1 Introduction

Within the pages of Herodotus is a passage that begins

> Beyond the Pillars of Hercules there is a race of men . . .

The Pillars are a pair of mountains that flank the entrance to the Straits of Gibraltar. According to Plato, the lost city of Atlantis was to be found beyond this point. The race of men were the Carthaginians, who were frequent visitors to this unnamed country lying beyond Gibraltar:

> . . . where they no sooner arrive but forthwith they unlade their wares, and, having disposed them after an orderly fashion along the beach, leave them, and, returning aboard their ships, raise a great smoke. The natives, when they see the smoke, come down to the shore, and, laying out to view so much gold as they think the worth of the wares, withdraw to a distance. The Carthaginians upon this come ashore and look. If they think the gold enough, they take it and go their way; but if it does not seem to them sufficient, they go aboard ship once more, and wait patiently. Then the others approach and add to their gold, till the Carthaginians are content. Neither party deals unfairly by the other: for they themselves never touch the gold till it comes up to the worth of their goods, nor do the natives ever carry off the goods till the gold is taken away.

Dumb barter; trade with neither sight nor sound of the other. Could it be true? Herodotus is known to have told some whoppers in his time. His account of gold-digging ants bigger than foxes but smaller than dogs beggars belief.

From the fourteenth century, the explorer Ibn Battuta sends us an account from the Volga River of a land of darkness, 40 days journey hence where those

> who go there do not know whom they are trading with or whether they be jinn or men, for they never see anyone.

A century later, the prelate Paulus Jovius reports that dumb barter was common among the Lapps, writing

> They bargain in simple faith with absent and unknown men.

These accounts, taken at face value, raise at least three questions:

1. If trade is anonymous, why doesn't one party steal the goods offered by the other?
2. If trade was mutually profitable and longstanding wouldn't this be an inducement to communication?

3. How do the parties decide the location of the "trading post" in the absence of communication?

If you find these questions of interest, then the study of economics is for you.

What is economics? The essayist Thomas Carlyle damned it as the *dismal science*. The art critic and social reformer John Ruskin called it the *bastard science*, which for him was a mild reproof.[1] The poet Matthew Arnold referred to economists as a *one-eyed race*.[2] In those days cyclopean economists were concerned with the causes of wealth and the exchange of material things. In 1932 Lionel Robbins[3] changed the conception of the subject. "Economics," he wrote, "is the science which studies human behavior as a relationship between ends and scarce means which have alternative uses."[4] And so, economics was elevated from mere catallactics to the study of all human behavior. It is not for nothing that economics is called the imperial science!

The goal of this book is more modest. It is to show how the economist's perspective is useful for thinking about the trade-offs associated with setting the terms of trade. Note the use of the word "perspective" in this paragraph's second sentence. This book is not a laundry list of facts or "take aways." It is about a particular way of approaching problems.

A remark about method is in order. Natural scientists can, and do, run randomized controlled trials to determine whether A causes B as well as the precise quantitative nature of the relationship. Economists rarely enjoy this luxury.[5] However, they can and do run controlled *thought* experiments called models. These are caricatures of reality in which the superfluous or complicated is stripped away. In this artificially constrained environment it is possible to deduce exactly what a rational agent will do and its consequences. Like the Chorus in *Henry V*, one asks that you "piece out our imperfections with your thoughts."

At the end of the exercise, one can conclude that there is a combination of circumstances in which A causes B. The actual quantitative nature of that relationship will be elusive, as it depends intimately on the specifications of the model. The usefulness of such *qualitative* conclusions depends on two things. First, would you ever have imagined that A could cause B? Second, are the circumstances under which A causes B reasonable?

[1] Of the painter James Whistler, Ruskin wrote "I have seen, and heard, much of Cockney impudence before now; but never expected to hear a coxcomb ask two hundred guineas for flinging a pot of paint in the public's face." Whistler sued Ruskin for libel, and won a farthing and no costs.

[2] Wandering between two worlds, one dead
The other powerless to be born, . . .

[3] Later Lord Robbins of Clare Market (1898–1984). Famous in the larger world as the author of the Robbins Report that instituted some of the most sweeping reforms of British higher education ever seen.

[4] Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science*, 2nd edn, London, Macmillan.

[5] It is now very much the rage among some economists to run controlled trials. Like all fashions, this too will pass.

Now, some advice. Reading economics is *not* like reading fiction. One is not a spectator to be entertained while waiting to die. Instead, one must engage with the text. Wrestle with its arguments. Find flaws in the logic. Discuss them with others.

1.1 Is Economics a Science?

It's a bad question. Comparisons made to arrive at a demarcation are problematic. If science were a country, physics might be its capital. If one asks whether history is a science, the customary thing is to measure the proximity of history to science's capital city. Why to the capital and not one of the outlying settlements like geology and archaeology?

Second, the question "is X a science?" is of interest only if we believe that scientific knowledge is privileged in some way. Perhaps it alone is valid and useful while non-scientific knowledge is not. If so, the correct question is not whether X is a science, but whether X produces knowledge that is valid and useful. Now we have something interesting to discuss: what constitutes useful or valid knowledge? Does economics provide this? Read on and judge for yourself.

1.2 Rationality

Economics begins with the assumption that the participants in a trade are rational; suggesting that mad dogs and Englishmen aren't the only ones who run out into the noonday sun. The assumption attracts criticism and deserves discussion. What precisely does it mean to be *rational*?[6]

The economist's conception of rationality has three parts. First, trading agents have *preferences*. Second, their preferences are *consistent* at all times. For example, if on a given day you like oranges, then you should, other things equal, prefer 5 oranges at least as much as 4 oranges. In other words, more of a good thing is never worse and possibly preferred to less of it. If you say you prefer apples to oranges and oranges to cherries then, other things equal, you should prefer apples to cherries. These restrictions do not rule out the possibility that on hot days you prefer ice-cold beer to hot chocolate and the reverse on cold days. Notice the qualifier "other things equal." Under identical conditions, separated in time, one's likes and dislikes are invariant. The trading agent is as

constant as the northern star,
Of whose true-fixed and resting quality
There is no fellow in the firmament.

[6] The reader who thinks that rationality does not require a definition should ponder the following: I'll give you a million dollars to do something irrational.

Third, when asked to select from a menu of things, outcomes, or possibilities, one will choose the most preferred item from the menu. This is the assumption of self-interest. Francis Edgeworth (1845–1926) described it thus:

> the first principle of Economics is that every agent is actuated only by self-interest.[7]

Given the first two conditions, the third is inevitable. Why wouldn't one choose one's most preferred outcome.

Is the rationality assumption plausible?[8] This book takes the view that the plausibility of the assumption depends on the context. Further, these contexts are neither rare nor exceptional. Less well appreciated is that the rationality assumption is more interesting than the alternatives. When perverse things happen because agents are irrational, this is dull. Why? One can eliminate the perverse outcomes by replacing the agents concerned. When perverse things happen *because* agents are rational, this is interesting because the cause of the perverse outcomes cannot be laid at the agent's feet. One must look, instead, to the environment they inhabit.

Not all economists are wedded to the assumption of rationality. A subset, called behavioral, examines, with gusto (but no brio), the consequences of relaxing the rationality assumption. Invariably charming, they are full of delightful tales about human foibles. Their writings are more properly housed in the self-help sections of book stores, but that is my own prejudice. Should you read what they write, do so with suspicion.[9] To illustrate, consider the following from Daniel Ariely, a celebrated member of the tribe:

> If you spend three years in a hospital with 70% of your body covered in burns, you are bound to notice several irrationalities. The one that bothered me in particular was the way my nurses would remove the bandage that wrapped my body. Now, there are two ways to remove a bandage. You can rip it off quickly, causing intense but short-term pain. Or you can remove it slowly, causing less intense pain but for a longer time. My nurses believed in the quick method. It was incredibly painful, and I dreaded the moment of ripping with remarkable intensity. I begged them to find a better way to do this, but they told me that this was the best approach and that they knew the best way for removing bandages. It was their intuition against mine, and they chose theirs. Moreover, they thought it unnecessary to test what appeared (to them) to be intuitively right.

Or, the nurses lied because time is short and they had many patients to care for. Enough of prologue. Down to brass tacks.

1.3 Rational Buyer Model

The simplest transaction to contemplate is that between buyer and seller. To determine a price, the seller must have in mind how the buyer will respond. The model of buyer

[7] Edgeworth has been described as "adept at avoiding conversational English." He once asked T. E. Lawrence (of Arabia): "Was it very caliginous in the Metropolis?" Back came the reply: "Somewhat caliginous but not altogether inspissated."

[8] If you disagree, you are welcome to send me twenty dollars and stop reading.

[9] As you should with this book.

behavior used in the first portion of this book will be described here.[10] Called the rational buyer model, it is characterized by three assumptions.

1. **Assumption 1**
   A buyer is able to assign an immutable monetary value to *every* transaction. This value is called their **reservation price** (RP for short). It is the maximum price a buyer is willing to pay for an additional unit of the product (or service).

   Assigning a monetary value to some transactions is difficult. How much for one's grandmother? Argosies of gold, silver, and peacock's eggs? Be careful not to confuse an unwillingness to assign a hard dollar figure to the life of another with the inability to do so. Our actions can betray us, like purchasing a morning coffee rather than donating to Oxfam. Even were it impossible to place a dollar value on everything, this would not invalidate a model based on this assumption. A model can be useful without being universal. What matters is that there be a sufficiently important class of transactions in which such an assumption is plausible.

   If one accepts the assumption, one might wonder about the possibility of determining a buyer's RP. This is challenging but not impossible. There are a host of statistical and econometric tools that have been developed to do just that.
2. **Assumption 2**
   A buyer evaluates a transaction in terms of its **consumer surplus**, which is the difference between her RP and the price she pays. For example, suppose a buyer's RP for a pound of pepper is \$5 and we sell it to her for \$3. If she buys the pepper from us, she will enjoy a consumer surplus of \$5 − \$3 = \$2. A buyer will never purchase a product that yields negative consumer surplus.
3. **Assumption 3**
   In choosing between transactions, the buyer will choose the transaction which maximizes her consumer surplus. This assumption captures the idea that more money is better than less. For example, suppose our buyer has a choice between a pound of pepper or a pound of salt. For simplicity, assume she will buy one or the other but not both. Let her RP for pepper be \$5 and for salt be \$4. Pepper is sold at \$3 a pound while salt is sold at \$3.50 a pound. Which will she acquire? In this case, the consumer surplus on pepper is \$2 while on salt it is \$0.50. So, she will buy pepper.

   Implicit in this assumption is that a buyer is not cash constrained.
   We incorporate cash constraints in Chapter 5. Until further notice we ignore them. Imaginative sellers find ways around cash constraints, something that even Lenin recognized. When asked how capitalists were to be hanged, there being insufficient rope, he is supposed to have responded "they will sell it to us on credit."

Return to the pepper vs. salt choice. As the seller of salt, how can you induce the buyer to purchase salt instead of pepper? Obviously, by dropping the price, in this case to below \$2. At that point the consumer surplus from salt is larger than that from pepper. Alternatively, one can induce the buyer to increase their RP for salt by at least

[10] A more general model is described in Chapter 5.

$1.51. Put differently, get them to value salt more highly. This might require making the buyer more aware of the product's usefulness (advertising), or making changes in the product (or service) (i.e. adding value).

If a buyer is interested in more than one unit of the product, we can model this using *incremental* RPs.

Example 1 *Table 1.1 displays the incremental RPs for various amounts of the product Soma.*[11]

Table 1.1 Incremental RPs

Quantity	First unit	Second unit	Third unit	Fourth unit
A's RP	7	5	3	1

This table means that she values the first unit of Soma at \$7. She values the second unit of Soma at \$5 and so on. Notice that her incremental RPs are declining; she exhibits diminishing marginal returns. You may assume that the RP for the fifth and higher units is zero.

If each unit of Soma was priced at \$4, how many units would she buy? She will buy as many units as maximize her consumer surplus. Thus, if she buys one unit, her surplus will be $7 - 4 = 3$. *If she buys two units, her surplus will be* $7 + 5 - 4 \times 2 = 4$. *If she buys three units, her surplus will be* $7 + 5 + 3 - 4 \times 3 = 3$. *If she buys four units, her surplus will be* $7 + 5 + 3 + 1 - 4 \times 4 = 0$. *Hence, her surplus is maximized when she buys two units.* □

Frequently a buyer will purchase through an agent. Large companies, for example, employ purchasing specialists. In these cases the incentives of the agent and their superiors need not coincide.[12] In what follows we will not consider this possibility.

1.4 Demand Curves

Demand curves summarize how the demand of a single individual, or a collection of them, changes with the price paid (holding the price of other goods and services fixed). To illustrate why they are convenient, imagine a population of a million buyers

[11] A possibly fictitious plant whose juice was used in India to produce an intoxicating drug. It appears in Aldous Huxley's *Brave New World* as a narcotic that is distributed by the state to produce social harmony.

[12] In some industrial settings an equipment purchase locks the buyer into the purchase of spare parts and services. If the buyer's agent is evaluated on initial expenditures, the shrewd seller will price the original equipment low and the spare parts high.

each interested in purchasing at most one unit of a particular good. Each buyer is endowed with an RP for the good. At a given price, p say, we would have to count up the number of buyers with an RP of at least p. Sorting the 1 million RPs on file would not be a burden. However, having to store the 1 million RPs would be inconvenient. Instead, we summarize the information needed, the number of buyers at a given price, using an algebraic function.[13] An example of a demand curve is $D(p) = 100 - 2p$. The left-hand side denotes the demand as a function of the unit price p faced by the consumer. The right-hand side is the precise functional form of that demand. In this example, if the unit price p is \$3, demand will be $100 - 2 \times 3 = 94$ units. We interpret a price p at which $D(p) < 0$ to mean demand is zero. The lowest price p at which $D(p) = 0$ is called the **choke price**. Strictly speaking we should express a demand curve as $D(p) = \max\{100 - 2p, 0\}$ to emphasize the fact that for a price above the choke price, demand is zero. We will usually not express demand curves this way with the understanding that one will remember that demand is zero for prices above the choke price.

Demand is a function of the price faced by the buyer *not* the price set by the seller. The two prices are not always identical. For example, suppose a sales tax of 5% is imposed on the price set by the seller. If the seller sets a unit price of p *exclusive* of tax, the price paid by the buyer is $1.05p$.

An important feature of a demand function is that as the price paid by the buyer of the product rises (holding other prices fixed), the demand for it declines. If you sold a good or service for which this relationship was reversed, you should be sitting on a beach somewhere earning 10% instead of reading this book.

1.4.1 Price Sensitivity

Fundamental to understanding how prices are set is the sensitivity of demand to a change in price. This is measured using **elasticity of demand**. It measures the sensitivity of demand to a change in price *holding other prices fixed*. Formally, it is the percentage change in the amount demanded for an infinitesimally small percentage change in the unit price *other things being equal*. Overlooking, for the moment, what an infinitesimally small percentage change in price is, we can express the elasticity of demand as

$$-\left(\frac{\%\text{ change in quantity}}{\%\text{ change in price}}\right).$$

While there is a negative sign in the expression for elasticity of demand, it is a non-negative number between 0 and ∞.[14] This is because a percentage increase in price will result in a percentage decrease in demand, making the numerator negative. The negative sign makes everything positive again. This definition of elasticity,

[13] In the case of a single individual, a demand curve summarizes the information contained in the table of incremental RPs (see Example 1).

[14] This is a departure from convention, where the negative sign is absent. This would make elasticity a negative number. However, no one ever refers to an elasticity of -3, for example.

while easy to digest, does not make explicit the fact that the elasticity of demand will vary with the current price. For example, if chocolate cost 1 cent a pound, we would not expect a 50% increase in price to have a significant effect on demand. On the contrary, if it cost 1 dollar a pound, a 50% increase in price might have more of an effect on demand. So, the sensitivity of demand to a change in price doesn't just depend on the magnitude of the change, but the base price from which the change is made.

Now let us turn to the issue of what an infinitesimally small percentage change in price is.

1. Suppose the current unit price is p. Therefore, the current demand is $D(p)$.
2. Increase the price by a small amount, h say, to $p + h$.
3. Demand at the new price is $D(p + h) < D(p)$.
4. % drop in demand is $100 \times \frac{D(p)-D(p+h)}{D(p)}$.
5. % increase in price is $100 \times \frac{(p+h)-p}{p}$.
6. Divide the percentage change in demand by the percentage change in price to get elasticity and let h go to zero. This step can only be executed if $D(p)$ is differentiable.

Step 6 leads to

$$\begin{aligned}\frac{\frac{D(p)-D(p+h)}{D(p)}}{\frac{(p+h)-p}{p}} &= \frac{D(p)-D(p+h)}{D(p)} \times \frac{p}{(p+h)-p} \\ &= \frac{D(p)-D(p+h)}{D(p)} \times \frac{p}{h} = \frac{p}{D(p)} \times \frac{D(p)-D(p+h)}{h} \\ &= -\frac{p}{D(p)} \times \frac{D(p+h)-D(p)}{h}. \end{aligned} \tag{1.1}$$

Now, let h go to zero. The term to the right of the product sign in (1.1) becomes

$$\lim_{h\to 0} \frac{D(p+h)-D(p)}{h} = \frac{dD(p)}{dp}.$$

Therefore, if $D(p)$ is differentiable, the elasticity of demand is $-[\frac{p}{D(p)}]\frac{dD(p)}{dp}$.

Example 2 *Suppose demand $D(p)$ as a function of unit price p is given by $3 - 0.5p^2$. Then, the elasticity of demand (as a function of p) is*

$$\frac{-p}{3-0.5p^2}(-p) = \frac{p^2}{3-0.5p^2}.$$ □

Demand as a function of unit price p, as Example 3 shows, need not be differentiable.

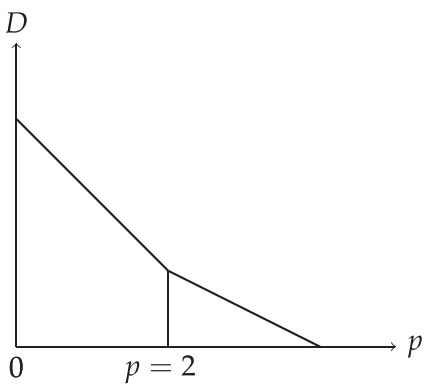

Figure 1.1 Non-differentiable demand

Example 3 *Suppose* $D(p) = 3 - p$ *for* $0 \leq p < 2$ *and* $D(p) = 2 - 0.5p$ *for* $2 \leq p \leq 4$. *A sketch of the demand curve appears in Figure 1.1.*

When $p < 2$, *the derivative of demand with respect to price is* -1, *however, for* $p \geq 2$ *the derivative is* -0.5. *Hence, the slope of the demand curve depends on whether one is to the left of* $p = 2$ *or to the right of it. For* $p < 2$, *we use the expression* $3 - p$ *to compute the elasticity of demand, which comes out to be* $\frac{p}{3-p}$. *For* $p > 2$ *we use the expression* $2 - 0.5p$ *to compute the elasticity of demand, which comes out to be* $\frac{0.5p}{2-0.5p}$. *Notice, at* $p = 2$, *the first expression for elasticity yields a value of* 2 *while the second yields a value of* 1. *Therefore, the percentage change in demand from a 1% change in the \$2 price will depend on whether one is contemplating a 1% increase or decrease in price.*

How might such a demand curve arise? Imagine two distinct markets for a hypothetical good. In market #1, demand as a function of price is given by $D_1(p) = 1 - 0.5p$ *and in market #2 it is* $D_2(p) = 2 - 0.5p$. *The choke price in market #1 is \$2 a unit. For a price p in each market that is at most \$2,* total *demand will be* $D_1(p) + D_2(p) = 1 - 0.5p + 2 - 0.5p = 3 - p$. *When the unit price exceeds \$2 a unit, the demand in market #1 falls to zero. Total demand will be* $2 - 0.5p$. □

1.5 Inverse Demand Curve

While natural to express demand as a function of price, it is frequently convenient to represent this relationship via an **inverse** demand curve. That is, write price as a function of quantity demanded, denoted $p(q)$. For example:

$$p(q) = \max\{7 - 2q, 0\}.$$

As in the case of demand curves, we will write the inverse demand curve as $p(q) = 7 - 2q$, keeping in mind that $p(q) = 0$ when $7 - 2q < 0$. The corresponding demand

curve will be $D(p) = \frac{7-p}{2}$. If we were to order buyers by decreasing RP, we can think of $p(q)$ as the RP of the qth buyer in the ordered list. Equivalently, if you sell q units, the RP of every buyer who purchases is at least $p(q)$.

1.6 Relevant Costs

In deciding the profitability of a particular choice, the only costs that matter are those that are relevant to the choice under consideration. Consider the following hypothetical situation. There is a machine capable of making two kinds of widget; red and blue.[15] Red widgets cost \$1 a unit to make and can be sold for \$2 a unit. At that price there are 100 buyers for red widgets. Blue widgets cost \$2 a unit and can be sold to a market of 300 buyers for \$2.50 a unit. The purchase price of the machine is \$X. There are two questions one can ask:

1. Given that you own the machine, which color widget should you make to maximize profit?
2. Should you buy the machine if you don't own it?

The answer to the first question is clearly blue. The blue widgets generate a profit of \$150. More importantly, the purchase price of the machine was irrelevant for deciding this issue. The purchase price of the machine was not a relevant cost in deciding on the color of widgets to make. The machine could have cost a billion dollars or nothing and the answer to the first question would be unchanged.

The answer to the second question is more involved. First, we must decide on how much profit we could make if we owned the machine (answer to the first question) and then we must verify whether that profit is sufficient to cover the purchase of the machine. If the purchase price \$X were smaller than \$150, we would buy the machine, otherwise not. So, for the second question the purchase price of the machine is a relevant cost.

A particular cost can be relevant for one set of choices and irrelevant for another set. We ignore throughout irrelevant costs when deciding on a profit-maximizing option.

1.7 Cost Function/Curve

The technological characteristics of a firm can be summarized by its total **cost function** or curve $C(q)$. The function $C(q)$ denotes the *minimum* total cost needed to

[15] The *Oxford English Dictionary* defines a widget to be any gadget or mechanical contrivance. The plastic container at the bottom of some beer cans used to produce a head of beer is called a widget as well.