Monopoly[1]

Elmar Wolfstetter

September 1996
Institut f. Wirtschaftstheorie I
Humboldt–Universität zu Berlin
Spandauer Str. 1
10178 Berlin
Germany
e–mail: wolf@wiwi.hu–berlin.de

[1]Comments by Anne Bussmann, Uwe Dulleck, Christian Hafner and Dorothea Kübler are gratefully acknowledged.
This research was supported by the Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich 373, "Quantifikation und Simulation Ökonomischer Prozesse", Humboldt–Universität Berlin.

Abstract

This is the first chapter of a graduate text entitled *Topics in Microeconomics.* It covers the basics of monopoly theory. Most of the material is kept at an intermediate level to serve as a bridge between the intermediate level training and the graduate level focus of the book. However, some sections, identified with a *, are at an advanced level.

The Chapter begins with the simple economics of Cournot monopoly, adding the quality dimension, the assessment of the welfare loss of monopoly in the face of rent seeking behavior, and the dynamics of pricing and inventory when demand is subject to unpredictable fluctuations.

Turning to price discrimination, the distinction between first–, second–, and third–degree price discrimination is introduced. The incomplete information theory of second–degree price discrimination is worked out, first for two and then for a continuum of customer types. Next, it is shown how the frequently observed *intertemporal* price discrimination gives rise to a time consistency problem ("durable goods monopoly"), and how the basic theory of third–degree price discrimination needs to be modified accordingly.

The Chapter closes with the noncooperative bargaining theory of bilateral monopoly and suggests a further marriage of monopoly and bargaining theories. The regulation of monopoly is covered in the separate Chapter on *Regulation of Monopoly* which was already circulated as a Discussion Paper.

Contents

"The best of all monopoly profits is a quiet life."
Sir John Hicks

1 Introduction

In this chapter we analyze the supply and pricing decisions of a pure, single–product monopolist, facing a large number of price taking buyers. We take the firm's choice of product as given and assume that consumers know all about product characteristics and quality. Moroever, we assume that the monopolist's market is sufficiently self–contained to allow us to neglect the strategic interdependency between markets. The strategic interdependency between markets is the subject matter of the theory of oligopoly with product differentiation.

Monopolies do exist. In the early days of photocopying, Rank Xerox was the exclusive supplier — we still use the word xeroxing as a synonym for photocopying. Postal and rail services are (or have been) monopolized (things are changing fast in these sectors), and so are public utilities (gas and electricity), to name just a few. One can even find perfectly inconspicuous products that are subject to monopolization. For example, in Germany matches were exclusively supplied by a single Swedish supplier who had acquired a monopoly license from the German government during WW I, when the German government was hard pressed for foreign currencies. Similarly, gambling licenses are often issued by states to raise revenue. Moreover, there are many local monopolies, like the single hardware store in a small community, the busline exclusively served by Greyhound, or the flight route, say from Ithaca to New York City, served by a single airline.

As these examples suggest, monopolization has a lot to do with the size of a market, but also with licensing, patent protection, and regulation — supported by law. If entry into a monopolized market is not prohibited, a monopoly has little chance to survive unless the market is too small to support more than one firm. Monopoly profits attract new entrants. And even if entry is prohibited, patent rights expire,[1] rival firms spend resources to develop similar products and technologies or even to gain political influence to raid the monopoly license. Therefore, a monopoly is always temporary

[1] The duration of patents varies, ranging from 16 to 20 years in most countries. In some countries, like France, certain types of patents are given shorter terms because the inventions have an overall general usefulness. Incidentally, the U.S. grants patents to the party that is "first to invent" whereas most European countries award to the party that is "first to file".

unless it is continuously renewed through innovations, patents, or political lobbying.

Monopolies — weak and strong A monopolist has exclusive control of a market. But to what extent a monopoly is actually turned into a fat profit depends upon several factors, in particular:

- the possibility of price discrimination,
- the closeness to competing markets,
- the ability to make credible commitments.

A "strong" monopolist has full control over his choice of price function. He can set linear or non–linear prices, he can even charge different prices from different buyers. In other words, the strong monopolist can use all his imagination to design sophisticated pricing schemes to pocket the entire gain from trade, restricted only by consumers' willingness to pay. No one will ever doubt the credibility of his announced pricing policy.

In contrast, the "weak" monopolist is restricted to linear prices.[2] He cannot even price discriminate between consumers.

Monopolists come in all shades, between the extremes of "weak" and "strong". For example, a monopolist may be constrained to set linear prices, but he may be able to price discriminate between some well identified groups of consumers. Or a monopolist may be restricted to set a menu of non–linear prices, just like the ones you are offered by your long–distance telephone company and your public utilities suppliers.

In the following pages you will learn more about these and other variations of the monopoly theme. We will not only analyze the monopolist's decision problem under various pricing constraints, but also attempt to explain what gives rise to these constraints from basic assumptions on technology, transaction costs, and information structures.

We begin with the simplest analysis of the weak monopoly, also known as *Cournot monopoly*, in homage to the French economist Antoine Augustin Cournot (1801–1877) who laid the foundations for the mathematical analysis of noncompetitive markets. Most of this analysis should be familiar from your undergraduate training. Therefore, you may quickly skim through these first pages, except when we cover the relationship between rent seeking and the social loss of monopoly, the Keynesian "price rigidity" property

[2] A price function $\mathcal{P}$ is called linear if it has the form $\mathcal{P}(x) := px$ where $p > 0$ is the unit price.

of monopolist pricing in the face of demand fluctuations, the durable goods monopoly problem, and the analysis of regulatory mechanisms.

Finally, keep in mind that there are really two opposite ways to model pure monopoly. The most common approach — exclusively adopted in this chapter — describes the monopolist as facing a given market demand function and ignores potential actions and reactions by the suppliers of related products. The other, opposite approach faces the strategic interdependency of markets head on and views monopolist pricing as an application of the theory of oligopoly with product differentiation. While we stick, in this chapter, to the conventional approach, you should nevertheless keep in mind that there are many examples where oligopoly theory gives the best clues on the monopolist's decisions.[3]

2 Cournot Monopoly — Weak Monopoly

We begin with the weak or Cournot monopolist who can only set a linear price function that applies equally to all customers. The demand function, defined on the unit price p, is denoted by $X(p)$, and the cost function, defined on output x, by $C(x)$. Both $X(p)$ and $C(x)$ are twice continuously differentiable; also $X(p)$ is strict monotone decreasing, and $C(x)$ strict monotone increasing. The inverse demand function, defined on total sales x, exists (due to the monotonicity of $X(p)$) and is denoted by $P(x)$. The rule underlying this notation is that capital letters like X and P denote functions, whereas the corresponding lower case letters x and p denote supply and the unit price.

In a nutshell, the Cournot monopolist views the market demand function as his menu of price–quantity choices from which he picks that pair that maximizes his profit. We will now characterize the optimal choice.

At the outset, notice that there are two ways to state the monopolist's decision problem: one, in terms of the demand function:

$$\max_{p,x} \quad px - C(x), \quad \text{s.t.} \quad X(p) - x \geq 0, \quad p, x \geq 0,$$

and the other in terms of the *inverse* demand function:

$$\max_{p,x} \quad px - C(x), \quad \text{s.t.} \quad P(x) - p \geq 0, \quad p, x \geq 0.$$

Obviously, both are equivalent. Therefore, the choice is exclusively one of convenience. We choose the latter. Also, notice that the constraint is binding

[3] Of course, also the opposite may hold, where standard monopoly theory gives the best clues on oligopolistic pricing. This is the case when reaction functions are horizontal.

(the monopolist would forgo profits if he did sell a given quantity below the price customers are willing to pay). Therefore, the monopolist's decision problem can be reduced to the unconstrained program:

$$\max_{x\geq 0} \quad \pi(x) := R(x) - C(x) \tag{2.1}$$

where R denotes the revenue function:

$$R(x) := P(x)x. \tag{2.2}$$

2.1 Cournot Point

Suppose, for the time being, that X and C are continuously differentiable on $\mathbb{R}_+$, that revenue $R(x)$ is bounded, and that profit is strictly concave.[4] Then, the decision problem is well behaved, and we know that there exists a unique solution that can be found by solving the Kuhn–Tucker conditions:[5]

$$\pi'(x) := R'(x) - C'(x) \leq 0 \quad \text{and} \quad x\pi'(x) = 0, \quad x \geq 0. \tag{2.3}$$

In principle, one may have a corner solution ($x = 0$). But if $P(0) - C'(0) > 0$, an interior solution is assured, which is characterized by the familiar condition of equality between marginal revenue and marginal cost $R'(x) = C'(x)$. Denote the solution by $x^M, p^M := P(x^M)$. The graph of the solution is called "Cournot point", and illustrated in Figure 1.

Example 1 *Suppose $P(x) := a - bx$, $a, b > 0$, and $C(x) := \frac{1}{2}x^2$, $x \in [0, a/b]$. Then, profit is a strictly concave function of output, $\pi(x) := ax - bx^2 - \frac{1}{2}x^2$. From the Kuhn–Tucker condition one obtains:*

$$0 = \pi'(x) = a - 2bx - x. \tag{2.4}$$

Therefore, the Cournot point is ($x^M = \frac{a}{1+2b}, p^M = \frac{a(1+b)}{1+2b}$), and the maximum (or indirect) profit function is: $\pi^*(a, b) := \frac{a^2}{2(1+2b)}$.

[4] Concavity of the revenue and convexity of the cost function — at least one strict — are sufficient, but not necessary.

[5] In case you are unsure about this, prove the following: 1) strong concavity implies strict concavity; 2) if a solution exists, strict concavity implies uniqueness; 3) the Weierstrass–theorem implies existence of a solution (you have to ask: is the feasible set closed and bounded?); 4) the Kuhn–Tucker theorem implies that every solution solves the Kuhn–Tucker conditions, and *vice versa*. Consult Appendix C and D.

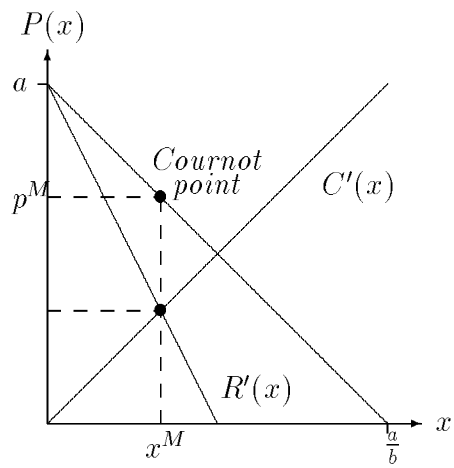

Figure 1: Cournot–point

Obviously, the monopolist's optimal price exceeds marginal cost. But by how much? The answer depends upon how strongly demand responds to price. If demand is fairly inelastic, the monopolist has a lot of leeway; he can charge a high mark–up without suffering much loss of demand. But if demand responds very strongly to a price hike, the best the monopolist can do is to stay close to marginal cost pricing. This suggests a strong link between monopoly power and price responsiveness of demand.

The conventional measure of price responsiveness of demand is the *price elasticity of demand:*

$$\varepsilon(p) := X'(p)\frac{p}{X(p)}. \tag{2.5}$$

We now use this measure to give a precise statement of the conjectured explanation of monopoly power.

As you probably recall from undergraduate micro, marginal revenue is linked to the price elasticity of demand as follows (see also Fig 2):[6]

$$\begin{aligned} R'(x) &= P'(x)x + P(x) \\ &= P(x)[1 + P'(x)\frac{x}{P(x)}] \\ &= P(x)[1 + \frac{x}{X'(P(x))P(x)}] \\ &= P(x)[1 + \frac{1}{\varepsilon(P(x))}] \\ &= P(x)[\frac{1+\varepsilon(P(x))}{\varepsilon(P(x))}]. \end{aligned} \tag{2.6}$$

Therefore, marginal revenue is positive if and only if demand responsiveness is low, in the sense that the price elasticity of demand is less than -1.

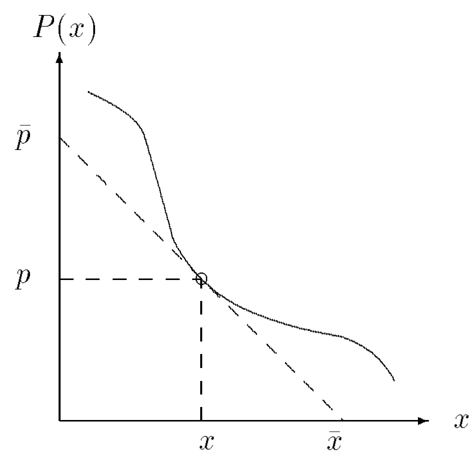

Figure 2: Relationship between marginal revenues and price elasticity

$|\varepsilon| = \frac{p}{\bar{p}-p}$, $R'(x) = p - (\bar{p} - p)$,
$E'(p) = x - (\bar{x} - x)$, $E(p) := pX(p)$.

Using this relationship together with the Kuhn–Tucker condition (2.3) for an interior solution, one obtains the following optimal "mark–up rule"

$$P(x) = \frac{\varepsilon(P(x))}{1+\varepsilon(P(x))}C'(x). \tag{2.7}$$

[6] One has $\frac{1}{P'(x)} = X'(P(x))$ because, by a known result, the first derivative of the inverse of a function is equal to the inverse of the first derivative of that function (provided $P'(x) \neq 0$.

Another frequently used variation of this form is the "Lerner index" of monopolization

$$\frac{P(x) - C'(x)}{P(x)} = \frac{1}{-\varepsilon(P(x))}. \tag{2.8}$$

These convenient forms should also remind you that the Cournot point always occurs at a point where the price elasticity of demand is less than minus one, that is where an increase in output raises revenue.

Monopoly and Mark–Up Pricing In the applied literature on Industrial Organization it is claimed that monopolistic firms often stick to a rigid mark–up pricing rule. This practice is sometimes quoted as contradicting basic principles of microeconomics. Notice, however, that (2.7) is consistent with a constant mark–up. All it takes is a constant elasticity demand function.

Another issue in this literature concerns the problem of measurement. Usually, one has no reliable data on firms' cost functions. So how can one ever measure such a simple thing as the Lerner index? As in other applications, a lot of ingenuity is called for to get around this lack of data.

A nice example for this kind of ingenuity can be found in Peter Temin's study of the German steel cartel in Imperial Germany, prior to WW I.[7] He noticed that the cartel sold steel also at the competitive world market. Temin concluded that the world market price, properly converted using the then current exchange rate, should be a good estimate of the steel cartel's marginal cost. And he proceeded to use this estimate to compute the Lerner index. Make sure you understand the economic reasoning behind this trick.

Monopoly and Cost–Push Inflation In economic policy debates it is sometimes claimed that monopolists contribute to the spiraling of "cost–push" inflation because — unlike competitive firms — monopolists apply a mark–up factor greater than 1. To discuss this assertion, it may be useful if you plot the mark–up factor $\frac{\varepsilon}{1+\varepsilon}$ for all $\varepsilon < -1$. Notice that it is always greater than 1, increasing in ε, and approaching 1 as ε goes to minus infinity and infinity as ε approaches -1.

Software Pirates and Copy Protection As a brief digression, consider a slightly unusual Cournot monopoly: the software house that faces competition from illegal copies and in response contemplates to introduce copy protection.

[7] See Temin, P. [1976]. *Did Monetary Forces Cause the Great Depression?*, Norton.

Legally, the copying of software is theft. Nevertheless, it is widespread, even among otherwise law abiding citizens. Software houses complain that illegal copies rob them of the fruits of their labor and pose a major threat to the industry.

Suppose copy protection is available at negligible cost. Should the monopolist apply it and, if so, how many copies should he permit? A copy protected program can only be copied $N \geq 0$ times, and copies cannot be copied again. Therefore, each original copy can be made into $N + 1$ *user copies.*

To discuss the optimal copy protection, we assume that there is a perfect secondary market for illegal copies. For simplicity, users are taken to be indifferent between legal and illegal user copies, and marginal costs of copying are taken to be constant.

Given these admittedly extreme assumptions, the software market is only feasible with some copy protection. Without it, each original copy would be copied again and again, until the price equals the marginal cost of copying. Anticipating this, no customer would be willing to pay more than the marginal cost of copying, and the software producer would go out of business because he knew that he could never recoup the fixed cost of software development.

An obvious solution is full copy protection ($N = 0$) combined with the Cournot point (p^M, x^M). However, this is not the only solution. Indeed, the software producer can be "generous" and permit any number of copies between 0 and $x^M - 1$ without any loss in profit. All he needs to do is to make sure that N does not exceed $x^M - 1$ and that the price is linked to the number of permitted copies in such a way that each original copy is priced at $N + 1$ times the Cournot equilibrium price, $(N + 1)p^M$.

Given this pricing plus copy protection rule, each customer anticipates that the price per user copy will be equal to the Cournot equilibrium price p^M; exactly $x^M/(N + 1)$ original copies are sold; each original copy is copied N times; exactly x^M user copies are supplied; and profits and consumer surplus are the same as under full copy protection.

At this point you may object that only few software houses have introduced copy protection;[8] nevertheless, the industry is thriving. So what is missing in our story?

One important point is that copy protection is costly, yet offers only temporary protection. Sooner or later, the code will be broken; there are far too many skilled "hackers" to make it last. Another important point is that

[8] *Lotus* is one of the few large software houses that rely on copy protection.

illegal copies are often imperfect substitutes, for example because handbooks come in odd sizes (not easily fit for xeroxing) or because illegal copies may be contaminated with computer viruses. In lieu of adding complicated copy protection devices, the monopolist may actually plant his own virus contaminated copies into the second hand market. Alas, computer viruses are probably the best copy protection.

Leviathan, Hyperinflation, and the Cournot Point We have said that monopoly has a lot to do with monopoly licensing, granted and enforced by the legislator. Of course, governments are particularly inclined to grant such licenses to its own bodies. This suggests that some of the best applications of the theory of monopoly should be found in the public sector of the economy.

A nice example that you may also come across in macroeconomics concerns the "inflation tax" theory of inflation and its application to the economic history of hyperinflations. A simple three ingredient macro model will explain this link.[9]

1) Government has a monopoly in printing money, and it can coerce the public to use it by declaring it "fiat money". Consider a government that finances all its real expenditures G by running the printing press. Let p be the price index, M^S the stock of high powered money, and suppose there are no demand deposits. Then, the government's budget constraint is

$$pG = \Delta M^S \quad \text{(budget constraint).} \tag{2.9}$$

.

2) Suppose the demand for real money balances M^d/p is a monotone decreasing and continuously differentiable function of the rate of inflation $\hat{p}$[10]

$$\frac{M^d}{p} = \phi(\hat{p}) \quad \text{(demand for money).} \tag{2.10}$$

3) Assume the simple quantity theory of inflation

$$\hat{p} := \frac{\Delta p}{p} = \frac{\Delta M^S}{M^S} = \frac{\Delta M^d}{M^d} \quad \text{(quantity theory of money).} \tag{2.11}$$

[9] The classic reference is Cagan, P. [1956]. "The monetary dynamics of hyperinflation", in: Friedman, M. (ed.). *Studies in the Quantity Theory of Money.* University of Chicago Press, 25-117.

[10] In macroeconomics it is often assumed that the demand for real money balances is a strict monotone decreasing function of the nominal interest rate. The latter is usually strongly correlated with the rate of inflation.

Putting all three pieces together it follows that the real expenditures that can be financed by running the printing press are a function of the rate of inflation

$$G(\hat{p}) = \hat{p}\phi(\hat{p}). \tag{2.12}$$

The government has the exclusive right to issue money, and it can force people to accept this money in exchange for goods and services (this is the origin of the term "fiat money"). However, even though it can set the speed of the printing press, the real expenditures that it can finance in this manner are severely limited. Therefore, the "inflation tax" is only a limited substitute for conventional taxes.

To determine these limits, simply compute the Cournot point rate of inflation $\hat{p}^M$, defined as the maximizer of $G(\hat{p})$ over $\hat{p}$. Since the government's maximization problem is equivalent to that of a Cournot monopolist subject to zero marginal costs, it follows immediately that real government expenditures reach a maximum at that rate of inflation where the elasticity of the demand for real money balances ε is equal to -1

$$\varepsilon(\hat{p}) := \phi'(\hat{p})\frac{\hat{p}}{\phi(\hat{p})} = -1. \tag{2.13}$$

Of course, this revenue maximizing rate of inflation imposes a deadweight loss upon society, just like any other Cournot monopoly. The socially optimal rate of inflation is obviously equal to zero. However, alternative methods of taxation tend to impose their own deadweight loss, in addition to often high costs of collecting taxes. Keeping these considerations in mind, it may very well be that some inflation is optimal, depending upon tax morale and other insitutional issues. Indeed, different countries with their different institutions may very well have different optimal inflation rates. Incidentally, these considerations are the background of current discussions on optimal currencies areas.

Another interesting application of the inflation tax concerns the theory of hyperinflations, like the one in Weimar Germany in 1923 (or most recently in Serbia, after the breakup of former Yugoslavia). Here, a government was in desperate need for funds, due to a fatal combination of events, from the exorbitantly high demands for reparations imposed by the Versaille treaty (aggravated by the French occupation of the Ruhr area in 1923), and a parliament torn between cooperation and conflict. Unable to finance its expenditures to any significant degree by explicit taxes, the government took recourse to the printing press. But the faster it set its speed, the fewer real expenditures

Exchange Rate (monthly averages)		
		Mark/$
January	1921	64
January	1922	191
January	1923	17 972
July	1923	353 412
August	1923	4 620 455
September	1923	98 860 000
October	1923	25 260 208 000
November	1923	4 200 000 000 000

Table 1: German Hyperinflation 1923

Source: Stolper, G. [1964].
Deutsche Wirtschaft Seit 1870, Mohr & Siebeck.

it could finance in this manner. The result was a rapidly exploding rate of inflation, reflected in the catastrophic devaluation of the Mark relative to the Dollar reported in Table 1 below, and a complete breakdown of government financing.[11]

Some Comparative Statics How does the Cournot point change if the marginal cost or the demand function shifts? As always, such questions are meaningful only if uniqueness of the Cournot point is assumed. This is one reason why comparative statics is always pursued in a framework of relatively strong assumptions.

As an example, suppose C is a continuously differentiable function of a cost parameter α in such a way that higher α represents higher marginal costs $C''_{x\alpha}(x, \alpha) > 0$. Also assume that the profit function is *strongly* concave in output and that the Cournot point is an interior solution.[12] Then, the optimal output is a differentiable function of α, described by the function

[11] At some point, the Reichsbank employed 300 paper manufacturers and 2,000 printing presses, day and night.

[12] Recall, strong concavity is strict concavity plus the requirement that the determinant of the Hessian matrix of the profit function (which is here simply the second derivative of this function) does not vanish. Strong concavity is always invoked if one wants to make sure that the solution functions are differentiable in the exogenous parameter, which is a prerequisite for the calculus approach to comparative statics.

$x^*(\alpha)$. And we can pursue comparative statics using calculus.

We now show that the monopolist's supply is strict monotone decreasing in α, $x^{*'}(\alpha) < 0$. For this purpose, insert the solution function $x^*(\alpha)$ into the Kuhn–Tucker condition (2.3), and one obtains the identity

$$\pi'_x(x^*(\alpha), \alpha) := R'(x^*(\alpha)) - C'_x(x^*(\alpha), \alpha) \equiv 0. \tag{2.14}$$

Differentiating it with respect to α gives, after a bit of rearranging

$$x^{*'}(\alpha) = \frac{-\pi''_{x\alpha}(x^*(\alpha), \alpha)}{\pi''_{xx}(x^*(\alpha), \alpha)} = \frac{C''_{x\alpha}(x^*(\alpha), \alpha)}{R''(x^*(\alpha)) - C''_{xx}(x^*(\alpha), \alpha)} < 0. \tag{2.15}$$

This proves that the monopolist's optimal supply is strict monotone decreasing in the marginal cost parameter, as asserted.

Two Technical Problems We close the analysis of the Cournot point with two slightly technical problems. The first one concerns the existence of the Cournot point in the face of plausible discontinuities of demand or cost functions. The second explains how you should proceed if the profit function is not strictly concave. If you are in full control of your undergraduate micro, you may skip this exposition, and move directly to Subsection 8.8.2.

An Existence Puzzle Suppose demand is unit elastic ($\varepsilon = -1$, for all $x > 0$), and the cost function is strictly convex with positive profits at some outputs. Then the profit function is strictly concave. Yet, the monopolist's decision problem has no solution.

The explanation is very simple. First, notice that revenue is constant for all positive x whereas cost is strictly increasing. Therefore, profit goes up as x is reduced (less output means higher profit), except if x is reduced all the way down to $x = 0$. Second, notice that there is no smallest positive rational number (there is no smallest positive output). Combine both observations, and it follows that there is no profit maximizing choice of x. So which of our assumptions has failed?

As you check the assumptions one by one, you will see that almost all of them are satisfied. The only exception is the continuity of the revenue function which is violated at precisely one point ($x = 0$).[13] This seemingly minor deviation changes it all.

[13]This discontinuity rules out the application of the Weierstrass theorem which was invoked in the proof of existence of the Cournot point sketched in an earlier footnote.

The discontinuity of the demand function at $x = 0$ is something that one would not like to rule out. For example, applied economists often work with constant elasticity demand functions, all of which share this discontinuity property.

Another frequently encountered discontinuity that should not be excluded either concerns the cost function. Recall, costs are usually decomposed into fixed and variable where fixed costs are defined as $\lim_{x\to 0} C(x)$. Some fixed costs are *reversible* (or quasi–fixed), and some are irreversible or *sunk*. Whenever some fixed costs are reversible, one has $C(0) < \lim_{x\to 0} C(x)$, so that the cost function has a discontinuity at $x = 0$. In the face of it there is always a reasonable chance that the corner point $x = 0$ may be optimal. Therefore, watch out for a corner solution.

So, what shall you do if the demand or the cost function has such a discontinuity, and how can one assure existence of the Cournot point even in these cases? As in other applications, a safe procedure is to break up the search for a solution into three steps: 1) search for a solution in the restricted domain $\mathbb{R}_{++}$ (an interior solution); 2) evaluate profit at the corner point $x = 0$; 3) choose the solution (either corner or interior) with the highest profit.

Since this procedure is cumbersome, one would of course like to know in which case existence of an interior solution is guaranteed so that the procedure can be stopped after round 1). A simple and often used sufficient condition is the following

$$\lim_{x\to 0}[R'(x) - C'(x)] > 0, \quad \lim_{x\to \infty}[R'(x) - C'(x)] < 0. \tag{2.16}$$

Make sure that you understand why this condition is indeed sufficient.

Example 2 *Suppose the demand function has a constant elasticity $\varepsilon < -1$. Then it must have the form $X(p) = ap^{\varepsilon}$ (show this), so that the inverse demand function is $P(x) = \left(\frac{x}{a}\right)^{\frac{1}{\varepsilon}}$. Twice differentiate the revenue function $R(x) := P(x)x$, and you see that the revenue function is strictly concave. Now add the assumption that the cost function is convex and that condition (2.16) holds. Then, the Cournot point has a unique interior solution.*

The Cournot Point without Concavity Let us get another technical problem out of the way: characterizing the Cournot point if the profit function is not strictly concave. Concavity, as a local property, assures that a stationary point is indeed a maximum, and strict concavity, as a global property, assures uniqueness. But concavity is far too strong a requirement.

A popular weaker requirement is quasiconcavity. But, as a global property, quasiconcavity is often difficult to confirm or reject. Like in other optimization problems, if a maximization problem is not concave (that is if either the objective function is not concave or the constraint set is not convex), it is often a better procedure to look for some transformation of variables that leads to a concave problem. The trouble is, however, that there are no simple rules of thumb and that you have to be imaginative to find a transformation that does the job.[14]

In many applications one can safely assume that the cost function is convex. But one may feel less comfortable assuming concavity of the revenue function. So you may wonder whether one could not assume instead that the revenue function is quasiconcave and then obtain a quasiconcave profit function, which is really enough for a well–behaved decision problem. The answer is no. Just recall that the sum of a concave and a quasiconcave function need not be quasiconcave. Consult Appendix C if you are not entirely sure about this matter.

So what shall you do if the profit function is continuous but not quasiconcave in output or in any conceivable transformation of this variable? Well, you cannot avoid the tedious job of checking out all stationary points and all corners. Of course, only those stationary points can qualify where the profit function is (locally) concave. Therefore, you need only consider those stationary points at which the "second order" or local concavity condition

$$\pi''(x) := R''(x) - C''(x) \leq 0 \quad \textit{(second order condition)} \tag{2.17}$$

is satisfied. But you may still be left with fairly extensive computations to compare the profits at the remaining stationary and corner points.

Example 3 *Suppose the cost function is S–shaped (strictly concave for low and strictly convex for high outputs), and suppose demand is linear. Then, the profit function has two stationary points. But the profit function is only locally concave at the one point with the higher output. Therefore, only one stationary point survives the second order or local concavity condition. However, this point need not be a profit maximum either. Indeed, if fixed costs are sufficiently high, it is always optimal to close down the firm and choose the corner point $x = 0$.*[15] *Draw a diagram to illustrate this case.*

[14] An example for such a transformation of variables was spelled out in detail in our analysis of the labor managed firm, in Chapter 2.

[15] It is useful to distinguish two cases: 1) Suppose average cost is higher than the price at the qualifying stationary point. Then, the corner point $x = 0$ is definitely optimal if

2.2 Deadweight Loss of Monopoly

Compared to a competitive firm, the Cournot monopolist earns higher profits (if he did not, the price would have to be equal to marginal cost, at the Cournot point). This shows that monopoly power redistributes welfare from buyers to sellers. But redistribution alone does not indicate any loss of social welfare, in the sense of the Pareto criterion. However, since the Cournot monopolist can only extract more of the consumers' willingness to pay by charging a higher unit price, the monopolist reduces welfare, unless demand is completely inelastic.

If the unit price rises above the competitive level, the consumers who continue to buy at the now higher price suffer a loss in consumer surplus that is however exactly offset by the seller's gain. However, those who quit buying at the higher price suffer a loss not offset by any gain to the seller. This "deadweight loss" of Cournot monopoly is illustrated in Figure 3, by the shaded area D.

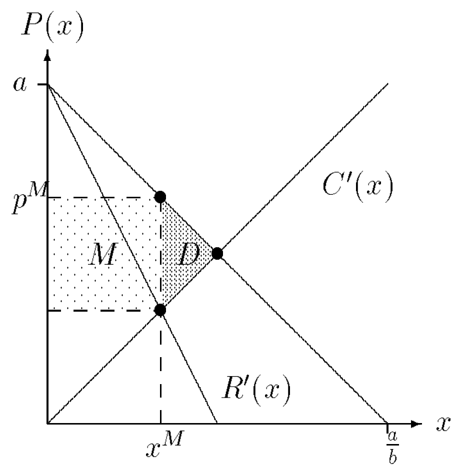

Figure 3: Deadweight Loss of Monopoly

As always, a deviation from the welfare optimum suggests that, with a bit of imagination, one can design Pareto improving trades. For example, starting from the Cournot point, the monopolist could propose to his customers to supply an additional $(x^* - x^M)$ units in exchange for an additional payment equal to the cost increment (measured by the area under the marginal

fixed costs are reversible (not sunk). 2) Suppose average variable cost is higher than the price at the qualifying stationary point. Then $x = 0$ is optimal even if fixed costs are irreversible (sunk).

cost function, between x^M and x^*), plus some small bonus. Miraculously, both buyers and sellers would be better–off. However, the weak Cournot monopolist cannot take advantage of these gains from trade because he is restricted to simple linear pricing schemes, for reasons that we will have to be explained from basic assumptions concerning technology and information structures.

Essentially, the deadweight loss of monopoly is the same as the deadweight loss of taxation. In the Middle Ages, it was popular to tax real estate on the basis of the size of windows. Due to the conspicuously high price of glass, the size of windows was correlated with wealth. Just like consumers reduce their demand when a monopolist raises the unit price, medieval citizens responded to the window tax by reducing the size of windows. In the end, they paid their dues in any case. But on top of the direct reduction of wealth due to taxation, they sat in the dark — a visible example of the deadweight loss of taxation.

Can a government reduce or even eliminate the deadweight loss of monopoly by means of corrective taxes? If the government has complete information about cost and demand functions, the task is easily accomplished. For example, a simple linear subsidy based on output — a negative excise tax — will do the job. The intuition is simple. An output subsidy smoothly reduces the effective marginal cost. By result 2.15, it follows immediately that the subsidy increases the Cournot equilibrium output. Therefore, one only needs to set the subsidy at the right level, and the monopolist is induced to produce the socially optimal level of output.

An obvious objection is that such a subsidy makes the monopolist even richer. However, this side effect of the output subsidy scheme can easily be eliminated by adding an appropriate lump–sum tax into the package.

To compute the appropriate subsidy rate and lump–sum tax, you should proceed as follows. In a first step, solve the monopolist's decision problem, given a subsidy rate s per output unit and a lump–sum tax T. Of course, the lump–sum tax does not affect the optimal output, but the subsidy does. Then, impose the requirement that the optimal output be equal to the competitive output x^M, implicitly defined by the condition $P(x^M) = C'(x^M)$. After a bit of rearranging the first–order condition you will find that the subsidy rate has to be set as follows

$$s = \frac{-C'(x^M)}{\varepsilon(P(x^M))} > 0. \qquad (2.18)$$

Finally, make the subsidy self–financing, by setting the lump–sum tax equal to

$$T = sx^M. \tag{2.19}$$

It is as simple as that.[16]

However, in most applications the regulation of Cournot monopoly is considerably more difficult. The main reason is that monopolists usually have private information about costs and sometimes even about their demand function. This raises a challenging mechanism design problem. We will address this issue in some detail in section ?? of the next chapter.

Another problem has to do with the fact that monopolies are often the product of government regulation. It is hard to imagine that those agencies that restrict entry and thus permit monopolization will also tightly monitor these monopolies and direct them toward maximizing social welfare. And indeed, many economists are inclined to view regulation as industry dominated and directed primarily to the industry's benefit. As Stigler[17] put it: *" ... as a rule, regulation is acquired by the industry and is designed and operated primarily for its benefit."*

2.3 Social Loss of Monopoly and Rent Seeking

The deadweight loss of monopoly D in Figure 3, however, tends to underestimate the social loss of monopoly. As Posner observed:

> *"The existence of an opportunity to obtain monopoly profits will attract resources into efforts to obtain monopolies, and the opportunity costs of those resources are social costs of monopoly too."*[18]

Under idealized conditions, the additional loss of monopoly is exactly equal to the monopoly profit measured by the area M in Figure 3. Therefore, the social cost of monopoly is the sum of the deadweight loss D and the monopoly profit M.

The additional loss may easily outweigh the traditional deadweight loss. For example, if consumers are identical and demand is perfectly inelastic,

[16] An even simpler mechanism is to impose sufficiently high penalties on any deviation from marginal cost pricing. This just shows that the regulation of monopoly is a trivial task if the regulator has complete information.

[17] Stigler, G. J. [1971]. "The theory of economic regulation", *Bell Journal of Economics*, 2: 3-21.

[18] Posner, R. [1975]. "The social cost of monopoly and regulation", *Journal of Political Economy*, 83: 807-827.

the deadweight loss vanishes, but the monopoly profit is as large as consumers' aggregate willingness to pay. This suggests that the additional cost component deserves close scrutiny.

The key assumption underlying the proposed inclusion of the monopoly profit as part of the social loss of monopoly is that obtaining a monopoly is itself a competitive activity. Even though there is perhaps no *competition in the market*, there is almost always *competition for the market*. The contestants spend resources to such an extent that, at the margin, the cost of obtaining the monopoly is exactly equal to the expected profit of being a monopolist. For if a monopoly could be acquired at a bargain, others would try to take it away until no net gain can be made. As a result, monopoly profits tend to be transformed into costs, and the social cost of monopoly is made equal to D plus M.

A simple argument illustrates this point. Suppose n identical firms spend resources, each at the level z, to obtain a lucrative monopoly with the monopoly profit $\pi^M > 0$. Then each firm has a $\frac{1}{n}$ chance to win π^M. In equilibrium n and z are such that the expected value of profit from participating the contest is equal to zero

$$\frac{1}{n}\pi^M - z = 0. \tag{2.20}$$

And, therefore, the monopoly profit is exactly equal to the overall cost of competition for the market

$$\pi^M = nz, \tag{2.21}$$

as asserted.

Assuming competition for the market is reasonable in many applications. For example, if monopoly is based on patents, many firms can enter the patent race for this monopoly.[19] Or if monopoly is based on public licensing, many firms can enter into the political lobbying or perhaps even bribery necessary to obtain a license or raid an existing one.[20]

[19]Incidentally, Plant, A. [1934]. "The economic theory concerning patents", *Economica*, 1: 30-51, criticized the patents' system precisely on the ground that it draws greater resources into inventions than into activities that yield only competitive returns.

[20]The case of bribery poses an intriguing problem. At first glance, one is inclined to argue that bribery is purely redistributional and therefore cannot qualify as a social loss component. However, if a political office is the recipient of substantial bribes, it is itself a lucrative monopoly, subject to its own competition for office. As a result, people will spend resources, for example in education, to be put in office and stay in office. Ultimately, it is these costs associated with competition for office that represent the social loss of monopoly.

2.4 Monopoly and Product Quality

A Cournot monopolist supplies insufficient output relative to the welfare optimum. Can one extrapolate and claim that a Cournot monopolist supplies also insufficient quality?

In order to answer this question, suppose quality can be described by a single index, measured by the real valued variable q. Let cost and inverse demand be functions of output x *and* quality q, denoted by $C(x, q)$ and $P(x, q)$, and let C be monotone increasing in x and q and P decreasing in x and increasing in q. Also, choose C and P in such a way that profit

$$\pi(x, q) := P(x, q)x - C(x, q)$$

and total surplus

$$T(x, q) := \int_0^x P(\tilde{x}, q)d\tilde{x} - P(x, q)x + \pi(x, q) = \int_0^x P(\tilde{x}, q)d\tilde{x} - C(x, q),$$

are strictly concave.[21] Then, the choice of output and quality can be described by first order conditions.

We compare the welfare optimal output and quality (x^{PO}, q^{PO}) with the generalized Cournot point (x^M, q^M). The welfare optimum maximizes total surplus whereas the Cournot point maximizes profit. Assuming interior solutions, the Cournot point solves the first–order conditions

$$\pi'_x := P + P'_x x - C'_x = 0, \tag{2.22}$$

$$\pi'_q := P'_q x - C'_q = 0 \tag{2.23}$$

whereas the welfare optimum solves

$$T'_x := P - C'_x = 0, \tag{2.24}$$

$$T'_q := \int_0^x P'_q(\tilde{x}, q)d\tilde{x} - C'_q = 0. \tag{2.25}$$

From (2.25) it is immediately obvious that the monopolist's choice of quality is *locally* inefficient if and only if the effect of a quality increment on the marginal willingness to pay differs across average and marginal customers

$$\frac{1}{x}\int_0^x P'_q(\tilde{x}, q)d\tilde{x} \neq P'_q(x, q) \quad \text{(local inefficiency condition)}.$$

Therefore, strict monotonicity of P'_q in x is always sufficient for a locally inefficient choice of quality, as summarized by the following proposition.

[21] A sufficient condition is that C is strictly convex and P concave.

Proposition 1 (Local Inefficiency) *The Cournot monopolist endows his output x^M with insufficient quality if $P''_{qx} < 0$ and with excessive quality if $P''_{qx} > 0$.*[22]

Proof Suppose $P''_{qx} < 0$. Evaluated at the Cournot point (x^M, q^M), one has $\pi'_q = 0$ and therefore, by (2.25),

$$\begin{aligned} T'_q &= x(\frac{1}{x}\int_0^x P'_q(\tilde{x}, q)d\tilde{x} - P'_q) \\ &> x(P'_q - P'_q) \quad (\text{by } P''_{qx} < 0) \\ &= 0. \end{aligned}$$

This proves that, at the Cournot point, welfare could be increased if quality were raised. Therefore, quality is locally too low. The proof of the other case is similar. ■

These results indicate that the Cournot monopolist chooses inefficient quality. But they do not tell you anything about the global comparison of q^M and q^{P0}. The latter is also influenced by the gap between the two outputs, x^M and x^{P0}, in addition to the gap between the impact of quality on the marginal willingness to pay of the average and the marginal customer.

Altogether, the global inefficiency gap of quality $(q^M - q^{P0})$ has the same sign as the local gap summarized in the above Proposition if the output gap $(x^M - x^{PO})$ is relatively small. However, if the output gap $(x^M - x^{PO})$ is substantial, the monopolist may actually provide more than the socially optimal level of quality $(q^M > q^{P0})$, even if the monopolist's quality level is locally insufficient because P'_q is strict monotone decreasing in x.

We close with a simple example in which the local and the global inefficiency gap have the same sign.

Example 4 *Suppose $P(x,q) := (1-x)q$ and $C(x,q) = \frac{1}{2}q^2$. Then, both profit and total surplus are strictly concave in x and q. Therefore, (x^M, q^M) and (x^{PO}, q^{PO}) can be found by solving the above first–order conditions.*

$$\begin{aligned} x^M &= \frac{1}{2}, \qquad q^M = \frac{1}{4} \\ x^{PO} &= 1, \qquad q^{PO} = \frac{1}{2}. \end{aligned}$$

[22]This result was observed, for the first time, by Spence, A. M. (1975). "Monopoly, regulation, and quality", *Bell Journal of Economics*, 6: 417–429. See also the follow–up article by Sheshinski, E. [1976]. "Price, quality and quantity regulation in monopoly situations", *Economica*, 143: 127-137.

2.5 Inventory and "Price Rigidity"*

In macroeconomics it is sometimes claimed that monopolies contribute to price rigidity which subsequently aggravates fluctuations in output and employment. We close our discussion of the Cournot monopoly with a few remarks on this topic, inspired by a contribution by Reagan.[23]

We inject three modifications of the basic Cournot model. First, the monopolist is assumed to serve the market repeatedly. Second, demand is taken to be uncertain. Third, production is assumed to take time.[24] For simplicity, the cost function is taken to be linear (constant unit costs c).

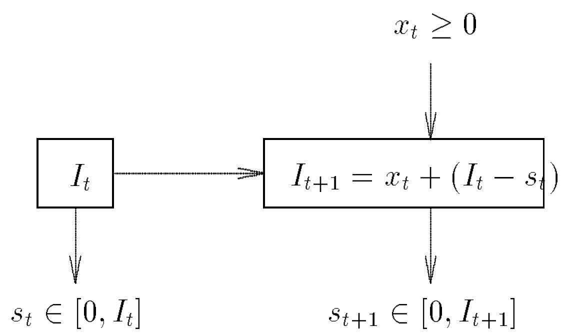

Figure 4: Sales–Output–Inventory (OSI) Dynamics

Take the production period as time unit. Due to the one–period lag between inputs and outputs, current production cannot be used to serve contemporaneous demand. Instead, all sales have to come out of inventory, and period t production is only available for shipment at the beginning of the subsequent period $t+1$. We denote the inventory held at the beginning of period t by I_t, sales in period t by s_t, and the output produced in period t (and available at $t+1$) by x_t. Therefore, feasible Output–Sales–Inventory (OSI) plans $\{(x_t, s_t, I_t)\}$ must satisfy the conditions (for all t)

$$0 \leq s_t \leq I_t \tag{2.26}$$

$$I_{t+1} = x_t + (I_t - s_t) \tag{2.27}$$

$$x_t \geq 0. \tag{2.28}$$

[23] Reagan, P. [1982]. "Inventory and price behavior", *Review of Economic Studies*, 49: 137-142.

[24] One sometimes distinguishes between point vs. flow input or output time structures. In this section, we assume a point–input point–output structure which means that inputs have to be invested at the beginning, and outputs are available at the end of the period.

At the beginning of each period, the monopolist finds out about the current period inverse demand function; but future demand remains uncertain. The demand uncertainty is represented by a parameter θ in the inverse demand function $P(s, \theta)$. The $\theta's$ are independently and identically *(iid)* distributed random variables.

Revenue is monotone increasing in sales, and in the parameter θ, and strictly concave in sales. Moreover, the $\theta's$ are such that the condition which describes the usual Cournot point, $R_1'(s, \theta) = c$, has a positive solution $s > 0$ for all θ.

The monopolist maximizes the discounted expected profit ($1 > \delta$: discount factor)

$$P(s_t, \theta_t)s_t - cx_t + E\left(\sum_{i=t+1}^{\infty} \delta^{i-t}[P(s_i, \theta_i)s_i - cx_i]\right) \tag{2.29}$$

subject to the feasibility constraints (2.26)–(2.28). The solution $\{x_i, s_i, I_i\}$ is the optimal OSI plan.

As a point of reference, consider the following plan to which we will refer as the "price flexibility OSI plan". Under this plan, the monopolist targets a certain inventory k. Once on target (the speed of adjustment depends upon the initial inventory and the random sequence of state of demand realizations)[25], the monopolist sets an inelastic supply of k units in each period ($s_t = k$) and produces just enough to replenish the inventory ($x_t = k$). The target inventory itself is set in such a way that the maximum discounted expected marginal return from having the k-th unit available for sale in the succeeding period equals the marginal cost of producing an additional unit in the present period.

Obviously, the "price flexibility *OSI* plan" is feasible. The nice feature is that it uses the price mechanism to the fullest extent. Output remains constant, and no inventory is ever carried over into the subsequent period. This makes it desirable from a macroeconomic stability perspective. But it is not optimal from the monopolist's point of view.

Optimal OSI Plan The optimal *OSI* plan has the following properties which will be made plausible without a formal proof.[26] The monopolist

[25]In case you wonder about it, the dynamics of adjustment is just like in the optimal plan described below.

[26]The proof uses standard properties of dynamic optimization problems and is spelled out in Reagan [16]. Some of the tools of dynamic optimization will be explained toward the end of this chapter, in our discussion of regulatory mechanism design under asymmetric information.

sets a target inventory k which is exactly the same as the target under the "price flexibility *OSI*", but only sells out inventory up to the point where the marginal cost of replenishing the inventory is equal to the marginal revenue of selling, unless the inventory is binding.

It is useful to distinguish between the dynamics of adjustment to the target inventory, and the steady–state *OSI* plan, implemented once the target has been reached.

Steady–State OSI Plan Once the target inventory has been reached, the optimal *OSI* plan prescribes a constant inventory at the target level k and therefore output equal to sales $x_t = s_t$. This is just like under the "price flexibility *OSI* plan". The difference is between the level of sales (and thus output). Whereas under the "price flexibility *OSI* plan", the inventory is completely turned over in each period, under the optimal *OSI* plan, sales (and thus output) are determined by the rule

$$x_t = s_t = \min\{k, \bar{s}_t\} \tag{2.30}$$

where $\bar{s}_t$ is implicitly defined by the familiar Cournot condition

$$R_1(\bar{s}_t, \theta_t) = c. \tag{2.31}$$

This rule is optimal because, if inventory has to be kept at the constant target level k, the return from selling an extra unit out of current inventory, $R_1(s, \theta)$) has to be weighed against the cost of having to replenish an extra unit of inventory, c.

Evidently, in "high" states of demand when it pays to sell the entire inventory in the present period because of ($\bar{s} \geq k$), the optimal *OSI* plan is just the same as the "price flexibility *OSI* plan"; inventory is completely turned over and output is constant at the rate k. However, in "low" states of demand, if $\bar{s}_t < k$, some of the inventory is carried over into the next period. Essentially, the monopolist speculates on higher future demand and does not deplete the entire inventory in the current period. Consequently, in low states of demand, output is lower and the price higher than under the "price flexibility *OSI* plan".

Dynamic Adjustment Rule Generally, the initial inventory differs from the target level k. Therefore, the optimal *OSI* plan includes an optimal rule to be applied during the adjustment period, until the target inventory is reached. This adjustment rule is easily described and made plausible as a simple modification of the steady–state *OSI* plan.

If the initial inventory is sufficiently low (or the demand state sufficiently high), the dynamic rule prescribes the steady–state level of sales $s_t = \min\{k, \bar{s}_t\}$, but replaces the steady–state output rule ($x_t = s_t$) by the rule

$$x_t = (k - I_t) + \min\{k, \bar{s}_t\}. \tag{2.32}$$

It applies whenever $(k - I_t) + \min\{k, \bar{s}_t\} \geq 0$, in which case the target inventory is reached in just one period.

But if the initial inventory is so high that $(k - I_t) + \min\{k, \bar{s}_t\} < 0$, it is optimal to close the plant, $x_t = 0$, and reduce the inventory by sales s_t to such an extent that the current marginal revenue $R_1(s_t, \theta_t)$ is made equal to the discounted maximum expected marginal return from having the $(I_t - s_t)$-th unit available for sale in the succeeding period. This rule is applied until the target inventory k is reached.

Altogether, the optimal adjustment rule is

$$x_t = \max\{0, k - I_t + \min\{k, \bar{s}_t\}\}. \tag{2.33}$$

And, once the target inventory $I_t = k$ is reached, the firm follows the optimal OSI plan.

Conclusions The bottom line of this dynamic extension of the Cournot monopoly is that the price mechanism is fully used to handle demand fluctuations only in high states of demand. In intermediate and low states, some of the inventory is not sold, but carried over into the next period. Essentially, the monopolist speculates on higher future demand, and output is subsequently lowered and the price is stabilized. This shows how, given exogenous demand fluctuations, the monopolist's optimal policy attenuates price fluctuations, and aggravates fluctuations in output and employment.

Remark 1 *The assumed linearity of the cost function is a crucial ingredient of this story. Why? Suppose, the cost function is strictly convex. Then it is cheaper to produce a certain average output by a constant rather than by a fluctuating output rate. Therefore, convexity of the cost function makes the "price flexibility OSI plan" more favorable.*

3 Price Discriminating or Strong Monopoly

A "strong" monopoly is not restricted to charge all customers the same linear price function, but may price discriminate and set nonlinear prices or discriminate between individuals or groups of customers.

Price discrimination is often defined to be present if the same good is sold at different prices to different customers. In this vein Joan Robinson defined it as "*. . . the act of selling the same article, produced under single control, at different prices to different buyers.*"[27]

However, this definition tends to fail if one interprets the "same good" too loosely. When delivery costs differ, as for example in the delivery of clay bricks, different prices may have nothing to do with price discrimination, whereas equal prices are discriminatory. On the other hand, the definition tends to become void of meaningful applications if one views two goods as the "same" only if they share the same physical characteristics and are available at the same time, place, and state of nature, as is common in general equilibrium theory.

In view of these difficulties it is preferable to adopt a pragmatic notion of price discrimination. The emphasis should be on the monopolist's motive to base the price on customers' willingness to pay rather than simply on cost. The typical price difference between hardcover and paperback books and between first– and second–class flights is a case in point. In both instances the price difference cannot be explained by the difference in cost alone. Instead, it reflects predominantly an attempt to charge buyers according to their willingness to pay. It is this pattern that qualifies observed price differences as discriminatory.

3.1 First–Degree Price Discrimination

First–degree price discrimination occurs if the monopolist charges different prices both across units and across individual customers. Ever since Pigou[28] it has been common to equate first–degree price discrimination with perfect discrimination where the monopolist generates the maximal gain from trade and captures all of it. But this identification leaves out the possibility of imperfect first–degree price discrimination.

A perfectly discriminating monopoly must know the marginal willingness to pay of each potential customer, prevent customers from engaging in arbitrage transactions and unambiguously convey to customers that it is pointless to haggle. Within this general framework the monopolist can succeed with one of two simple pricing schemes.

[27]Robinson, J. [1933]. *The Economics of Imperfect Competition,* Macmillan.

[28]Pigou, A. C. [1920]. *The Economics of Welfare,* Cambridge University Press, introduced the standard distinction between first–, second– and third–degree price discrimination.

One way to go is a customized *take–it–or–leave–it–sales plan* S_i

$$S_i := \{(T_i, x_i), (0,0)\}. \tag{3.1}$$

There, each customer i is offered to either buy the stipulated x_i units for the total price T_i or "leave it" $(0,0)$. To capture the maximum gain, the monopolist only needs to offer the efficient quantities, implicitly defined by the "Price = Marginal Cost" requirement (where $X := \sum_j x_j$)

$$P_i(x_i) = C'(X), \quad \text{for all } i, \tag{3.2}$$

and set T_i equal to $i's$ maximum willingness to pay for x_i.

Another way is to offer customized *two–part tariffs*

$$S_i' := \{(t_i, f_i), (0,0)\} \tag{3.3}$$

that prescribe a certain unit price t_i plus a lump–sum price f_i. In order to capture the maximum gain, t_i needs to be set equal to the marginal cost $C'(\sum_j x_j)$, and f_i at the level where the total price equals $i's$ maximum willingness to pay. Then allow each customer to buy as many units as he likes, unless he chooses the no–buy option $(0,0)$.

Both methods of sale are equivalent as long as complete information and exclusion of arbitrage prevail.

An immediate implication of optimal first–degree price discrimination is that it maximizes social surplus. Hence, one arrives at the somewhat paradoxical conclusion that the strong monopolist gives rise to efficiency of output, whereas the monopolist that exercises restraint and adopts a uniform linear price function contributes to a welfare loss.

However, keep in mind that higher monopoly profits give rise to more wasteful expenditures in the course of the competition for the market. On page 18 we showed that monopoly profit is a good statistic of this underlying waste. This suggests that first–degree price discrimination is the least efficient among all conceivable market forms. True, once a monopoly position has been acquired, perfect price discrimination maximizes the social surplus. But, the entire social gain is completely eaten up by wasteful expenditures in the course of the preceding competition for the market.

3.2 Second–Degree Price Discrimination

Second–degree price discrimination occurs if unit prices vary with the number of units bought, but all customers are subject to the same nonlinear price function. It is by far the most frequently observed form of price discrimination.

For example, many products are sold in different sized packages at a quantity discount, airlines offer frequent flyer bonuses, and public utilities and amusement parks alike charge two–part tariffs. The opposite of a quantity discount — a quantity premium — also occurs. For example, supermarkets often impose a "three—per–customer" rule on discount priced items, where customers are charged the regular price except for the first three units.

The main reason for the popularity of second–degree price discrimination is the lack of precise information about individual customers' willingness to pay. Since it is so important, we cover it in detail in Section 4 and again, at a more advanced level, in Section **??**. The particularly nice feature of models based on incomplete information is that they give an endogenous explanation of second–degree price discrimination.

3.3 Third–Degree Price Discrimination

Third–degree price discrimination occurs if different submarkets are charged different linear price functions. Each customer pays a constant unit price, but unit prices differ across submarkets.

Common examples of third–degree price discrimination are senior citizen and student discounts, lower prices at certain shopping hours, like the "happy hour" in bars and restaurants, the infamous "coupons" in U.S. supermarkets, and intertemporal price discrimination.

Another example is price discrimination across national markets, as in the European car market. Table 2 compares the markup on costs for a sample of cars across some European countries. The spread of markups is remarkably high. The table also suggests that loyalty is a true "luxury"; the Italians are fond of their Fiats and the Germans of their VW Golf (in the U.S. known as the unpopular "rabbit") — and they pay extra for this national brand loyalty.

Of course, the optimal third–degree price discrimination is a straightforward extension of the standard Cournot monopoly solution. It simply extends the Cournot mark–up rule to each and every submarket i:

$$P_i(x_i)\left(\frac{1+\varepsilon_i}{\varepsilon_i}\right) = C'(\sum_j x_j), \tag{3.4}$$

(ε_i : price elasticity of demand in submarket i).

Table 2: Relative Markups of Selected Cars in %, Year 1990

Model	Belgium	France	Germany	Italy	UK
Nissan Micra	8.1	23.1	8.9	36.1	12.5
Fiat Tipo	8.4	9.2	9.0	20.8	9.1
Toyota Corolla	9.7	19.6	13.0	24.2	13.6
VW Golf	9.3	10.3	12.2	11.0	10.0
Mercedes 190	14.3	14.4	17.2	15.6	12.3
BMW 7–series	15.7	15.7	14.7	19.0	21.5

Source: Verboven, F. [1996]. "International price discrimination in the European car market", *Rand Journal of Economics*, 27: 240–268.

Of course, a full account of price discrimination cannot be given in terms of price alone. Pricing is interconnected with product design, quality, and product bundling. Often product design is a prerequisite of price discrimination. As a particularly extreme example, Scherer[29] reports that a manufacturer considered to add arsenic to his industrial plastic molding powder methyl methacrylate in order to prevent its use in dentures manufacture.[30]

Third–degree price discrimination comes–up again in our discussion of the time inconsistency of optimal intertemporal price discrimination, also known as the durable goods monopoly problem, and later in the book when we deal with optimal auctions.

3.4 Limits of Price Discrimination

The possibilities of price discrimination are limited for at least three reasons:

1. arbitrage,
2. hidden information,

[29]Scherer, F. [1970]. *Industrial Market Structure and Economic Performance,* Rand McNally.

[30]This method of separating markets is also common in taxation. For example, diesel fuel can be used to heat your home or to run a Diesel engine powered automobile. In Europe both uses are widespread. Governments wanted to tax engine fuel at a higher rate than heating fuel. This was made feasible by adding a substance to heating fuel that generates an obnoxious fume if burnt in an engine.

3. limited commitment power.

If different prices are charged across units or across customers, arbitrage transactions tend to be profitable. For example, if all customers are charged the same two–part tariff, customers can gain if they buy through an intermediary. This saves participating customers all except one lump–sum fee.

As a rule, the possibility of arbitrage erodes price discrimination. But, due to transaction costs, it does not usually rule it out altogether. The European car market is a case in point. In German newspapers one sees ads for low–priced "reimports" of new German cars. Car manufacturers run their own campaigns to warn potential buyers of reimports from alleged fraud. The combination of fear and bother seems to scare away most customers.

Moreover, there are many products where arbitrage is intrinsically difficult to achieve. Did you never wish you could send someone else to the dentist to have your teeth drilled? There are many examples of products and services where a transfer of ownership is seriously inhibited. Therefore, price discrimination has many applications, and indeed it flourishes in real–world markets.

The second limitation, hidden information, has to do with the fact that the monopolist typically does not know the marginal willingness to pay of each and every customer. The statistical distribution of customers' characteristics may be fairly well known. But when customers walk in, it is difficult to identify their type.

The third limitation has to do with limited commitment power and the credibility of threats. If a monopolist makes a take–it–or–leave–it offer, the "leave–it" threat may pose a problem. Suppose a customer has refused the initial offer and starts haggling. Then the monopolist is tempted to enter negotiations in order to avoid the loss of a profitable customer. When both sides of a transaction gain, both tend to have some bargaining power. The seller can capture the entire gain from trade only if he can make a reliable commitment to always break off negotiations after an offer has been refused. But such commitment is difficult to achieve. In some cases delegation is effective (just try to negotiate the price of a soap bar in a department store), but if the gains are substantial, one can always ask to see the manager.

4 Hidden Information and Price Discrimination

Second–degree price discrimination is the most widely observed kind of price discrimination. We now take a closer look at it and elaborate on a model that explains why this kind of discrimination emerges, and how it should be done.[31]

Consider a profit maximizing monopolist who faces two types of customers with known payoff functions in equal proportions.[32] Arbitrage transactions between customers are not feasible. Therefore, the monopolist may price discriminate.

However, the pricing problem is complicated by the fact that customers' type is their private or hidden information. The monopolist knows all payoff functions; but he cannot tell customers apart (he does not know who is either type 1 and or type 2). Therefore, price discrimination requires a somewhat sophisticated sorting device.

The market game is structured as follows:

The monopolist sets a uniform nonlinear price function, in the form of a menu of price–quantity combinations, (T, x), called "sales plan", from which each customer is free to select one

$$S := \{(T_1, x_1), (T_2, x_2), (0, 0)\}. \tag{4.1}$$

The $(0, 0)$ combination is included because market transactions are voluntary; customers are free to abstain from buying. Of course, $x_1, x_2 \geq 0$.

Customers observe the sales plan and pick that price–quantity combination that maximizes their payoff. Payments are made, and the market game ends.

Without loss of generality, the component (T_1, x_1) is designated for customer 1, and (T_2, x_2) for customer 2 ("incentive compatibility").

Of course, the monopolist could also live with a sales plan where for example customer 2 picks (x_1, T_1) and 1 picks (x_2, T_2), as long as he makes no error in predicting customer's rational choice. But then, incentive compatibility can be restored simply by relabeling the components of the sales plan.

---

[31] Here we present a two–type version of the continuous type model by Maskin, E. and J. G. Riley [1984]. "Monopoly with incomplete information", *Rand Journal of Economics*, 15: 171-196.

[32] Equal proportions are invoked only in order to avoid a glut of notation and obvious case distinctions.

Therefore, the restriction to incentive compatible sales plans is without loss of generality. This is the essence of the well–known "revelation principle".[33]

Assumptions Four assumptions are made:

A1 (Cost function) Unit costs of production are constant and normalized to zero.

A2 (Payoff functions) The monopolist maximizes profit[34]

$$\pi := T_1 + T_2. \tag{4.2}$$

Customers maximize consumer surplus

$$U_i(x, T) := \int_0^x P_i(y)dy - T, \quad \text{for} \quad i = 1, 2, \tag{4.3}$$

where $P_i(x)$ denotes $i's$ marginal willingness to pay for the quantity x.

A3 (Declining Marginal Willingness to Pay) $P_i(x)$ is strict monotone decreasing and $P_i(0) > 0$, $i = 1, 2$.

A4 (Single–Crossing)

$$P_2(x) > P_1(x), \quad \text{for all } x. \tag{4.4}$$

A4 is called "single–crossing" assumption for the following reason: Pick an arbitrary point in (x, T) space, say x_1^*, T_1^*, and draw the two types' indifference curves that pass through this point. Since the slope of indifference curves is equal to $P_i(x)$, A4 assures that these curves cross only once, at this given point, as illustrated in Figure 5.

Optimal Sales Plan The optimal sales plan maximizes π subject to the following *participation constraints*

$$U_1(x_1, T_1) \geq U_1(0, 0) = 0 \tag{4.5}$$

$$U_2(x_2, T_2) \geq U_2(0, 0) = 0 \tag{4.6}$$

---

[33] For an explicit proof of the revelation principle, in the framework of auction theory, see page **??**.

[34] Alternative interpretation: there is only one customer; this customer is either type 1 or type 2 with equal chance; the monopolist maximizes expected profit.

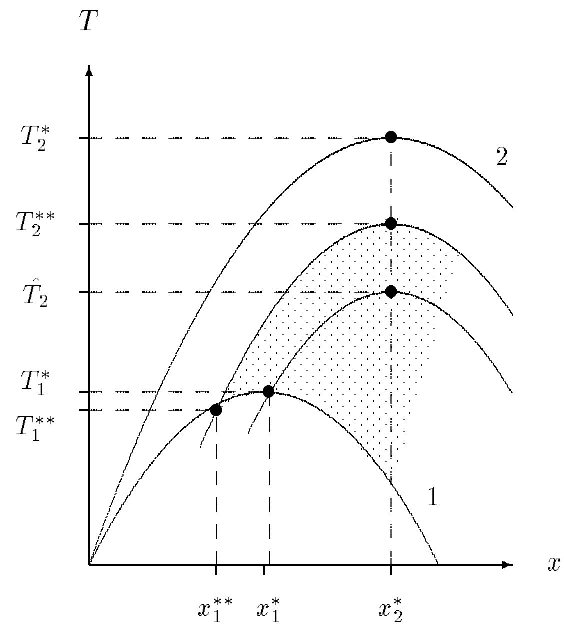

Figure 5: Customers' Indifference Curves

and *incentive constraints*

$$U_1(x_1, T_1) \geq U_1(x_2, T_2) \tag{4.7}$$

$$U_2(x_2, T_2) \geq U_2(x_1, T_1). \tag{4.8}$$

Conditions (4.5) and (4.7) assure that the component of the sales plan designated for customer 1, (x_1, T_1), is neither dominated by the $(0, 0)$ nor by the (x_2, T_2) option. Similarly, conditions (4.6) and (4.8) assure that (x_2, T_2) is neither dominated by $(0, 0)$ nor by (x_1, T_1).

Some Preliminaries Luckily, the program can be simplified by eliminating two constraints. Indeed, among the participation constraints only the lower type's participation constraint (4.5) binds. And among the incentive constraints only the upper type's incentive constraint (4.8) binds.

A constraint does not "bind" if eliminating it from the optimization program does not affect the solution. Therefore, it is claimed that one can eliminate constraints (4.6) and (4.7), without loss of generality.

What makes us come to this conclusion? At this point, just take it as a working hypothesis. Of course, it is only justified if it turns out that the solution of the thus restricted optimization program also satisfies the eliminated constraints — which will be confirmed later on.

Moreover, note that one cannot also eliminate incentive constraint (4.8) or participation constraint (4.5). Because if one also eliminates the incentive constraint (4.8), it is obviously optimal to give each customer the efficient quantity, implicitly defined by the condition $P_i(x_i) = 0$, at a price equal to their maximum willingness to pay $T_i = \int_0^{x_i} P_i(y)dy$ — illustrated by the one–star variables in Figure 5. But in that case type 2 is evidently better–off by choosing (x_1, T_1) in lieu of the designated (x_2, T_2) — violating incentive constraints. And if one eliminates both participation constraints, the monopolist could exploit customers without limit — violating participation constraints.

4.1 Solution of the Restricted Program

The restricted program, restricted by eliminating constraints (4.6) and (4.7), can be further simplified due to the following results.

Lemma 1 *The optimal sales plan exhibits*

$$T_1 = \int_0^{x_1} P_1(y)dy \tag{4.9}$$

$$T_2 = T_1 + \int_{x_1}^{x_2} P_2(y)dy. \tag{4.10}$$

Proof We have noted (but not yet proved) that the upper type's incentive constraint and the lower type's participation constraints are binding. If a constraint binds then it is satisfied with equality, at the optimal sales plan. Therefore, (4.5) entails (4.9). Using this result concerning T_1, when (4.8) and (4.5) bind one has

$$\begin{aligned} 0 &= U_2(x_2, T_2) - U_2(x_1, T_1) \\ &= \int_0^{x_1} P_1(y)dy + \int_{x_1}^{x_2} P_2(y)dy - T_2, \end{aligned} \tag{4.11}$$

which entails (4.10), as asserted. ■

These price functions have a nice interpretation:

1. The low type is charged his maximum willingness to pay for x_1.
2. The high type pays the same for the first x_1 units plus his own maximum willingness to pay for the additional $x_2 - x_1$ units. Therefore, the

high type makes a net gain iff $x_1 > 0$, simply because he obtains the first x_1 units at a bargain price. (Note, this presumes $x_2 \geq x_1$, which we confirm in Proposition 2.)

In view of Lemma 1 we can now eliminate the T variables in the monopolist's objective function and state the "restricted program" in the form of the unconstrained (except for nonnegativity) optimization problem

$$\max_{x_1, x_2 \geq 0} \left(2 \int_0^{x_1} P_1(y) dy + \int_{x_1}^{x_2} P_2(y) dy \right). \tag{4.12}$$

The Kuhn–Tucker conditions of the restricted program are

$$(2P_1(x_1) - P_2(x_1)) \leq 0 \quad \text{and} \quad (\dots)\, x_1 = 0 \tag{4.13}$$

$$P_2(x_2) \leq 0 \quad \text{and} \quad P_2(x_2) x_2 = 0. \tag{4.14}$$

And the $T's$ are obtained by inserting the optimal $x's$ into (4.9), (4.10).

4.2 The Optimal Sales Plan

Proposition 2 *The optimal sales plan exhibits*

$$P_2(x_2) = 0, \quad x_2 > 0 \quad \textit{(no distortion at top)} \tag{4.15}$$

$$x_2 \geq x_1, \quad T_2 \geq T_1 \quad \textit{(monotonicity)} \tag{4.16}$$

$$P_1(x_1) > 0 \quad \textit{(distortion at bottom)} \tag{4.17}$$

$$U_1(x_1, T_1) = 0 \quad \textit{(no surplus at bottom)} \tag{4.18}$$

$$U_2(x_2, T_2) \geq 0 \textit{ with } > \iff x_1 > 0 \tag{4.19}$$

$$\textit{(surplus at top unless } x_1 = 0\textit{)}$$

The optimal prices are computed in (4.9) and (4.10).

Proof First we characterize the solution of the restricted program (4.12) and then show that it also solves the unrestricted program.

1) Obviously, $x_2 > 0$, because if $x_2 = 0$ one would have $P_2(0) \leq 0$ which, however, violates A3. Therefore, inequality (4.14) can be replaced by an equality. We conclude: the high type gets the efficient quantity, $P_2(x_2) = 0, x_2 > 0$.

2) Suppose $x_1 > x_2$, contrary to what is asserted. Since $x_2 > 0$ one has also $x_1 > 0$. Therefore, (4.13) is satisfied with equality, and one has, using

the single–crossing assumption A4,

$$
\begin{aligned}
0 &= 2P_1(x_1) - P_2(x_1) && (4.20)\\
&\leq 2P_2(x_1) - P_2(x_1) && (4.21)\\
&= P_2(x_1) && (4.22)\\
&< P_2(x_2). && (4.23)
\end{aligned}
$$

But this contradicts (4.15) which was already proven in 1). Therefore, $x_2 \geq x_1$. Using Lemma 1, this also implies $T_2 \geq T_1$.

3) If $x_1 = 0$ one has $P_1(x_1) > 0$, by A3. And if $x_1 > 0$, condition (4.14) combined with monotonicity (4.16) entails

$$P_1(x_1) = \frac{1}{2}P_2(x_1) > \frac{1}{2}P_2(x_2) = 0. \tag{4.24}$$

In either case, the low customer gets less than the efficient quantity, $P_1(x_1) > 0$ (distortion at bottom).

4) $U_1(x_1, T_1) = 0$ is obvious from (4.9). And $U_2(x_2, T_2) \geq 0$, with $>$ if $x_1 > 0$, follows immediately from (4.10) and monotonicity.

5) Finally, we need to confirm that the reduced program also satisfies the two omitted constraints (4.6) and (4.7). The omitted participation constraint (4.6) is obviously satisfied by (4.19). And the omitted incentive constraint (4.7) holds for the following reasoning (the last step uses the monotonicity property $x_2 \geq x_1$ and the single–crossing assumption A4)

$$
\begin{aligned}
U_1(x_2, T_2) - U_1(x_1, T_1) &= \int_{x_1}^{x_2} P_1(y)dy - (T_2 - T_1) && (4.25)\\
&= \int_{x_1}^{x_2} \left(P_1(y) - P_2(y)\right) dy && (4.26)\\
&\leq 0. && (4.27)
\end{aligned}
$$

This completes the proof. ■

4.3 Why it Pays to "Distort" Efficiency

Why is it optimal to deviate from efficiency in dealing with the low type but not the high type? The intuition is simple. The high type has to be kept indifferent between (x_2, T_2) and (x_1, T_1). This is achieved by charging the high type the price T_1 for the first x_1 units, and a price equal to his maximum willingness to pay for $x_2 - x_1$ units. From this observation it follows immediately that profit is maximized by expanding x_2 to a level where the marginal willingness to pay equals the marginal cost, $P_2(x_2) = 0$ (see Figure 5). In turn, starting from $P_1(x_1) = 0$ (see point (x_1^*, T_1^*) in that Figure), a small reduction in x_1 is costless in terms of foregone profits from the low type (the marginal profit is zero, at this starting point). But, as a side effect, it extends the domain where the high type is charged a price equal to his maximum willingness to pay. Altogether, it thus pays to introduce a downward distortion at the bottom, illustrated by the two–star variables in Figure 5.

Both Figures 5 and 6 provide illustrations of these considerations. Note, in Figure 6 the shaded area is the high type's consumer surplus (the explanation was already provided in Lemma 1). That surplus is always lowered if one deviates from the efficient level of x_2.

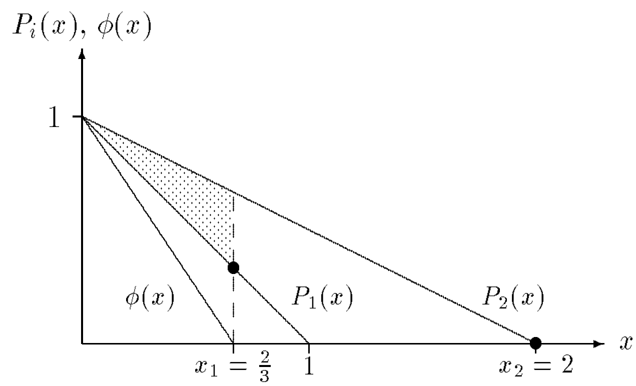

Figure 6: Optimal Sorting with Two Customers if $P_i(x) := 1 - \frac{1}{i}x$

4.4 Sorting, Bunching, and Exclusion

Finally, note that it is not always optimal to serve both customer and discriminate between them. Depending upon the properties of inverse demand functions, it may be optimal to either not serve the low type at all, and exclusively serve the high type, or treat both types the same.

Altogether, the optimal price discrimination falls into either one of three categories:

1. "Sorting" or true discrimination, with $0 < x_1 < x_2$, $T_1 < T_2$.
2. "Bunching" or no discrimination, with $x_1 = x_2 > 0$, $T_1 = T_2$.

3. "Exclusion", where only the high type is served at a price equal to his maximum willingness to pay, $0 = x_1 < x_2$, $T_2 > T_1 = 0$.

Note carefully, that Proposition 2 excludes neither case.

Example 5 *Here we illustrate that all three cases may occur.*

1. *Suppose* $P_i(x) := 1 - x/i, i = 1, 2$. *Then, the optimal price discrimination exhibits sorting, with* $x_1 = \frac{2}{3}, T_1 = \frac{4}{9}$, $x_2 = 2, T_2 = \frac{8}{9}$, *and, incidentally, a declining unit price (quantity discount)* $\frac{T_2}{x_2} = \frac{4}{9} < \frac{6}{9} = \frac{2}{3} = \frac{T_1}{x_1}$.

2. *Suppose* $P_i(x) := \theta_i(1 - x), i = 1, 2$, *and* $1 = \theta_1 < \theta_2 \neq 2$. *Then, it is optimal to abstain from discrimination (bunching). Specifically,* $x_1 = x_2 = 1, T_1 = T_2 = 2$.

3. *Suppose* $P_i(x) := i - x, i = 1, 2$. *Then, it is optimal to serve only the high type (exclusivity) and take away his entire surplus* $x_1 = T_1 = 0, x_2 = 2$, $T_2 = 2$.

Digression: Two–Part Tariffs A somewhat less effective price discriminating scheme is to offer a menu of two–part tariffs. A two–part tariff is an affine price function

$$T_i(x) := t_i x + f_i, \tag{4.28}$$

with the constant unit–price component t_i and the lump–sum component f_i. This pricing scheme is observed in many regulated industries, for example in public utilities and in the taxi business.

The two–part tariff discriminating monopolist offers a sales plan

$$S' := \{(t_1, f_1), (t_2, f_2), (0, 0)\}, \tag{4.29}$$

asks each customer to pick one component, and then lets each customer buy as many units as he wishes, unless he chose the no–buy option $(0, 0)$.

Using the revelation principle, one can again restrict attention to prices that satisfy the corresponding participation and incentive constraints, and then compute the optimal two–part tariffs. Again only the low type's participation and the high type's incentive constraint bind. Therefore, the solution procedure is quite similar.

Two–part tariffs are evidently less effective. The profit earned with a menu of two–part tariffs can always be replicated by appropriately chosen price–quantity combinations, but not *vice versa*. You may wish to confirm this by computing the optimal two–part tariffs for the $P_i(x)$ functions assumed in the complete sorting case of Example 5 and show that two–part tariffs are less profitable.

Generalization* The above generalizes in a straightforward manner to $n \geq 2$ types with the single–crossing marginal willingness to pay functions

$$P_1(y) < P_2(y) < \ldots < P_n(y). \tag{4.30}$$

In particular, if complete sorting is optimal one can show that the optimal price discrimination exhibits

1. Zero consumer surplus for the lowest type only.
2. No distortion at the top only.
3. Only local downward incentive constraints bind (customer $i \geq 2$ is indifferent between (T_i, x_i) and (T_{i-1}, x_{i-1}); all other price–quantity combinations in the optimal sales plan are inferior).

Moreover, the optimal sales plan is then completely characterized by the following rules

$$(n+1-i)P_i(x_i) = (n-i)P_{i+1}(x_i), \quad i \in \{1, \ldots, n-1\} \tag{4.31}$$

$$P_n(x_n) = 0 \tag{4.32}$$

$$T_1 = \int_0^{x_1} P_1(y)dy \tag{4.33}$$

$$T_i = T_{i-1} + \int_{x_{i-1}}^{x_i} P_i(y)dy, \quad i \in \{2, \ldots, n\}. \tag{4.34}$$

The proof of these assertions is a fairly straightforward extension of the above analysis of the two types case.

5 Generalization*

In the previous section we showed how hidden information may give rise to second–degree price discrimination. There we covered the simplest case of two types of customers. Here, in this section, we extend this analysis to a continuum of customers and spell out some fairly general properties optimal mechanism design. You can either read this now, if you are ready to exercise some more advanced techniques, or come back to it at a later stage.[35]

[35] We follow the seminal contribution by Maskin, E. and J. G. Riley [1984]. "Monopoly with incomplete information", *Rand Journal of Economics*, 15: 171-196. Some proofs are simpler, and the exposition is more accessible.

The basic notation and assumptions of Section 4 are maintained. However, the monopolist now faces a continuum of customer types from the type space $\Theta := [0, \bar{\theta}]$, $\bar{\theta} > 0$. Customers' type is their private or hidden information. The monopolist only knows that types are drawn from the probability distribution $F : \Theta \to [0, 1]$, which is common knowledge.

From the *revelation principle* we know that any sales plan is equivalent to a *direct incentive compatible sales plan.*[36] Accordingly, the monopolist announces a sales plan composed of the functions $x(\theta)$ (quantity) and $T(\theta)$ (total payment):

$$S := (x(\theta), T(\theta)), \tag{5.1}$$

and invites customers to choose a combination of quantity and total payment, $(x(\tilde{\theta}), T(\tilde{\theta}))$ by self–declaring their type $\tilde{\theta} \in \Theta$. Since the monopolist does not know customers' type, they may cheat and declare some $\tilde{\theta} \neq \theta$. However, an incentive compatible sales plan induces all customers not to cheat and declare their true type, $\tilde{\theta} = \theta$.

As in Section 4 customers' payoff function is

$$U(x, T; \theta) := W(x; \theta) - T. \tag{5.2}$$

Thereby, $W(x; \theta)$ denotes *θ's willingness to pay* for the quantity x

$$W(x; \theta) := \int_0^x P(y; \theta) dy \tag{5.3}$$

and $P(x; \theta) = \partial W / \partial x$ the associated *marginal willingness to pay.* Customers maximize their payoff by self–declaring their type.

Assuming constant unit costs that are normalized to zero (as in Section 4), the monopolist's profit from a certain customer type is that type's total payment $T(\theta)$. Therefore, the optimal sales plan maximizes the monopolist's expected profit (where $F(\theta)$ is the probability distribution (c.d.f.) of types)

$$\max_{\{x(\theta), T(\theta)\}} \int_0^{\bar{\theta}} T(\theta) dF(\theta), \tag{5.4}$$

subject to the following *participation* (6.5) and *incentive constraints* (6.6):

$$U(x(\theta), T(\theta); \theta) \geq U(0, 0; \theta) = 0, \quad \forall \theta \in \Theta \tag{5.5}$$

$$\theta \in \arg\max_{\tilde{\theta}} U(x(\tilde{\theta}), T(\tilde{\theta}); \theta), \quad \forall \tilde{\theta}, \theta \in \Theta. \tag{5.6}$$

[36] The revelation principle was stated and proved on page ??.

Further Assumptions In Section 4 we assumed that marginal willingness to pay functions are decreasing and single–crossing and that unit costs are constant. We now adapt and strengthen these assumptions and add assumptions concerning the probability distribution:

A1 $P(x;\theta)$ is twice continuously differentiable with the partial derivatives $P_1 < 0$, $P_2 > 0$, $P_{22} < 0$, and unit costs are constant and normalized to zero.

A2 The price elasticity of demand is increasing in θ; stated in terms of inverse demand P:

$$x, P > 0 \Rightarrow \frac{\partial}{\partial\theta}\left(\frac{-x}{P}\frac{\partial P}{\partial x}\right) < 0.$$

A3 The c.d.f of θ, $F(\theta)$, is continuously differentiable with $F'(\theta) > 0$ everywhere on Θ and $F(0) = 1 - F(\bar{\theta}) = 0$.

5.1 Auxiliary Results

Fortunately, the optimization problem can be simplified, due to the following auxiliary results.

Let $U^*(\theta)$ denote the indirect utility function of customer θ:

$$U^*(\theta) := \max_{\tilde{\theta}\in\Theta} U(x(\tilde{\theta}), T(\tilde{\theta}); \theta).$$

Lemma 2 (Monotonicity) *The sales plan S is incentive compatible if and only if:*

$$U^{*'}(\theta) = \frac{\partial W}{\partial\theta}(x(\theta);\theta) \geq 0 \tag{5.7}$$

$$\textit{and} \quad x'(\theta) \geq 0. \tag{5.8}$$

Proof 1) Necessity: Applying the envelope theorem to the customers' maximization problem (6.6) gives

$$U^{*'}(\theta) = \frac{\partial W}{\partial\theta}(x(\theta);\theta),$$

which is positive because $P_2 > 0$. Note, (6.7) incorporates the first–order condition of customers' maximization problem.

Moreover $U^*(\theta)$ is convex;[37] therefore,

$$\begin{aligned} 0 &\leq U^{*''}(\theta) \\ &= \frac{\partial^2 W}{\partial x \partial \theta}(x(\theta);\theta)x'(\theta) + \frac{\partial^2 W}{(\partial \theta)^2}(x(\theta);\theta) \\ &= P_2(x(\theta);\theta)x'(\theta) + P_{22}(x(\theta);\theta). \end{aligned}$$

And since $P_{22} \leq 0$ and $P_2 > 0$, we conclude: $x'(\theta) \geq 0$.

2) The proof of sufficiency is by contradiction.[38] Consider a sales plan that satisfies the two conditions, yet type θ_1 prefers to "cheat" and declare $\theta_2 \neq \theta_1$.

Define customers' payoff as a function of their true and declared type $(\theta, \tilde{\theta})$, for a given sales plan, as: $\bar{U}(\theta, \tilde{\theta}) := U(x(\tilde{\theta}), T(\tilde{\theta}); \theta)$. Then, $\bar{U}(\theta_1, \theta_2) > \bar{U}(\theta_1, \theta_1)$, and after a few tautological rearrangements, also using the first–order condition $\bar{U}_2(\theta, \theta) \equiv 0$ (third line), one obtains:

$$\begin{aligned} 0 &< \bar{U}(\theta_1, \theta_2) - \bar{U}(\theta_1, \theta_1) \\ &= \int_{\theta_1}^{\theta_2} \bar{U}_2(\theta_1, z)dz \\ &= \int_{\theta_1}^{\theta_2} \left(\bar{U}_2(\theta_1, z) - \bar{U}_2(z, z)\right) dz \\ &= \int_{\theta_1}^{\theta_2} \int_{z}^{\theta_1} \bar{U}_{12}(r, z)drdz \\ &= \int_{\theta_1}^{\theta_2} \int_{z}^{\theta_1} P_2(x(z); r)x'(z)drdz. \end{aligned}$$

Let $\theta_2 > \theta_1$; then, $z \geq \theta_1$ for all $x \in [\theta_1, \theta_2]$ and the inequality cannot hold. Similarly, let $\theta_1 > \theta_2$; then, $z \leq \theta_1$ for all $z \in [\theta_2, \theta_1]$, which is again a contradiction. ■

The monotonicity of $x(\theta)$ is illustrated by the indifference curves of two distinct types, $\theta_2 > \theta_1$, in Figure 8. There, if type θ_1 gets $(x(\theta_1), T(\theta_1))$ the price–quantity combination offered to type θ_2 must be in the shaded area, above $\theta_1's$ and below $\theta_2's$ indifference curves that pass through $(x(\theta_1), T(\theta_1))$. Hence, $x(\theta_2)$ must be either the same as $x(\theta_1)$ or exceed it.

[37] The proof of convexity is a standard exercise in microeconomics; it was sketched before on page ??.

[38] The proof is adapted from Laffont, J.–J. and J. Tirole [1993]. *A Theory of Incentives in Procurement and Regulation,* MIT Press, p. 121.

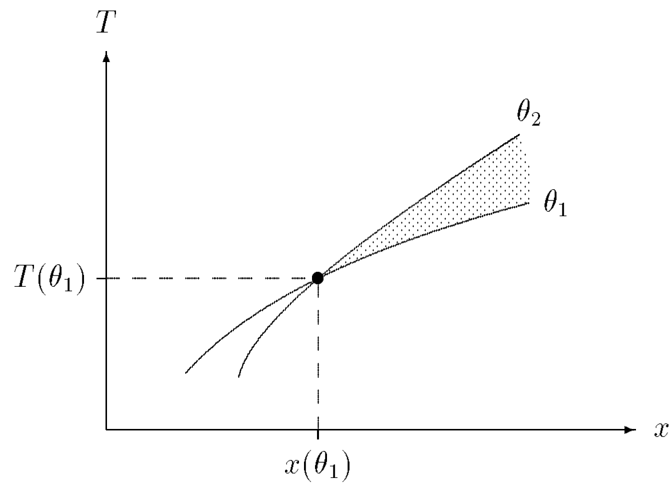

Figure 7: Incentive Compatiblity and the Monotonicity of $x(\theta)$

Lemma 3 *An incentive compatible sales plan satisfies all participation constraints if and only if* $U^*(0) \geq 0$.

Proof Follows immediately by the monotonicity of $U^*(\theta)$. ∎

The monotonicity of $x(\theta)$ implies that $x(\theta)$ is differentiable almost everywhere. In the following we consider the smaller class of piecewise differentiable functions. This is a prerequisite of using the calculus of variations or optimal control.[39]

Lemma 4 (Total payment) *Consider a nondecreasing function* $x(\theta)$. *Incentive compatibility is assured if and only if:*

$$T(\theta) = W(x(\theta);\theta) - \int_0^\theta \frac{\partial W}{\partial z}(x(z);z)\,dz - U^*(0). \tag{5.9}$$

Proof Since $x(\theta)$ is nondecreasing, incentive compatibility is assured if and only if $U^{*'}(\theta)$ is set as in (6.7). Integrate $U^{*'}(\theta)$, and one obtains:

$$U^*(\theta) \equiv \int_0^\theta U^{*'}(z)dz + U^*(0) = \int_0^\theta \frac{\partial W}{\partial z}(x(z);z)dz + U^*(0).$$

Solving $U^*(\theta) = W(x(\theta);\theta) - T(\theta)$ for $T(\theta)$ gives (6.9). ∎

[39] A piecewise differentiable function has continuous derivatives almost everywhere, and when it has no derivative it has always left and right derivatives.

Proposition 3 *The monopolist's decision problem is equivalent to the variational problem:*

$$\max_{\{x(\theta)\}}\{\int_0^{\bar\theta}\Pi(x(\theta);\theta)dF(\theta)\mid x'(\theta)\geq 0\} \tag{5.10}$$

$$\textit{where}\quad \Pi(x(\theta);\theta):=W(x(\theta);\theta)-\frac{1}{h(\theta)}\frac{\partial W}{\partial\theta}(x(\theta);\theta) \tag{5.11}$$

$$\textit{and}\quad h(\theta):=\frac{F'(\theta)}{1-F(\theta)}\quad \textit{(hazard rate)}; \tag{5.12}$$

the associated $T(\theta)$ is obtained from (6.9) together with $U^*(0)=0$.

Proof The monopolist maximizes the expected value of $T(\theta)$. By the above Lemmas, the incentive and participation constraints are equivalent to choosing $T(\theta)$ as in (6.9) combined with $U^*(0)\geq 0$, $x'(\theta)\geq 0$.

Leaving a surplus to the lowest type, $U^*(0)>0$, is costly. Therefore, the monopolist sets $U^*(0)=0$. Together with (6.9) the expected profit is hence equal to:

$$\int_0^{\bar\theta}W(x(\theta);\theta)dF(\theta)-\int_0^{\bar\theta}\int_0^{\theta}\frac{\partial W}{\partial z}(x(z);z)dzdF(\theta). \tag{5.13}$$

Apply integration by parts to the second term and substitute the hazard rate $h(\theta)$, defined in (6.12). After a few rearrangements this gives:

$$\int_0^{\bar\theta}\int_0^{\theta}\frac{\partial W}{\partial z}(x(z);z)dzdF(\theta)=\int_0^{\bar\theta}\frac{\partial W}{\partial\theta}(x(\theta);\theta)\frac{1}{h(\theta)}dF(\theta). \tag{5.14}$$

Finally, insert (6.14) into (6.13), and one has the program stated in the proposition. ∎

5.2 Complete Sorting Solution

Consider the program obtained from (6.10) by omitting the monotonicity constraint $x'(\theta)\geq 0$. Call it *restricted program* and denote its solution by $\bar x(\theta)$, $\bar T(\theta)$.

Lemma 5 (Restricted Program) *The solution of the restricted program, $\bar x(\theta)$ is characterized by the conditions*

$$\bar x(\theta)>0\Rightarrow P(\bar x(\theta);\theta)=\frac{1}{h(\theta)}P_2(\bar x(\theta);\theta). \tag{5.15}$$

It exhibits distortions, $P(\bar{x}(\theta);\theta) > 0$ everywhere on Θ, except at the top where $P(x(\bar{\theta});\bar{\theta}) = 0$ ("no distortion at the top").

Proof In a variational problem the principle of optimality is that of "pointwise optimality" (Euler equation).[40] In other words, $\bar{x}(\theta)$ maximizes Π for every θ. Hence, either $\bar{x}(\theta) = 0$ or $\bar{x}(\theta)$ satisfies the first–order condition $\partial\Pi/\partial x = 0$. Therefore,

$$\bar{x}(\theta) > 0 \Rightarrow P(\bar{x}(\theta);\theta) - \frac{1}{h(\theta)} P_2(\bar{x}(\theta);\theta) = 0. \tag{5.16}$$

Note, $P_2 > 0$, $h \geq 0$, and $1/h = 0 \iff \theta = \bar{\theta}$. Therefore, $P(\bar{x}(\theta);\theta) > 0$ everywhere except at $\theta = \bar{\theta}$ where $P(\bar{x}(\bar{\theta});\bar{\theta}) = 0$. ■

We now show that the solution of the restricted program also solves the full program, provided the probability distribution function satisfies a certain concavity condition. That condition is weaker than the assumption of a monotone increasing hazard rate that plays a prominent role in the incomplete information literature.

Proposition 4 (Complete Sorting) *Suppose the probability distribution function $F(\theta)$ is log–concave.*[41] *Then, the solution of the restricted program, $\bar{x}(\theta)$, also solves the full program (6.10).*

Proof F is log–concave if and only if

$$\rho(\theta) := \theta - \frac{1}{h(\theta)} \tag{5.17}$$

is monotone increasing. Since $\bar{x}(\theta)$ maximizes $\Pi(x(\theta);\theta)$ either $\bar{x}(\theta) = 0$ or $\frac{\partial\Pi}{\partial x}(\bar{x}(\theta);\theta) = 0$. Totally differentiating the latter gives:

$$\bar{x}'(\theta) = -\frac{(\partial^2\Pi)/(\partial x \partial\theta)}{(\partial^2\Pi)/(\partial x)^2}. \tag{5.18}$$

[40] For an introductory exposition of the calculus of variations and optimal control consult Kamien, M. I. and N. L. Schwartz [1991]. *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, Second Edition, North–Holland.

[41] F is *log–concave* if $\ln(F)$ is concave. A monotone increasing hazard rate $h(\theta)$ is sufficient (though not necessary) for log–concavity. Many standard distributions are log–concave; examples are normal, uniform, and exponential distributions. See Bagnoli, M. and T. Bergstrom [1989]. "Log–concave probability and its applications", *Working Paper*, University of Michigan.

Therefore, if at $\bar{x}(\theta)$ $(\partial^2\Pi)/(\partial x)^2 < 0$ and $(\partial^2\Pi)/(\partial x\partial\theta) \geq 0$, the solution of the restricted program $\bar{x}(\theta)$ also solves the full program.

From now on all terms are evaluated at $\bar{x}(\theta) > 0$. By (6.15) one has $1/h = P/P_2$. Compute $(\partial^2\Pi)/(\partial x)^2 = P_1 - P_{21}/h$, substitute $1/h$, and use the monotonicity of the elasticity of demand (A2), and one confirms:

$$\frac{\partial^2\Pi}{(\partial x)^2} = P_1 - \frac{P_{21}P}{P_2} = \frac{P^2}{xP_2}\frac{\partial}{\partial\theta}\left(-\frac{xP_1}{P}\right) < 0.$$

Similarly, compute $(\partial^2\Pi)/(\partial x\partial\theta) = P_2 - (P_{22}h - P_2h')/h^2$, substitute ρ, and utilize the monotonicity of ρ. And after a few rearrangements (also using $P_2 > 0$, $P_{22} \leq 0$) one confirms:

$$\frac{\partial^2\Pi}{(\partial x\partial\theta)} = P_2\left(\rho' - \frac{P_{22}}{P_2}\frac{1}{h}\right) \geq 0.$$

■

Note, in Section 4 we showed that the optimal second degree price discrimination may also exhibit bunching. Evidently, the fairly strong assumptions imposed in this section exclude bunching over a measurable set of types. However, if these assumptions are weakened bunching occurs in some subsets of Θ. In that case, one must reintroduce the monotonicity constraint into the optimization program and state it as an optimal control problem. Optimal control is better suited to deal with such inequality constraints than the calculus of variations.

5.3 Implementation by Nonlinear Pricing

At first glance the direct incentive compatible sales plan discussed here seems hopelessly unrealistic. Or have you ever seen a monopolist playing a direct revelation game with customers? Therefore, it is important to realize that the frequently observed second–degree price discrimination, where all customers are charged the same nonlinear price function and are free to buy as much as they want, is nothing but a direct incentive compatible sales plan in disguise.

Proposition 5 (Second–Degree Price Discrimination) *The optimal incentive compatible sales plan is equivalent to second–degree price discrimination with the following nonlinear price function:*

$$\tilde{T}(x) := T(\phi(x)) = \int_0^x P(z;\phi(z))dz, \tag{5.19}$$

where $\phi(z)$ is the (generalized) inverse of $x(\theta)$:

$$\phi(z) := \min\{\theta \mid x(\theta) = z\}. \tag{5.20}$$

Proof We apply a change of variables, using the inverse function $\phi := x^{-1}$. That inverse exists since $x(\theta)$ is monotone.

Start from the optimal $T(\theta)$ function (6.9) using $U^*(0) = 0$, and employ the *change of variable theorem*[42] (first line). After a few more manipulations one obtains:

$$\begin{aligned} T(\phi(x)) &= W(x;\phi(x)) - \int_0^x \frac{\partial W}{\partial(\phi(z))}(z;\phi(z))\phi'(z)dz \\ &= W(x;\phi(x)) - \int_0^x \left(\frac{dW}{dz}(z;\phi(z)) - \frac{\partial W}{\partial z}(z;\phi(z))\right) dz \\ &= \int_0^x \frac{\partial W}{\partial z}(z;\phi(z))dz \\ &= \int_0^x P(z;\phi(z))dz, \end{aligned}$$

as asserted. ■

An interesting follow–up problem is to look for necessary and sufficient conditions for the frequently observed quantitity discounts in nonlinear pricing. Quantity discounts are equivalent to strict concavity of the price function $\tilde{T}(x)$.[43]

6 Price Discrimination and Public Goods*

Monopolies tend to price discriminate but not all price discrimination is due to monopoly. Public goods are a case in point. There, price discrimination is a prerequisite for an efficient allocation, even if the supplier does not exercises monopoly power.

Many public goods happen to be supplied by monopolies. Therefore, an introduction to monopoly should include a few remarks on the relationship between price discrimination and public goods.

Usually we consider *private goods* which are defined by two properties:

[42] See Bartle, R. G. [1976]. *Elements of Real Analysis*, Wiley & Sons, p. 234, or in any other good calculus text.

[43] Some sufficient conditions are in Maskin and Riley.

1. *rivalry in consumption:* a good consumed by one agent is no longer available to others;

2. *excludability.*

However, some goods are *nonrival* in the sense that one person's consumption does not reduce the amount available to others. And some goods are *nonexcludable* because it is impossible or too costly to exclude customers. Goods that are both nonrival and nonexcludable are often called *common goods.* However, it is useful to distinguish more narrowly between *public goods* and *club goods,* as in the classification in Table 3.

Table 3: Private, Public and Club Goods

	Rivalry	
	yes	no
Excludability yes	private good	club good
Excludability no	no name	public good

A good example of a nonrival good is the reception of a radio or TV broadcast. In the absence of exclusion devices one has a *public good.* But if the broadcast is coded and only those who have a decoder can listen to or watch the broadcast, one has a *club good.*

Lindahl Prices: Price Discrimination without Monopoly Suppose there are two types of customers, as in Section 4, except that the good is public rather than private, and suppose marginal cost is constant at the rate $0 < c < \min\{P_1(0), P_2(0)\}$. What allocation is welfare optimal? And can it be implemented by a set linear prices? These questions are standard exercises in (undergraduate) microeconomics. The answers are:

1. Choose that output x^* at which the *sum* of customers' marginal willingness to pay is equal to marginal cost:[44]

$$P_1(x^*) + P_2(x^*) = c.$$

[44] The efficiency conditions for public goods were introduced by Samuelson, P. A. [1954]. "The pure theory of public expenditure", *The Review of Economics and Statistics,* 64:387–389.

2. Charge each customer a unit price p_i equal to this customer's marginal willingness to pay *(Lindahl prices)*[45]:

$$p_i = P_i(x^*).$$

An illustration is provided in Figure 7.

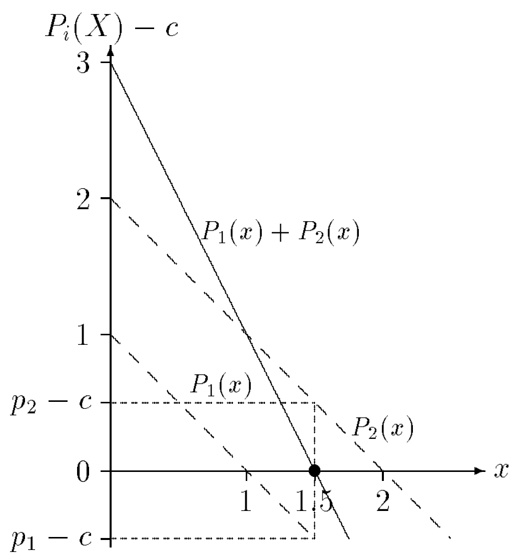

Figure 8: Lindahl Prices for $P_i(x) := i - x$, $i = 1, 2$

First–Degree Price Discrimination Suppose the public good is provided by a monopoly. If the monopolist has complete information concerning customers' marginal willingness to pay and if arbitrage transaction among customers are excluded, first–degree price discrimination is feasible. As you can easily confirm, the profit maximizing monopolist implements the welfare optimum by setting two–part tariffs with a unit price equal to the Lindahl–price and a lump–sum fee equal to customer's maximum consumer surplus. Therefore, Lindahl–prices are also part of first–degree price discrimination.

Second–Degree Price Discrimination If the monopolist operates under incomplete information, as in the Section 4 on hidden information and price

[45]Lindahl prices were introduced by Lindahl, [1919]. "Just taxation — a positive solution", in: Musgrave, R. and A. Peacock (eds.), *Classics in Theory of Public Finance.* London (Macmillan).

discrimination, he cannot price discriminate because he cannot induce sorting by supplying different price/quantity packages. Therefore, the monopolist either supplies both customers or exclusively customer 2, whichever is more profitable, and earns the maximum profit:

$$\Pi := \max\{\max_x 2\int_0^x P_1(y)dy - cx, \max_x \int_0^x P_2(y)dy - cx\}.$$

However, if the good is a *club good* because exlusion is feasible second–degree price discrimination comes back. And the optimal price discrimination scheme is characterized by the sales plan $S := \{(T_1, x_1), (T_2, x_2)\}$ that solves

$$\max_{x_1,x_2} \left(2\int_0^{x_1} P_1(y)dy + \int_{x_1}^{x_2} P_2(y)dy - c\max\{x_1, x_2\}\right)$$

$$T_1 : = \int_0^{x_1} P_1(y)dy$$

$$T_2 : = T_1 + \int_{x_1}^{x_2} P_2(y)dy.$$

7 Intertemporal Price Discrimination

Third–degree price discrimination is frequently observed in markets for consumer durables in the form of intertemporal price discrimination. For example, when new computer hardware, digital video–telephone receivers, or even new books are introduced, the seller typically starts out with a high unit price in order to skim the impatient high demand customers, and then gradually lowers the price.

The rules of optimal third–degree price discrimination are a straightforward extension of the Cournot monopoly rule, which was already summarized on page 28. However, the application to intertemporal price discrimination for durable goods poses an intriguing new problem: that of the time inconsistency of optimal plans. In the face of this problem the standard formula for optimal third–degree price discrimination needs to be modified along the following lines.

Time Inconsistency of Optimal Plans An optimal plan that calls for a sequence of actions is called "time inconsistent" if, in the course of time, one can gain by deviating from it. Essentially such gains may occur if others rely in an important way on the execution of that plan and the decision maker takes advantage of breaching that "trust".

If an optimal sequence of actions is time inconsistent, inflexibility and the power to make a commitment to stick to these actions become valuable. However, such commitment is difficult to achieve.

The problem of the time inconsistency of optimal plans and the search for commitment mechanisms is a matter of concern in many decision problems (think of "smoking" and the notorious difficulties of quitting). It already engaged the attention of Greek mythology. In his famous story of *"Odysseus and the Sirens"*, Homer reports of two Sirens — creatures half bird and half woman — on the rocks of Scylla who lured sailors to destruction by the sweetness of their song. The Greek hero Odysseus escaped the danger of their song by stopping the ears of his crew with wax so that they were deaf to the Sirens. Yet he was able to hear the music and had himself tied to the mast so that he could not steer the ship out of course.

7.1 Durable Goods Monopoly

We now consider the simplest possible durable goods monopoly problem with a time horizon of only two periods. Although the exposition is kept at an elementary level, the raised issues show up in general settings and have a bearing on many other topics in economics, from oligopoly to monetary policy.

Suppose a monopolist produces a durable good that can be used for two periods. The average cost is constant in output and time and thus normalized to zero. There is no wear and tear. The durable good can only be sold but not rented; leasing is not feasible. The monopolist is free to deviate from an earlier plan and cannot commit to a sequence of outputs or prices.

The demand side of the market is characterized by the time invariant inverse *user demand* (or *marginal willingness to pay for use*) per period $t \in \{1, 2, \dots\}$, defined on the consumption flow of the durable good in period t, $X_t \in [0, 1]$:

$$P(X_t) := \begin{cases} 1 - X_t & \text{for } t = 1, 2 \\ 0 & \text{for } t > 2. \end{cases} \tag{7.1}$$

Note: $P(X)$ is a marginal willingness for consumption or use, not ownership, and customers are interested in this good only during the first two periods.

Denote the monopolist's outputs by x_1, x_2 and normalize the rate of consumption per unit of the durable good to be equal to one. Then, since there is no wear and tear, the consumption flow available in each period bears the following relationship to past outputs:

$$X_1 := x_1, \quad X_2 := x_1 + x_2. \tag{7.2}$$

Customers can only buy, not rent. Buying in period 1 provides a consumption flow in two periods, whereas buying in period 2 provides valuable consumption only in period 2. From these facts one can compute the *inverse demand for ownership,* as follows.

Inverse Ownership Demand In period 2, the marginal willingness to pay or inverse ownership demand, P_2, coincides with the marginal willingness to pay for use since the durable good is useless thereafter. Using the stock flow conversion assumption (7.2) this entails

$$P_2 = P(X_2) = P(x_1 + x_2) = 1 - x_1 - x_2. \tag{7.3}$$

In period 1, the marginal willingness to pay for ownership, P_1, takes into account that the acquired good provides services during two periods. Using the discount factor $\delta \in (0, 1]$ and assumption (7.2), one obtains the inverse ownership demand in period 1:

$$\begin{aligned} P_1 :&= P(X_1) + \delta P(X_2) \\ &= P(x_1) + \delta P(x_1 + x_2) \\ &= (1 - x_1) + \delta(1 - x_1 - x_2). \end{aligned} \tag{7.4}$$

Evidently, in order to make the right demand decision in period 1, customers have to correctly predict the monopolist's output in period 2.

Interpretation Many students who have difficulties with this computation of the inverse ownership demand P_1, find the following interpretation useful.

Imagine those who buy the durable good in period 1 sell it at the end of the period and then buy again the quantity they want to consume, at zero transaction costs. Since new and old durable goods are perfect substitutes, the unit price at which the used good is sold at the end of period 1 is equal to $P(x_1+x_2)$. Therefore, the inverse demand for ownership in period 1 must be equal to the marginal willingness to pay for use in period 1 plus the present value of the price earned from selling it at the end of that period which gives $P_1 = P(x_1) + \delta P(x_1 + x_2)$, as asserted.

Payoff Functions After these preliminaries it follows immediately that the monopolist's present values of profits in periods 1 and 2 are

$$\begin{aligned}\pi_2 &:= P_2(x_1, x_2)x_2 \\ &= (1 - x_1 - x_2)x_2 \qquad (7.5)\\ \pi_1 &:= P_1(x_1, x_2)x_1 + \delta\pi_2 \\ &= (1 - x_1 + \delta(1 - x_1 - x_2))\, x_1 \\ &\quad + \delta(1 - x_1 - x_2)x_2. \qquad (7.6)\end{aligned}$$

7.2 Time Inconsistency Problem

The optimal output plan maximizes the present value of profits

$$\max_{x_1, x_2 \geq 0} \pi_1(x_1, x_2). \qquad (7.7)$$

Proposition 6 (Optimal Price Discrimination) *The optimal output plan and associated third–degree price discrimination and maximum present value of profits is*

$$x_1 = 0.5, \quad x_2 = 0 \qquad (7.8)$$

$$p_1 = 0.5 + \delta p_2, \quad p_2 = 0.5 \qquad (7.9)$$

$$\pi_1 = 0.25(1 + \delta). \qquad (7.10)$$

Proof The objective function is strictly concave. Therefore, the solution is uniquely characterized by the Kuhn–Tucker conditions

$$((1 - 2x_1)(1 - \delta) - \delta 2x_2)) \leq 0 \quad \text{and} \quad (\ldots)x_1 = 0 \qquad (7.11)$$

$$(1 - 2x_1 - 2x_2) \leq 0 \quad \text{and} \quad (\ldots)x_2 = 0. \qquad (7.12)$$

The output plan (7.8) satisfies these two conditions. The associated prices and maximum present value of profits follow easily. ■

Alternative Solution Procedure Of course, the optimal price discrimination follows already from the standard formula for optimal third–degree price discrimination. Note, if marginal cost is equal to zero, that formula requires user prices (the unit cost of consumption) to be set at the level where the corresponding price elasticities of demand are equal to -1 or equivalently (see (3.4))

$$\frac{1}{\varepsilon_i} := \frac{\partial P_i}{\partial X_i}\frac{X_i}{P_i} = -1, \quad i \in \{1, 2\}.$$

Since $P_i(X_i) = 1 - X_i$, it follows immediately that

$$\frac{X_i}{1 - X_i} = 1, \quad i \in \{1, 2\}.$$

Hence, $X_1 = X_2 = \frac{1}{2}$, and $x_1 = X_1 = \frac{1}{2}$, $x_2 = X_2 - x_1 = 0$, user prices $= 1/2$, prices for ownership $p_1 = (1 + \delta)1/2$, $p_2 = 1/2$, as asserted in Proposition 6.

Proposition 7 (Time Inconsistency) *The optimal third–degree price discrimination characterized in Proposition 6 is time inconsistent. In the absence of commitment power the monopolist deviates from that plan.*

Proof Suppose the monopolist starts out with $x_1 = \frac{1}{2}$ and customers believe that the monopolist will continue with the optimal output plan. Then, the market clearing price in period 1 is equal to $p_1 = \frac{1}{2}(1 + \delta)$. However, when period 2 has arrived, the monopolist faces the inverse residual demand function $P_2 = (1 - \frac{1}{2} - x_2)$. The maximizer of the associated profit function $\pi_2(x_2) = \frac{x_2}{2} - x_2^2$ is $x_2 = \frac{1}{4} > 0$. Since the monopolist is free to deviate from any plan, he will serve the market again at the rate $x_2 = \frac{1}{4}$ — violating the optimal plan. ∎

At first glance you may wonder: how can the monopolist raise his profits by deviating from the optimal plan? Of course, this works only if customers believe incorrectly that the monopolist sticks to the optimal plan and will not erode the price of the durable good in period 2. Of course, rational customers anticipate that the optimal plan is not time consistent, adjust their marginal willingness to pay for ownership accordingly, and shift purchases to later periods when prices are lower — to the monopolist's dismay.

7.3 Optimal Time Consistent Price Discrimination

If customers anticipate correctly at what rate the monopolist will serve the market in period 2, once they have observed the output rate in period 1, the monopolist is restricted to time consistent output plans. Therefore, the monopolist's decision problem is to find the optimal time consistent third–degree price discrimination.

Can one find a time consistent plan? A simple way to achieve time consistency is to serve the market only during the last relevant period: $(x_1 = 0, x_2 = \frac{1}{2})$. However, such "end loading" is surely not the optimal time consistent plan since it completely destroys the benefits of durability.

In order to find the optimal time consistent plan, one has to solve a dynamic programming problem. The solution procedure is that of "backward induction". In a first step, one has to find the reaction function $x_2^*(x_1)$ that maps the first period output into the optimal second period output. That reaction function is employed by customers in their prediction of x_2 and in computing the optimal first period output, given customers' predictions.

Suppose the monopolist has supplied x_1 in period 1 and customers have observed this. Then, in period 2 the monopolist will supply that quantity which solves

$$\max_{x_2} \pi_2(x_1, x_2). \tag{7.13}$$

The solution is described by the reaction function

$$x_2^*(x_1) := \max\{\frac{1-x_1}{2}, 0\}. \tag{7.14}$$

Customers anticipate the monopolist's reaction and, after observing x_1, adjust their marginal willingness to pay for ownership in period 1 to

$$P_1(x_1, x_2^*(x_1)). \tag{7.15}$$

In turn, the monopolist anticipates that customers make these predictions. Therefore, the monopolist sets that output rate x_1 that maximizes the reduced form profit function $\pi_1^*(x_1)$

$$\max_{x_1} \pi_1^*(x_1) := \pi_1(x_1, x_2^*(x_1)). \tag{7.16}$$

Proposition 8 (Optimal Time Consistent Price Discrimination) *The optimal time consistent output plan, associated third–degree price discrimination, and maximum present value of profits is*

$$x_1 = \frac{2}{4+\delta}, \quad p_1 = \frac{\delta}{2} + \frac{2}{4+\delta} \tag{7.17}$$

$$x_2 = \frac{2+\delta}{2(4+\delta)} \quad p_2 = \frac{2+\delta}{2(4+\delta)}, \tag{7.18}$$

$$\pi_1 = \frac{(2+\delta)^2}{4(4+\delta)}. \tag{7.19}$$

Altogether, both prices and the present value of profits are lower than in the optimal plan.

Proof Inserting (7.14) into (7.6) gives the reduced form profit function (for $x_1 \leq 1$):

$$\pi_1^*(x_1) = x_1 \left(1 - x_1 + \delta(1 - x_1 - \frac{1-x_1}{2})\right) + \delta\frac{(1-x_1)^2}{4}.$$

$\pi_1^*(x_1)$ is strictly concave and the first derivative of $\pi_1^*(x_1)$ vanishes at $x_1 = 2/(4+\delta)$. The asserted x_2 is the monopolist's best response to this x_1, by (7.14). And the associated prices and maximum present value of profits follow immediately. ■

Leasing as Commitment Mechanism? As Bulow[46] observed, *short term leasing* may serve as a commitment mechanism that supports optimal price discrimination. This follows immediately by the fact that the solution of (recall, the $X's$ denote the consumption flow and the $x's$ acquisition of ownership)

$$\max_{X_1, X_2} P(X_1)X_1 + \delta P(X_2)X_2$$

is $X_1 = X_2 = \frac{1}{2}$, and that this plan is also time consistent because

$$\arg\max_{X_2} P(X_2)X_2 = \frac{1}{2}.$$

However, notice that observed leasing arrangements are typically long term and thus cannot solve the durable goods monopoly problem. Leasing can only set the right incentives to maintain the optimal price of the durable good if leasing rates are flexible and are always adjusted to the current market rate. Observed leasing arrangements are apparantly geared to save taxes, not to solve a commitment problem.

Example 6 *As an exercise, suppose the durable good is subject to physical decay at the rate $\rho \in (0,1)$ (after one period, for each unit only ρ units are left), and set the discount factor equal to 1. Then, one has*

$$\pi_1 = (1 - x_1)x_1 + (1 - \rho x_1 - x_2)\rho x_1 + (1 - \rho x_1 - x_2)x_2. \tag{7.20}$$

Therefore, the following output plan is optimal

$$x_1 = \frac{1}{2}, \quad x_2 = \frac{1-\rho}{2}. \tag{7.21}$$

And the optimal time consistent output plan is

$$x_1 = \frac{2}{4+\rho^2}, \quad x_2 = \frac{1-\rho x_1}{2}. \tag{7.22}$$

[46]Bulow, J. [1982]. "Durable goods monopolists", *Journal of Political Economy*, 90: 314-332.

Comparative Statics

1. Switching to an inefficient technology with higher marginal costs and lower fixed costs can be profitable.
2. The monopolist has an incentive to reduce durability ("obsolescence planning").
3. If there is a constant inflow of a new generation of customers, the equilibrium price sequence can be cyclical.[47]
4. Incomplete information concerning the seller's marginal cost tends to benefit the monopolist.

7.4 Coase Conjecture

The durable goods monopoly problem was discovered by Coase[48] in one of his seminal contributions to economics. Coase did not only discover the inconsistency of the optimal plan. He also conjectured that the monopolist tends to lose all monopoly power if he can adjust prices faster and faster. The latter assertion has become known as the "Coase conjecture". It was later proved by Stokey for particular demand functions, and generalized by Gul, Sonnenschein and Wilson and by Kahn.[49]

The Coase conjecture is intuitively appealing. If the time span between trading periods within a given time period is reduced, the durable goods monopolist works himself down the demand function at a faster pace. Since consumers anticipate that prices will fall, fewer and fewer transactions take place at high prices, and more and more transactions are concentrated around the competitive price. In the limit, all transactions take place "at a twinkling of the eye", and the equilibrium present value of profits converges to zero.

[47] See Conlisk, J. and E. Gerstner and J. Sobel [1984]. "Cyclic pricing by a durable goods monopolist", *Quarterly Journal of Economics*, 99: 489-505.

[48] Coase, R. [1972]. "Durability and monopoly", *Journal of Law and Economics*, 15: 143-149.

[49] Stokey, N.[1979]. "Intertemporal price discrimination", *Quarterly Journal of Economics*, 93: 355-371; Stokey, N. [1981]. "Rational expectations and durable goods pricing", *Bell Journal of Economics*, 12: 112-128; Gul, F. and H. Sonnenschein and R. Wilson [1986]. "Foundations of dynamic monopoly and the Coase conjecture", *Journal of Economic Theory*, 39: 248-254; Kahn, C. M. [1986]. "The durable goods monopolist and consistency with increasing costs", *Econometrica*, 54: 275-294.

8 Bilateral Monopoly and Bargaining*

So far monopoly was equated with the power to unilaterally dictate either the unit price or some more sophisticated price discrimination scheme. In the language of game theory, monopoly pricing was modeled as an "ultimatum game" where the monopolist sets the price function and customers make their purchases accordingly.

An "ultimatum" is a final proposition or demand whose rejection will end negotiations. It requires the ultimatum player to have a reliable commitment to end negotiations if the ultimatum is rejected. Such commitment is, however, difficult to achieve. This suggests that the theory of monopoly pricing should be put into the framework of more general bargaining games between buyer and seller where both have some market power.

One way to generalize the simple ultimatum game is to permit several rounds of haggling, while maintaining that one party has ultimately the power to make an ultimatum, and assume that delay of reaching an agreement is costly. This is the perspective of "finite horizon" bargaining games.

In most bargaining settings one can always add another round of haggling. Therefore, the finite horizon framework is not entirely satisfactory. Moreover, the two parties may be in a "bilateral monopoly" position, where both sides have equal market power, and no one is in the position to make an ultimatum at any point in the relationship. In either case, "infinite horizon" bargaining games are appropriate where no one has ever the power to set an ultimatum.

In the following, we give a brief introduction to finite and infinite horizon bargaining games. The emphasis is on explaining basic concepts and to present simple proofs of existence and uniqueness of the bargaining solution. The main limitation is that we stick to the complete information framework. Incomplete information is, however, particularly important in order to understand why rational bargainers may end up with a breakdown of negotiations despite potential gains from trade.

8.1 A Finite Horizon Bargaining Game

Consider the following three stage bargaining game: Two players, 1 and 2, bargain over the division of a fixed sum of money, say 1$. Their utilities are linear in money, but there is discounting with discount factors $\delta_1, \delta_2 < 1$.

The bargaining has three rounds or stages; exactly one period passes between two consecutive rounds.

Stage 1 Player 1 asks for the share $x_1 \in [0, 1]$. Player 2 accepts or rejects;

if he accepts, the game is over, if he rejects, it continues.

Stage 2 Player 2 asks for a share $x_2 \in [0,1]$; if player 1 accepts, the game is over, if he rejects, it continues.

Stage 3 Player 1 obtains the default payment $d > 0$, and player 2 goes empty handed.

This game has many Nash equilibria (find at least two). But since it has a subgame structure, we invoke subgame perfection. This selection principle is effective. Indeed, the game has a unique subgame perfect Nash equilibrium, explained as follows.

1. If stage 3 is reached, player 1 gets his default payment d; player 2 gets nothing.

2. If stage 2 is reached, player 2 asks for the share $x_2^*(2) = 1 - \delta_1 d$, and player 1 accepts. Acceptance gives player 1 the payoff $1 - x_2^*(2) = \delta_1 d$; rejection gives d one stage later, which has the present value $\delta_1 d$. Therefore, $x_2^*(2)$ makes player 1 indifferent between acceptance and rejection.

3. In stage 1, player 1 asks for the share $x_1^*(1) = 1 - \delta_2(1 - \delta_1 d)$; player 2 accepts. The share $x_1^*(1)$ is chosen in such a way that player 2 is made indifferent between acceptance and rejection (the left–hand side is player 2's payoff from acceptance, the right–hand side that from rejection):

$$1 - x_1^* = \delta_2 x_2^* = \delta_2(1 - \delta_1 d). \tag{8.1}$$

   This gives $x_1^*(1) = 1 - \delta_2(1 - \delta_1 d)$, as asserted.

The associated equilibrium payoffs are

$$\pi_1^*(d, \delta_1, \delta_2) = 1 - \delta_2(1 - \delta_1 d) \tag{8.2}$$

$$\pi_2^*(d, \delta_1, \delta_2) = \delta_2(1 - \delta_1 d). \tag{8.3}$$

Evidently, each player would prefer to face an impatient rival with a high discount rate; impatience is costly. And if agent 2 is infinitely impatient, $\delta_2 \to 0$), the solution approaches that of the *ultimatum game.*

An important special case occurs when there are no payoffs in case of disagreement ($d = 0$). Then,

$$\pi_1^*(0, \delta_1, \delta_2) = 1 - \delta_2, \quad \pi_2^*(0, \delta_1, \delta_2) = \delta_2. \tag{8.4}$$

The general finite bargaining game, with possibly many rounds, was solved by Ståhl.[50] The extension to an infinite game — to which we turn next — is due to Rubinstein.[51]

8.2 Infinite Horizon Bargaining — the Rubinstein Solution

Finite horizon bargaining games are not convincing. No matter how long the parties have already bargained, they can always add yet another round of haggling. This suggests that, at the outset, one should not restrict the number of stages. This leads us to the bargaining game solved by Rubinstein, the solution of which is known as the "Rubinstein solution".

The analysis of this bargaining game utilizes an important property of subgame perfectness known as the "one–stage–deviation principle".

Proposition 9 (One–stage–deviation–principle) *Consider a multi–stage game with observed actions. A strategy profile is subgame perfect if and only if it satisfies the one–stage–deviation condition that no player can gain by deviating from it in a single stage while conforming to it thereafter.*

This principle applies to finite as well as to infinite horizon games, provided that events in the distant future are made sufficiently insignificant through discounting. Essentially, the one–stage–deviation–principle is an application of the fundamental dynamic programming principle of *point wise optimization* which says that a profile of actions is optimal if and only if it is optimal in each time period.[52]

Multiplicity of Nash Equilibria Like the finite horizon bargaining game, the infinite horizon game has many Nash equilibria; but only one is subgame perfect.

An example of a Nash equilibrium that is not also subgame perfect is the following profile of "tough" strategies:

"each player always asks for share $x = 1$ and never accepts less."

[50] See Ståhl, I. [1972]. *Bargaining Theory*, Stockholm.

[51] See Rubinstein, A. [1982]. "Perfect equilibrium in a bargaining game", *Econometrica*, 50: 97-109.

[52] For a proof of the one–time–deviation–principle see Fudenberg, D. and J. Tirole [1991]. *Game Theory*, MIT Press, pp. 109f.

Evidently, these strategies are a Nash equilibrium (given the rival's strategy, the own strategy is a best response). However, subgame perfection is violated because these strategies are improvable by one–stage deviations.

To see why this happens, one has to identify a particular history of the game at which it pays to engage in a one–stage deviation from the candidate strategy if the rival sticks to it. A case in point is the history described by the event that the rival has just asked for the share $0 < x < 1$.[53] Then, by sticking to the candidate solution strategy, one gets payoff equal to zero. Whereas, if one deviates just once and accepts, one earns $1 - x > 0$. Therefore, the candidate strategy is improvable by a one–stage deviation.

Proposer/Responder In the following, we call the player whose turn it is to propose the *proposer* and the player whose turn it is to either accept or reject the *responder*.

Subgame Perfect Equilibrium Rubinstein's noncooperative bargaining game has a unique subgame perfect Nash equilibrium; it exhibits stationary strategies, as follows. We first state and prove existence, and then turn to the proof of uniqueness. Since the proof of uniqueness includes the proof of existence, the following proposition is just another opportunity to exercise the one–stage deviation principle.

Proposition 10 (Existence) *The following strategy profile is a subgame perfect Nash equilibrium; it tells player i what to do, depending upon whether it is i's turn to propose or to accept/reject:*

1. *proposer's strategy: always ask for the share*

$$x_i^* = \frac{1 - \delta_j}{1 - \delta_i \delta_j} \tag{8.5}$$

2. *responder's strategy: always accept any share equal or greater than*

$$\frac{\delta_i (1 - \delta_j)}{1 - \delta_i \delta_j}. \tag{8.6}$$

[53]Note, this would not occur if both players played the candidate equilibrium strategy. However, a subgame perfect equilibrium strategy must not be improvable *for all* possible histories — not just the ones that occur in equilibrium. Therefore, the one–stage deviation principle must be applied on as well as off the equilibrium path.

Proof Consider the above strategy profile. We want to show that these strategies are indeed a subgame perfect Nash equilibrium. In view of the one–stage deviation principle, all one needs to show is that no one–stage deviation pays, for all possible histories of the game.

1) Suppose i is responder and i is offered at least $\frac{\delta_i(1-\delta_j)}{1-\delta_i\delta_j}$. The candidate strategy tells him to "accept". But what if he engages in a one–time deviation, "rejects" the offer, and then returns to the candidate strategy. Then the game enters into the next bargaining round where i is proposer, and since he returns to the candidate strategy, he asks for the share $\frac{1-\delta_j}{1-\delta_i\delta_j}$ which is accepted by the rival. Discounting this payoff shows that it would have been at least as good to accept the originally offered share in the first place.

By similar reasoning one can show that it does not pay to engage in a stage–deviation either if the rival has offered less than $\frac{\delta_i(1-\delta_j)}{1-\delta_i\delta_j}$.

2) Next, suppose i is proposer. Suppose he deviates once, asks for a higher share than $\frac{1-\delta_j}{1-\delta_i\delta_j}$ (lower shares would be accepted at a loss), and thereafter returns to the candidate equilibrium strategy. Then, j rejects and then offers i the share $1-\frac{1-\delta_i}{1-\delta_i\delta_j}$, which player i accepts. Properly discounting shows that this one–time deviation would lead to a lower payoff, since

$$\delta_i \frac{\delta_i(1-\delta_j)}{1-\delta_i\delta_j} < \frac{1-\delta_j}{1-\delta_i\delta_j}. \tag{8.7}$$

■

Proposition 11 (Uniqueness) *The stationary strategies described in Proposition 10 are the unique subgame perfect equilibrium.*[54]

Proof Suppose the game has not yet ended at time t. Denote the subgame perfect *continuation* payoff of the proposer at t by v and the corresponding continuation payoff of the responder by w. Define

$$\bar{v} := \sup(v), \quad \underline{v} := \inf(v) \tag{8.8}$$

$$\bar{w} := \sup(w), \quad \underline{w} := \inf(w). \tag{8.9}$$

We will show that the continuation payoffs are uniquely determined as follows

$$v_i = \bar{v}_i = \underline{v}_i = \frac{1-\delta_j}{1-\delta_i\delta_j} \tag{8.10}$$

$$w_i = \bar{w}_i = \underline{w}_i = \frac{\delta_i(1-\delta_j)}{1-\delta_i\delta_j}. \tag{8.11}$$

[54]The original proof by Rubinstein is rather involved. The following ingeniously simple proof was introduced by Sutton, J. [1986]. "Non–cooperative bargaining theory: an introduction", *Review of Economic Studies*, 53: 709-724.

Then it follows immediately that (8.5) and (8.6) are the unique subgame perfect equilibrium strategies. The associated equilibrium outcome is that player 1 opens the game asking for $x^* = \frac{1-\delta_2}{1-\delta_1\delta_2}$ which player 2 accepts.

1) *Assessing* $\underline{v}$*:* Suppose 1 is proposer. If he offers 2 at least as much as the present value of 2's "maximum" continuation payoff after rejection, $1 - x \geq \delta_2\bar{v}_2$, his proposal will definitely be accepted. Therefore, 1's continuation payoff cannot be below $1 - \delta_2\bar{v}_2$. The same reasoning applies if 2 is proposer. Hence,

$$\underline{v}_i \geq 1 - \delta_j\bar{v}_j. \tag{8.12}$$

2) *Assessing* $\bar{w}$*:* Suppose 1 is proposer. He will never offer the responder 2 more than $\delta_2\bar{v}_2$. Therefore, the responder's continuation payoff cannot exceed $\delta_2\bar{v}_2$. The same reasoning applies if 2 is responder. Hence,

$$\bar{w}_i \leq \delta_i\bar{v}_i. \tag{8.13}$$

3) *Assessing* $\bar{v}$*:* Suppose 1 is proposer. Since the responder will definitely reject any offer $1 - x \leq \delta_2\bar{v}_2$, player 1 cannot get more than $1 - \delta_2\bar{v}_2$ through acceptance of his current proposal. In turn, if 1's proposal is rejected, he cannot get more than $\delta_1\bar{w}_1$. The same reasoning applies if 2 is proposer. Hence,

$$\begin{aligned} \bar{v}_i &\leq \max\{1 - \delta_j\bar{v}_j, \delta_i\bar{w}_i\} && (8.14)\\ &\leq \max\{1 - \delta_j\bar{v}_j, \delta_i^2\bar{v}_i\} \quad \text{by (8.13)} && (8.15)\\ &= 1 - \delta_j\bar{v}_j. && (8.16) \end{aligned}$$

To prove the last step, suppose, *per absurdum,* that $1 - \delta_j\bar{v}_j < \delta_i^2\bar{v}_i$. Then, $\bar{v}_i \leq \delta_i^2\bar{v}_i$, and hence $\bar{v}_i \leq 0$. But then[55] $1 - \delta_j\bar{v}_j \geq 0 \geq \delta_i^2\bar{v}_i$, which is a contradiction. Again, similar reasoning applies if 2 is proposer. Hence,

$$\bar{v}_i \leq 1 - \delta_j\bar{v}_j. \tag{8.17}$$

4) Finally, combine all of the above, as follows:

$$\begin{aligned} \underline{v}_i &\geq 1 - \delta_j\bar{v}_j \quad \text{by (8.12)}\\ &\geq 1 - \delta_j + \delta_i\delta_j\underline{v}_i \quad \text{by (8.17)}\\ \Longrightarrow \underline{v}_i &\geq \frac{1-\delta_j}{1-\delta_i\delta_j}. && (8.18) \end{aligned}$$

[55]Note, $\delta_j, \bar{v}_j \leq 1$.

And

$$\begin{aligned} \bar{v}_i &\leq 1 - \delta_j \underline{v}_j \quad \text{by (8.17)} \\ &\leq 1 - \delta_j + \delta_i \delta_j \bar{v}_i \\ \Longrightarrow \bar{v}_i &\leq \frac{1-\delta_j}{1-\delta_i\delta_j}. \end{aligned} \tag{8.19}$$

Hence, $v_i = \bar{v}_i = \underline{v}_i = \frac{1-\delta_j}{1-\delta_i\delta_j}$, which confirms (8.10). Similar reasoning also confirms $w_i = \bar{w}_i = \underline{w}_i = \frac{\delta_i(1-\delta_j)}{1-\delta_i\delta_j}$(8.11). ■

Remark 2 (Nash Bargaining Solution) *Make the time lag between bargaining rounds arbitrarily small. Then the Rubinstein solution converges toward the well–known cooperative Nash bargaining solution.*[56] *If both players have the same discount factor, that solution gives rise to equal sharing,* $x* = \frac{1}{2}$.

Remark 3 (Three or More Bargainers) *Unfortunately, the uniqueness result holds only in bilateral bargaining. With three or more parties, one may end up with a multiplicity of subgame perfect equilibrium solutions.*

9 Concluding Remarks

The theory of monopoly has made important advances. However, we are still at a far distance from a unified theory that explains under which circumstances either price discrimination or Cournot monopoly tends to emerge. We close this introduction to monopoly with some tentative remarks on the intricate relationships between monopoly and bargaining, and the need for further research on this topic.

A peculiar feature of price discrimination is that the monopolist draws a positive profit not only from the collective body of customers but from each and every individual customer. If each customer is valuable, the monopolist is tempted to enter negotiations when a customer has refused his ultimatum. This, in turn, induces customers to question the credibility of the ultimatum and make counteroffers. Therefore, price discrimination can only work in the particular way in which it was analyzed if the monopolist has found a mechanism that reliably commits him to end negotiations when an ultimatum was rejected. And without such commitment, price discrimination is intricately linked with bargaining problems.

[56]See Nash, J. [1950]. "The bargaining problem", *Econometrica,* 18:155-162.

Not so under Cournot monopoly. Recall, at the Cournot point the marginal profit is equal to zero. If each customer contributes only a relatively small quantity to market demand, it follows that the marginal profit of serving individual customers is equal to zero. In that case, the monopolist need not worry about individual bargaining and problems of credibility. When an individual customer has refused the monopolist's ultimatum and starts haggling, the monopolist has no reason to enter negotiations because losing that customer is of no concern. Therefore, the Cournot point is "renegotiation proof", and hence can be maintained even if individual customers are in doubt about the monopolist's commitment power. This is one of the strong points of Cournot monopoly.

References

[1] M. Bagnoli and T. Bergstrom. Log–concave probability and its applications. Working paper, University of Michigan, 1989.

[2] R. G. Bartle. *Elements of Real Analysis.* Wiley & Sons, 1976.

[3] J. Bulow. Durable goods monopolists. *Journal of Political Economy*, 90:314–332, 1982.

[4] P. Cagan. The monetary dynamics of hyperinflation. In M. Friedman, editor, *Studies in the Quantity Theory of Money*, pages 25–117. Chicago University Press, 1956.

[5] R. Coase. Durability and monopoly. *Journal of Law and Economics*, 15:143–149, 1972.

[6] J. Conlisk, E. Gerstner, and J. Sobel. Cyclic pricing by a durable goods monopolist. *Quarterly Journal of Economics*, 99:489–505, 1984.

[7] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, 1991.

[8] F. Gul, H. Sonnenschein, and R. Wilson. Foundations of dynamic monopoly and the Coase conjecture. *Journal of Economic Theory*, 39:248–254, 1986.

[9] C. M. Kahn. The durable goods monopolist and consistency with increasing costs. *Econometrica*, 54:275–294, 1986.

[10] M. I. Kamien and N. L. Schwartz. *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management.* North–Holland, second edition, 1991.

[11] J. J. Laffont and J. Tirole. *A Theory of Incentives in Procurement and Regulation.* MIT Press, 1993.

[12] E. Maskin and J. G. Riley. Monopoly with incomplete information. *Rand Journal of Economics*, 15:171–196, 1984.

[13] J. Nash. The bargaining problem. *Econometrica*, 18:155–162, 1950.

[14] A. Plant. The economic theory concerning patents. *Economica*, 1:30–51, 1934.

[15] R. Posner. The social cost of monopoly and regulation. *Journal of Political Economy*, 83:807–827, 1975.

[16] P. Reagan. Inventory and price behavior. *Review of Economic Studies*, 49:137–142, 1982.

[17] J. Robinson. *The Economics of Imperfect Competition.* Macmillan, 1933.

[18] A. Rubinstein. Perfect equilibrium in a bargaining game. *Econometrica*, 50:97–109, 1982.

[19] F. M. Scherer. *Industrial Market Structure and Economic Performance.* Rand McNally, 1970.

[20] E. Sheshinski. Price, quality and quantity regulation in monopoly situations. *Economica*, 143:127–137, 1976.

[21] M. Spence. Monopoly, quality, and regulation. *Bell Journal of Economics*, 6:417–429, 1975.

[22] I. Ståhl. *Bargaining Theory.* Stockholm, 1972.

[23] G. J. Stigler. The theory of economic regulation. *Bell Journal of Economics*, 2:3–21, 1971.

[24] N. Stokey. Intertemporal price discrimination. *Quarterly Journal of Economics*, 93:355–371, 1979.

[25] N. Stokey. Rational expectations and durable goods pricing. *Bell Journal of Economics*, 12:112–128, 1981.

[26] G. Stolper. *Deutsche Wirtschaft seit 1870.* Mohr & Siebeck, 1964.

[27] J. Sutton. Non–cooperative bargaining theory: An introduction. *Review of Economic Studies*, 53:709–724, 1986.

[28] P. Temin. *Did Monetary Forces Cause the Great Depression?* Norton, 1976.

[29] F. Verboven. International price discrimination in the European car market. *Rand Journal of Economics*, 27:240–268, 1996.