 65
 space, 31
type space, 75

Uhlig, Harald, 11*n*
uninformed player, 63
utility
 quasi-linear, 5
 reservation, *see* reservation utility
 von Neumann-Morgenstern, 75

van Tiel, Jan, 127*n*, 130*n*
Varian, Hal R., 5, 7, 23*n*, 30, 35
verifiable, 119*n*

Whinston, Michael D., 5
Willig, Robert, 7
Wilson, Robert B., 30, 67