The Foundations of Economic Theory

Fred Foldvary

BEP BUSINESS EXPERT PRESS
cognella academic publishing
www.cognella.com 800-200-3908

The Foundations of Economic Theory

Fred Foldvary
Santa Clara University

cognella academic publishing
www.cognella.com 800-200-3908

BEP BUSINESS EXPERT PRESS

The Foundations of Economic Theory

www.cognella.com

ISBN-13: 978-1-63157-204-3 (e-book)
Collection ISSN: 2163-7628 (electronic)

www.businessexpertpress.com

A publication in the Business Expert Press Economics collection

Cover and interior design by S4Carlisle Publishing Services Private Ltd., Chennai, India

Contents

CHAPTER 1

Origins of Economic Theory

1. The Quest for Economic Prosperity with Justice

The most important reason for learning economics is to understand social reality. Things are not what they appear to be. There is often an implicit reality that is not visible, obvious, or apparent. We need economic understanding to be able to know the reality. Economics enables us to understand the implicit reality beneath the superficial appearances of economic activity.

Many people also study economics because they wish to understand why we have social problems such as unemployment and poverty, and how these problems can be remedied. Most people have similar visions of prosperity and justice. We want to live in social harmony, where everyone who wants to can work and make a good living. We would like to eliminate poverty and live in a healthy environment.

We can envision ideal worlds, but they must be founded on sound principles if our vision is to succeed. Too often, as was the case with totalitarian central planning, utopian dreams turn into horrific nightmares because they have been based on unsound premises. This book presents a "paradigm" or basic analysis of economics called "foundational economics," which will enable you to understand the major social issues of our time and, as importantly, to be able to analyze assertions made by people, including politicians and economic authorities, to determine whether they are sound or contain incomplete theories or outright fallacies. The primary aim of this book is to enable you to think for yourself and analyze economic issues in a fundamental, logical, scientific way. The only authorities in economics, as in any science, are logic and evidence. The only prerequisite or prior knowledge needed is an open but critical mind.

The basic principle of foundational economics, the paradigm analyzed in this book, is that the market economy works, providing efficiency, equity, and sustainability. The legal structure and the government fiscal or tax policy must be in harmony with economic and ethical principles for it to work properly.

2. The Concept of An Economic Model

A model boat is a small-scale replica of the larger real-world item. A fashion model demonstrates clothes which real-world people could wear. A map is a model of some physical territory. Science also has models. Like the boat, it is similar to the object being analyzed, but smaller in the sense of being simpler and more abstract. Real-world people are usually not as beautiful or slender as models, but the model still shows how the clothes would be worn. A scientific model is a set of concepts and propositions which, like maps, demonstrate the main features of the phenomenon being analyzed. Often a model will have some very restrictive premises that simplify the phenomenon in order to emphasize a few of its aspects (for example, ignoring friction in a physics model to focus on gravity), and these premises are later relaxed (friction is then added) to bring in some realistic complications after some conclusions have been made about the main features.

For example, the story of Robinson Crusoe alone in an island is a favorite model that illustrates some economic concepts concerning only one individual; then, the second person, Friday, can be brought in to complicate things and make it closer to real-world experience. A large amount of theory is explained with such models.

Many economic models consist of premises from which conclusions are logically deduced. If the premises are general, they constitute pure theory, often conditional on certain non-universal secondary premises. If the premises include data from history or current economic figures, they become specific theories, whether descriptive or predictive. Many economic models are mathematical, with premises consisting of algebraic variables, functions and equations, the deduction being mainly mathematical manipulation. Many economic models can be "quantified" and tested by statistical or "econometric" calculations using computers. Computer programs have also been used as models that simulate economies. However, such programming, statistical analysis and higher math

(which is avoided in this book) is not really needed to understand the basic concepts and theories of economics.

3. Positive and Normative Economics

"Positive" economics is the study of the actual phenomena of the world, including predictions about the future. "Normative" economics concerns what one thinks economic policy should be, or how an economy is best established. It judges economies or some economic process by some norm or moral standard, or by some standard of efficiency.

People, including economists and politicians, often make normative statements based on their personal values or the values of some group. But if values are all personal and arbitrary, then is it possible to have a scientific theory of normative economics? This would require a scientific, or non-arbitrary, moral standard. Such a universal ethic has been derived in my book *The Soul of Liberty* (1980). What is not so evident is that a rational ethic is important not only for policy, but for an understanding of a market economy as well. This concept is explained in the following section.

4. The Ethical Foundation of A Market Economy

It is not enough to say that a market consists of voluntary acts, since we can then ask, what exactly are such acts? If your better product leaves me with fewer profits, that competition would not likely have my consent, yet it is not considered to be involuntary. So we need some ethical rules to tell which acts are voluntary; indeed, to define the concept of "voluntary."

In the discussion of normative economics above, the possibility of a universal ethic, a moral standard that does not depend on any particular culture or personal view, was raised. The full treatment of this ethic is beyond the scope of this book; interested readers are referred to *The Soul of Liberty,* as mentioned above. A brief outline of the derivation is useful, however, so that the ethic is not simply presented as fiat.

The two premises of the universal ethic were recognized by the philosopher John Locke (1690): independence and equality. Each human being thinks and acts as an internally independent living being, although he is socially dependent on others. As a common species, the qualities that make us reasoning beings with the capacity of deliberate choice is

equal to all. These premises are based on human nature, hence the ethic derived from these premises is referred to as "natural moral law" or, if the context is clear, just "natural law."

As an independent mind, each person perceives that the acts of others are either pleasing or displeasing to oneself, hence they are personally and subjectively good or bad. All values originate in these individual valuations. Therefore, no act can be universally deemed as good unless the recipient of the act deems it as good, and if he considers it good, one may not declare that it is universally not good, since the recipient is a moral equal. Hence, acts that benefit others must be considered good by the universal ethic.

Similarly, an act which affects no other person cannot be designated as evil if the person doing it does not consider it evil. Such acts can be generally designated as neutral, neither good nor evil.

This leaves us with acts which affect others negatively. Let's divide them into two categories, those which depend only on the state of the recipient's mind, which will be called "offenses," and those, called "harms," which do not. For example, if X objects to Y's religious views, this depends only on the personal views of X. On the other hand, if Y stabs X with a knife, this injury does not depend only on X's beliefs; it is an invasion, an unwelcome entering from outside of X into the domain of X.

Harms are considered bad or evil by the recipient, and thus also designated as evil by the universal ethic, since the only source of values is the individual. But offenses cannot be designated as evil, since the universal ethic, as noted above, is by definition independent of personal views; if the injury is caused only by such a subjective view, then the universal ethic designates it as neutral. We then have two definitions and three rules that make up the universal ethic:

A benefit is an act the recipient deems to be pleasing.
A harm is a direct, actual, invasive injury.

1. Acts that benefit others are good.
2. Acts that coercively harm others are evil.
3. All other acts are neutral, neither good nor evil.

A "voluntary" act can now be defined as one which is not evil according to the universal ethic. The totality of such acts in a certain context constitutes a market in the broadest sense.

With this universal ethic we can now derive natural or human rights, and the concept of "liberty." Liberty is defined as the absence of any legal restrictions on human action, other than those prohibiting and penalizing coercive harm. Moral rights are just another way of stating what is morally wrong. Natural rights are moral rights based on the moral law expressed by the universal ethic. If you have a moral right to do or have something, it means that it is wrong for others to prevent you from doing or having it.

Since economic resources boil down to land and labor, rights to either are fundamental. Since any arbitrary restriction on human action is invasive, a person has self-ownership, including the right to labor without restrictions other than against force and fraud. Since one has a right to one's time and energy, hence labor, one has a right to the reward of the labor, or wages. One also then has a right to the product of labor, produced wealth or capital goods. The taxation of labor or capital is therefore a violation of natural rights.

Land is a product of nature, not human exertion, so self-ownership does not apply to natural resources. Here, the equality premise applies: the right to the benefits of natural resources prior to their alteration by human effort belongs equally to all in the relevant community. As will be analyzed below, the efficient economic implementation of equal benefits is not a physical division of land, nor necessarily a community control over it, but by the collection of its rent to fund community services, leaving possession in individual hands.

This ethic is implemented by a body of law. The most fundamental and supreme set of laws, from which all other laws derive their authority, is a "constitution." The constitution of a community having a free society and free markets needs to determine the rules for property rights, the resolution of disputes, and the laws concerning contracts. A constitution is what economist James Buchanan (1987, p. 58) called a "choice of constraints" in contrast to a choice within constraints.

5. The Meaning and Methodology of Economics

The first aspect of methodology is careful definition. The key words of the field need to be concisely and precisely defined, so that the meaning in that context is clear.

"Economics" has been defined in several ways. The classical definition is that economics is the social science dealing with real wealth. Wealth, in the economic sense, is not money (which has value as generally exchangeable for real wealth), but thegoods which people produce. Thus, economics is a social science, dealing with the activities of people, and it specializes in the production, distribution, exchange, and consumption of wealth, or the material needs and desires of human beings.

But wealth is not an end in itself. People desire goods for the use, enjoyment, and value that they generate. These qualities as a whole are labeled "utility" in economics. Leisure time also has utility, and economics also takes into account the choice of not producing something but rather enjoying free time instead. Therefore, a broader and deeper definition of the subject is: economics is the science of utility. In contrast, related social sciences can be characterized as follows. Political science is the study of governance; sociology is the study of relationships; psychology is the study of genetically caused behavior; and anthropology is the study of human evolution and culture. These related fields impact on utility, and so are interdependent with economics.

In producing and consuming wealth, or pursuing utility, people do not behave randomly. As the Austrian economist Ludwig von Mises (1949, p. 4) stated, "Human action is purposeful behavior." People tend to follow particular laws or regularities due to their human nature. "The ultimate goal of human action is always the satisfaction of the acting man's desire" (p. 14). These desires, and the values people place on things, are subjective. People choose among alternatives, which implies they are able to rank their desires into those of greater and lessor importance.

It is also a fundamental fact that the resources from which wealth is produced are scarce. Some economists, following Lionel Robbins (1932) have therefore defined economics as "the allocation of scarce resources among alternative uses in order to satisfy human desires." This is descriptive of what goes on, but it seems less than comprehensive and yet overly detailed for a foundational definition of the scientific field, like saying that physics is "the examination of space, particles, energy, and time in the determination of laws and measurements of the fundamental

phenomena of the universe" instead of the simpler but comprehensive meaning, "the science of fundamental universal phenomena." Hence, in my judgment, a more parsimonious yet comprehensive definition of economics is "the science of utility."

The "methodology" of a science is the methods, techniques, and scientific philosophy used in obtaining knowledge. The methodology presented in this book is "foundational." The basic principles of foundational economics include definitions of key terms, the "taxonomy" or division of a field into meaningful categories, the formulation of premises universal to the field, the deduction of pure theory by the use of logic, the discovery of empirical knowledge by observation, the realization that our perceptions are tinged by our interpretations, and the discovery of specific theories regarding events and particular people, areas, and histories with conjectures and hypotheses tested by data.

In economics as in any science, the only authorities are logic and evidence, not the views and sayings of any person as such. As Henry George (1883, p. 242) stated it, "I ask no one who may read this book to accept my views. I ask him to think for himself."

6. The Foundational Premises of Economics

After defining key terms, a field of science requires fundamental premises or first principles. These are propositions or statements that apply to the whole universe of the field; in social science, this means they are valid for all people, in all times, places, and cultures. The following universal propositions are foundational for economics:

Propositions about Physical Resources and Technology

1. Some natural resources are scarce relative to human desires.
2. Resources vary in quality.
3. After some level of use, given the existence of one of more fixed resources, the use of extra amounts of a variable resource will produce ever smaller extra amounts of output (the law of diminishing returns).
4. Different amounts and methods of production may produce different amounts of output for the same proportion of inputs, i.e. returns to scale and to techniques can vary.

Propositions about Human Biology

1. All human beings belong to the same species and have the same human nature.
2. The human lifetime is finite.
3. Human beings have children who need care, and in old age, may no longer be able to work.

Propositions about Human Behavior and Thought

1. Human beings have ends, i.e. goals, desires, and needs.
2. Human beings are able to rank their ends, i.e. into those of greater and lesser importance.
3. Human desires tend to be unlimited.
4. Human values, both moral and material, are subjective.
5. Human beings economize: they desire to obtain things with the least possible unpleasant exertion, or equivalently, with some level of exertion, to obtain as much as possible.
6. The desires and motivations of human beings include 1) self-interest, ends connected with their own survival, happiness, and power, and 2) the well being of those for whom they have sympathy.
7. People tend to have a "time preference," preferring goods at present to those in the future.

Proposition about The Future

1. The future is uncertain.

The tenth proposition was noted by Henry George, who stated (1879, p. 507) that desires "short of infinity can never be satisfied," and (1883, p. 33) that "Man is not like the ox. He has no fixed standard of satisfaction." George (1879, p. 204) also stated the twelfth premise as a "fundamental principle of human action" that "men seek to gratify their desires with the least exertion." Carl Menger (1871, pp. 95–6) defined the principle of "economizing" thus: "men endeavor . . . to obtain the greatest possible result with a given quantity of a good or a given result with the smallest possible quantity."

These universal propositions are warranted by empirical observation and logic; most of them are obvious. We can observe that these

phenomenon occur and that they are not limited to any particular time, place, or culture. This inductive analysis involves a subjective investigation of one's own behavior and sentiments as well as the observation of the acts of others (which we interpret also from a knowledge of our own interior motives).

From these universal propositions or premises, the next element of economic methodology is to logically deduce propositions which make up "pure theory," pure because this theory applies universally. "Theory" is a systematically organized collection or set of propositions, a "theorem" being a proposition or statement that is warranted by logic and evidence. Some of these conclusions or statements may be conditional, depending on secondary premises which may or may not be applicable to a particular time or place.

7. Deductive and Inductive Theory

As stated above, the 15 foundational propositions are based on empirical observations which any person may verify for oneself from experience. Empirical or factual observations are then also used to verify deduced theories, to test whether the logic is sound.(It should be noted, however, that induction itself is also a type of deduction, since it involves two sets of premises: 1) a collection of observed facts; 2) rules by which one generalizes from facts. Deductions are then made using the data and rules.)

Using pure theory, economists then obtain knowledge about specific phenomena, events, and regularities that do pertain to some culture, time, or place. In physical and life science, scientists can perform experiments, while in social science, we rely mainly on the evidence from history, although economists also conduct experiments (often with students as subjects).

The method of finding specific theory is called "hypothetical deductive." An economist uses principles of pure theory to make an educated guess about some specific phenomenon; this is called a "hypothesis." The hypothesis is then tested with evidence, usually from current practices, from present and past data, or historical documents. If the evidence contradicts the hypothesis, the hypothesis is rejected. If the evidence is consistent with the hypothesis, then perhaps other tests will be conducted,

and other economists will then study the results. If the tests seems sound, then the hypothesis becomes a "specific theory" about a particular phenomenon rather than a statement that is universally true. Such theory is subject to change and challenge. Pure theory is also subject to challenge, but this would be a much more fundamental and less common challenge.

An example of a specific theory is the explanation of the Great Depression, and an example of conditional specific theory is the effect on future inflation when the Federal Reserve increases the money supply, if there is no change in the velocity or turnover of money.

An important principle of all theory is that of "interpretive understanding." When we observe a phenomenon, like a wild duck or a person sawing a board, the facts do not simply speak for themselves. We do not see just the facts, but interpret them according to our previous ideas, beliefs, and values. Some people will see a wild duck and think of how good it would taste or what fun it would be to hunt; others might see a beautiful living being, part of our natural heritage that should be preserved.

When it comes to observing human action, we interpret the internal as well as external acts. The internal state of the actor includes how it thinks and feels; since we are human, we presume that other human beings act similarly to how we act. So we can imagine ourselves sawing the board, and interpret why the person would do it; perhaps he is building something. We can imagine people acting out of hunger, the desire for wealth or power, erotic stimulation, or from love and devotion. We understand these things, or think we do, because we interpret the acts from our own experience, emotions, and reasoning. But interpretations, whether of acts, art, or some text, also requires logic to make sure we on the right track. It helps to have dialogue with others about the phenomenon, to make sure we are not overly reflecting our own personal biases. That is why an important part of any science is letting others see and criticize our work.

8. The Factors of Production

Like in the Bible, let us start our economics with the creation of the universe. In the beginning, there was the universe, and the earth. Then came human beings. As indicated above, economics is a social science, which

means a science about persons. We separate the universe into human beings and "nature," defined here as everything prior to being affected by human beings. Since humans use nature to satisfy their needs and desires, nature becomes natural resources.

Things of general value that people create from nature are termed "economic wealth." This consists of things that others would value enough to be willing to offer something else in exchange. In other words, economic wealth has a market value, or value to others. Money is not economic wealth, because people do not want it for itself alone, but for the things it will buy. The term "financial wealth" is used for things with market value in general, including money, bonds, and other items that do not have intrinsic value (value due to their being wanted for their actual use). The term "wealth" can refer to either financial or economic wealth. Here, the term "wealth" will be used as a shorter synonym for economic wealth.

Since economic wealth is neither natural resources nor people, it forms a third category of economic elements, capital goods. This leaves things that people have produced, but which do not have social or market value, such as garbage. This we will call "valueless products."

"Land" is usually thought of as the solid surface of the earth. But economic land has a wider meaning, namely, all natural resources, including the waters, air, and underground resources. Here, "land" will be used as a short synonym for economic land. So, the universe is divided into land, human beings, wealth, and valueless products.

Human beings exert effort in order to produce wealth, and this time and energy is termed "labor." Labor includes all human exertion in the production of wealth, including that done by managers and entrepreneurs. Labor includes the education, skills, and training that makes workers more productive; these enhancements are called "human capital."

People also use produced goods, or tools, to produce wealth. Capital goods are wealth items which have been produced, but not yet consumed. Money or other financial instruments that are owned by a business and eventually get exchanged for wealth are called "financial capital." So the term "capital" can mean financial capital or capital goods.

These are the three "factors" or resource inputs of production. Wealth is produced using land, labor, and capital goods. This wealth is distributed or assigned to the owners of these factors, or to those who have property

rights to the income of yields of the factors. Owners of land receive land rent, owners of labor (as self-owners) get wages, and the owners of capital goods get a capital yield.

Note that "wages" include any earning from wages, whether it be in the form of a salary, commission, fee, or in-kind yield, such as the fish people catch or fruit they pick. Also, much income is due to two or three factors. A farmer who owns his land earns wages from his own labor, rent from the portion of produce due to the benefit of land, and a yield on his capital goods investments, such as machines, buildings, and irrigation canals. These are income even if you pay them to yourself.

Having examined the foundations of economic theory, we will in the following chapters analyze these factors in more detail and then see how they and their products and the organization of production operate in an economy.

CHAPTER 2

Labor

1. Labor as a Factor of Production

This and the following two chapters will analyze the three factors of production: labor, land, and capital goods. Some people prefer the listing and analysis to be in the order of land, labor, and capital goods, for the reason that land appeared prior to human beings or is logically prior. But to understand the economics of land, we must first delve into that of labor, since the rent of land depends on the productivity of labor, hence labor is treated here first.

(This book was originally written for use by followers of Henry George, and thus much if is based on and cites the thought of this American economist,)

Henry George (1879, p. 32) provided a concise yet comprehensive definition of labor and wages: "the term labor includes all human exertion in the production of wealth, and wages, being that part of the produce which goes to labor, includes all reward for such exertion." This exertion includes both mental and physical effort, and it encompasses the efforts of entrepreneurs, managers, and the self-employed. (The Austrian economist Carl Menger (1871, p. 172) noted that "Entrepreneurial activity must definitely be counted as a category of labor services.") "Wages" includes any return or yield to labor, whether it be a salary, commission, or the profit of the self-employed. Part of the value of crops grown by a farmer on his or her own property or the gold panned by a prospector are wages.

To be meaningful, the concept of labor must be distinct from capital goods and land. George (p. 39) noted that people often speak of a worker's skill and knowledge as being "capital"; economists call these "human capital." But this meaning of "capital," he noted, is "a metaphorical use of language," and not to be confused with the use of "capital" as a resource or factor of production. A human being is different from a machine or a

horse; persons are the subjects, not objects, of social science. One may consider all inputs into a productive process as "capital," but then one would still distinguish among the meaningful categories of such capital, one of which would be labor, distinct from natural resources and capital goods.

2. The Determination of The Wage Level

But how does the wage level get established in the first place. And how does one account for poverty? As Henry George (p. 205) pointed out, a free man would not agree to work for another for less than he could secure by working for himself. So wages are determined by what one could earn from self-employment elsewhere. But where?

Wages From The Product of Labor

The determination of the wages of different types of labor is the result of the supply and demand for that type and quality of labor. In a pure market economy, wages ultimately reflect the value put on a worker's product by consumers: if consumers put a low value on his product then his wages will be low; if it puts no value on them, then his wages will be nil, and he will have to switch to producing a product other people want. As George (p. 77) put it, "The demand for consumption determines the direction in which labor will be expended in production."

Wages vary among different occupations because what some workers have to offer is valued more highly than what others have to offer. In consideration of one particular kind of labor, if there exists a shortage of labor then through the action of supply and demand wages for that labor will tend to rise. Conversely, where a particular labor is in surplus, wages for that labor will be reduced. The laws of supply and demand determine the relative wages among different types of labor.

But this still does not tell us what determines the wage level in an economy. There is a tendency for wages in an area to be linked together, so that one can speak of wages being generally high in Japan and low in India. Barbers in the United States have had a higher wage, even relative to local prices, than those in Mexico or Latvia, although their quality of work is similar.

George (pp. 26–7) recognized the "fundamental truth" that the basic principles of economics evident in a simpler, primitive, society are still in effect in a more complex, more developed world. We can discover the principles of wages by first analyzing a primal economy. Suppose there is a village that gets its food from hunting and gathering, and that there is more than enough land and game to support the village. The land is owned by the village in common, and since land is abundant, there it no rental value. Suppose also that the village is really primal, so that they don't have significant capital goods—no tools, like baskets or spears.

If the villagers go naked into the bush and gather nuts and berries with their bare hands, the only resources are land and labor. Here we see an economy in its most fundamental form: a person "endeavoring to obtain from nature by the exertion of his powers the satisfaction of his desires" (George, p. 27). Since there is no rent, the fruit they gather is all wages. Leaving out capital goods for now, what the naked hunters get from the forest, catching animals with their hands, is also wages. It is clear here that the hourly wage of the hunter/gatherer is equal to the produce that one can obtain from an hour's labor. Also clear is the principle that production precedes consumption.

When a laborer receives his wages in money instead of goods, the principle is the same. A worker "really receives in return for the addition his labor has made to the general stock of wealth, a draft upon that general stock . . . " (p. 29). Thus, money wages too are not economically an advance but only a claim on the amount of value one's labor has added to.

George notes, further, that all workers contribute to the production of all wealth. For example, the person who repairs fish nets helps catch the fish as much as the ones who go out into the sea. But the one who made the boat also contributed to the catch. And so did the one who made the wood and steel for the boat. Extending this to its logical conclusion, everyone who labors helped catch those fish. You as a worker help produce bread and steel by demanding these products in exchange for the goods you helped to make as well as by contributing ultimately to the goods that the bread maker and steel maker need. As George (p. 77) put it, "in aiding in the production of what other producers want, he is directing other labor to the production of the things he wants—in effect, producing them himself."

If we then add tools and buildings and other capital goods, labor is able to obtain more product, but again, the portion earned by labor in general will be its marginal (extra) contribution to the product. If land rent is zero, then since capital goods are produced by labor, both the tool maker and consumer-goods maker obtain their wages from their product. In practice, individual workers might get paid more or less than the economic value of their product due to personal biases of bosses, imperfect knowledge, inadequate or superior negotiating skills, or luck, but the general tendency in a pure market economy is for wages to equal the marginal product that labor provides.

Now, perhaps because the forest is destroyed, the society turns to farming. Each family gets a plot that it farms. Suppose it can grow 10 bushels of corn per unit of land per some period of time. They have as much land as they want of that quality, so rent is still zero. Again leaving capital goods aside, the 10 bushels grown by a farmer is all wages. Clearly, wages are drawn from the goods it produces, and in any particular area, the amount and value of the goods that labor can produce determines the wage.

The fact that some goods are produced over a long period of time does not change the principle. As George (pp. 50–51) notes, if a shoemaker starts with leather and works it up into a pair of shoes, the labor has gradually added more and more value to the original capital good, the leather. Hence the wage comes from the value added rather than from the original capital. As George (p. 56) put it, "Production is always the mother of wages."

The Extensive Margin

Suppose now that the most productive land, where farmers can grow 10 bushels, is all taken up. Farmers will now cultivate the next best area, which we can consider to be 9-bushel land. Wages at the 9-bushel land are 9. What, then are wages now in the 10-bushel land? If someone offers a wage of 9.5, all the farmers in the 9-bushel land will come running to apply. Someone who wants to hire labor only needs to pay 9 bushels. If he offers any less, no one applies, since they can get 9 by working for themselves.

If someone owns a farm on the 10-bushel land and hires a worker instead of working on it himself, that extra bushel produced after paying 9 to the employee is therefore not wages, but goes to the owner as rent. So wages in all land are equalized, due to competition among the workers, and any extra product goes to the owners of the lands as rent.

The best available land that can be had for free is called the "extensive" margin of production. It is called "extensive" because people keep extending or moving it out to lands of ever lesser quality as the better land gets taken up. The wage level is determined at the extensive margin, where the best free land is available. This boundary is also called the margin of cultivation, or more generally, for lands of all uses, the margin of production.

It is only when the margin is pushed further and further away and people are located on worse and worse land that the base rate of wages will fall. As George (p. 206) stated, "the wages which an employer must pay will be measured by the lowest point of natural productiveness to which production extends, and wages will rise or fall as this point rises or falls." If people are pushed to production on the land on which one can barely survive, then wages will be at a subsistence level.

In Great Britain, at the time when people set off to colonize Australia, wages were low. With labor competing for limited opportunities and no free good land available, employers could afford to offer low wages. However, in Australia, New Zealand, and America, the situation was reversed. Land taken from the aboriginal inhabitants was available to European immigrants, and it had a much higher yield than the margin in Europe. Therefore, wages for other occupations had to be high to keep employees.

Suppose in some island all the land gets taken up. There is no more free land. The margin, however, is still there, if not for agriculture, then for something else. One can go to the sea and catch fish in waters where one does not have to pay rent. In towns, one can add an extra story or build a two-story instead of one-story building; the margin would then be vertical space, where the top stories of buildings are located, since another story can be built without paying any more for land. There is almost always some type of land, whether air, water, or surface soil, that is available.

If we now switch from a one-crop economy to many crops, different products, we see that labor can be used to produce one or the other. The value of the labor, its wages, will be determined by the values that the

customers and consumers of the crops place on those products. If one person grows sour apples and few people want them, then his wage will be low. If there is a high demand for the good, the wage, over the time of production, will be high.

Workers will then tend to move from low-paid to higher-paid products, if they can. If workers are growing mangos and lemons, and one mango is trading for one papaya, but it takes twice as much work to grow the papaya as the mango, the mango growers will have a wage twice that of the papaya growers. If some workers are willing to switch from one crop to another, wages will tend to equalize among the crops or, more generally, among products, resulting in some overall wage level. More mangos will be produced and fewer papayas, reducing the relative price of mangos until one papaya trades for two mangos.

Hence, the principle remains the same in complex production, where we have many products and industries. The "law of wages," as Henry George (p. 213) called it, is that "Wages depend upon the margin of production." More generally, the wage level is determined by the margin of production, the least productive land in use, and the intensive margin where the next worker can get the highest wage without having to pay extra rent.

The Marginal Product of Labor

What happens if one of the farm owners that has hired a worker wishes to hire a second worker? The second one is also paid the same wage as people can get by working on their own farms, but is the extra product of this worker the same? We now turn to the interaction between wages and the productivity of workers on the same lot of land, or in the same factory or enterprise.

The "marginal product of labor" is the increase in total output achieved by hiring one more laborer. If by raising the work force from 50 to 51 a firm raises total output from 1000 to 1010 units, then the marginal product of labor is 10. The "value of the marginal product" is the physical marginal product (the extra goods produced) times the price of the product.

Suppose we have a farm of 100 acres (40 hectares). One farmer by himself might be able to grow 100 bushels of corn during a certain time.

If a second farmer is hired, the total product might grow to 240. The marginal product of the second farmer is 140, since the two can do some things that the first could not by himself. A third farmer might raise total product to 350. He adds 110 to product, less that the second, since there is less marginal benefit from the added cooperation and work. Although in any particular case, the first few added workers may each add more to output than the previous, eventually, the added or marginal product must decline, since the fixed factor, in this case land, will not yield increased output forever. This was the third foundational proposition of economics, as presented in Chapter 1.

The phenomenon of each new laborer (or other factor) adding ever less marginal product is called the "law of variable proportions," or, more famously, the "law of diminishing returns," also the "law of diminishing marginal product." Eventually, the diminishing marginal product becomes negative as workers keep getting added to a fixed amount of land.

The wage at the internal margin, or "intensive" margin (since a given lot of land or a factory gets used more and more intensively), must equal the wage at the extensive margin, due to competition among workers. It is possible that due to the high costs of labor or enterprise imposed by taxes and restrictive regulations, the cost of labor to an employer can be higher than the marginal product of labor, so that no more labor is hired, resulting in unemployment.

When the marginal product is greater than the average product, it pulls the average up, and when it is less, it pulls the average down. Therefore, the marginal product equals the average product when the average product is at its maximum. A rational producer, who wishes to have as high a profit as possible, will only hire workers when the marginal product is less than the average product but still positive.

Getting back to our earlier example, suppose that the margin of production is still at 10-bushel land, and one of the owners wants to hire a second worker. If the marginal product of the first worker is 10 but that of a second worker is only 9, the owner would not offer him more than 9 bushels as wages. But no one would want to be a second worker, since one could earn 10 as a first worker on his own land.

But suppose that the population grows and all the 10-bushel land is taken up. The extensive margin moves to the 9-bushel land. The owner

will now be willing to hire an extra worker, and will be especially willing if the margin moves to just below 9, so he can pay a wage of less than 9 and get some extra rent. In general, an employer will hire more workers as long as their marginal product is greater than the wage. Since the marginal product eventually declines, workers are hired just up to the amount where the value of their marginal product, that extra revenue produced by that extra worker, just equals the wage.

Since a firm will hire labor at the amount that equals its value of marginal product, and since that marginal product declines with increasing numbers of workers, a firm's demand curve for labor is exactly the relation between its value of marginal product and the number of workers, whether depicted as a curve in a graph or a table of numbers. The firm's demand curve for labor will thus slope down, since it demands more workers as the wage declines.

It should be kept in mind that though conceptually the demand for labor seems to be a precise thing, in practice the marginal product is a fuzzy, uncertain, imprecise amount, so the demand curve or relationship for labor, like any demand or supply curve, is a fuzzy rather than sharply defined line or table. Also, as we know, other factors can affect the demand for a particular worker, such as his looks, personality, ethnic background, personal relationship with an employer, and just plain luck! So the equation of wage with marginal product is a general tendency rather than exact description for every worker.

What about the demand for labor by an entire economy? It is the result of the demand for labor by all firms, but this "demand" itself is derived from productivity, since self-employed workers are their own demanders. We can envision a "production function" for the entire economy, i.e. total output as a function of the number of workers.

Since labor exhibits diminishing returns relative to the land in any particular region or economy, total output goes up with increasing labor, but at a slowing rate of growth, each extra worker adding a bit less to output than the previous. That extra output is none other than the wage of the extra worker, so we have a downward-sloping demand for labor as a whole in an economy, wages declining with increasing labor at any particular moment. But it is important to note, and note well, that this is a static relationship between labor and output. It applies to the amount

of labor at any particular moment in time, not to the addition of workers in an economy over time, which could also increase the division of labor and dynamically increase output per worker.

The supply of labor for an entire economy, or the market supply curve, is the quantity of the labor force (all workers plus the unemployed who want to work) as a function of the wage level. In other words, it is a curve showing the number of workers at each wage level. Its exact shape depends on the culture and demographics (of age, sex, family size) in a particular economy. It is possible in some places for the curve to bend backwards, or be upward-sloping at some wage level, because with higher wages, the workers will not want to work so many hours, preferring leisure to more consumer goods.

Generally, one would expect the curve to be rather flat at the subsistence level, since every family needs to eat, up to the number of families. Then it would slope up as second or third members of a family are willing to work at higher wages, and workers are more willing to work overtime or take less leisure. But then at a very high wage, the curve would become very steep, vertical, and then slope back as workers have a greater marginal desire for leisure time rather than more goods.

As with any market, the wage level would be determined at the intersection of the market supply and demand curves. With that type of market supply curve for labor, wages would be high if the demand curve crosses it at the steeply rising area. This would occur if the marginal product of labor is high to begin with, in which case the supply curve would become steep or vertical after all families have applied their labor. Increasing demand for labor, or productivity, would only raise the wage without increasing the labor supply much. But if the demand curve for labor hits the supply curve at the horizontal section, then an outward shift (increase) in demand would increase employment without increasing wages.

3. How to Create Unemployment and Impoverish Workers

The above analysis assumes that there is no tax on wages. If wages are taxed, then the worker receives a lower net wage, so if the supply curve is sloping up, then there will be less labor supplied, since the worker

responds to the take-home wage net of taxes. Thus, a tax on wages, such as a payroll or income tax, shifts the supply-of-labor curve to the left. As the supply curve shifts up along the demand curve, this increases the cost of labor to the employers. Employers must pay the gross wage, including the tax. The result is less employment at a higher cost to employers, and a lower net wage for workers. The tax is a "wedge" between the net and gross wage, which distorts or skews the market wage to employers and employees from what a pure market would yield.

To see the effect of taxes on labor, suppose there were a tax of $1 million per worker. Almost all workers would be thrown out of work, including the self-employed. The effect of a smaller tax is the same; the difference is only in degree. The higher the wage tax, the less employment.

Henry George (1883, p. 152) stated, "The essence of slavery is the robbery of labor." With chattel slavery, as existed in the 19th century and earlier, the slave owner expropriates the product of the slave's labor, beyond what the slave keeps to live on. "Free" labor has a choice of whom to work for, but if wages are taxed, labor is also deprived. It is not working for an employer puts a worker at a disadvantage, since in a free economy, he has the option of working for himself. Rather, it is being forced to work for others to the extent that part of one's wages is taken by government.

Another type of intervention in the labor market is minimum wages. If the minimum wage set by the government is higher than that of a market wage level, the quantity supplied of labor is increased, since more people want to work, while the quantity demanded is decreased, since labor is more expensive. The result is an excess or glut of workers wanting to work but not finding it: unemployment. Minimum wages especially affect teenagers and those wanting to enter the labor market, since they are unable to get entry-level work that would give them experience to enable them to get better jobs later.

Thus, if a government wants to reduce employment and keep workers unemployed, a good way to do this is to tax labor heavily and also enact a minimum wage, in addition to restricting entry into some types of occupations. This has been the policy of the U.S. and other governments, and it has been quite effective in keeping many workers poor and unemployed.

4. Labor Unions

Low wages and bad working conditions are two reasons why labor unions have organized. Trade unions arose out of the conditions of the labor force during the Industrial Revolution. Workers could gain bargaining power through collective action, of which the most potent weapon is the strike. Unions also became mutual aid societies, offering various services to their members.

Unions can be effective in giving laborers more bargaining power in a particular industry, but they cannot increase the overall wage level, since, as discussed above, that level depends on the productivity of labor, which cannot be increased solely by a labor strike. If an economy is divided into two labor sections, one with unions and the other without, then where they have labor monopolies, unions raise the wages of workers in those industries, reducing employment there. The workers thrown out of work will then move to the non-union sector, increasing the labor supply and so decreasing the wage level there.

Thus the effect of unions is to transfer income from the non-union section to the union section. The pushed-up wages in unionized industries also increase the prices of those products, decreasing the quantity bought, so part of the cost of these union wages are borne by consumers.

This does not imply that labor unions are harmful in general, only that they reduce employment and output in an industry when they have a legally enforceable monopoly power. Labor unions can and have been useful in organizing social benefits for their members and in serving as a way to communicate in an organized way with the management of enterprises to negotiate better working conditions. But unions by themselves, whether voluntary or coercive, cannot raise the overall wage level or decrease unemployment. As analyzed above, the way to maximize wages and employment is to remove the barriers, wage controls, and tax costs imposed on labor.

5. Raising Wages Through Education

As noted above, "human capital," education and training which increases the productivity of labor, is part of the labor factor. The general wage level is based on that of unskilled labor. Workers obtain a wage premium

above the unskilled wage level for their skill, talent, charm, and personal connections, and the scarcity of workers in the field. There can also be premiums or discounts due to discrimination and legal restrictions.

An individual worker can make himself more marketable relative to others by increasing his skill, including his skill at job finding. But when most workers attain similar skills, the comparative advantage of the skill will be lost, although there will still be an absolute advantage in being better trained. As we know, for education to increase productivity, it needs to be geared either to general skills such as reading and writing, or to the specific requirements of a field. A general education is also useful over the long run both for personal consumption, to better enjoy life, and to be a good citizen.

Education presumes the freedom to make use of it. When opportunities are blocked off, education makes a person frustrated. In some less-developed countries, young people obtain a university education and then find no job opportunities, other than the civil service, which expands to give them jobs, but without any productive purpose. In a free society, employment opportunities are abundant, and education does not need to be highly subsidized, since most families can afford to pay tuition. Enterprises seeking skilled workers also offer training and scholarships. If government schooling is still provided, it is in equal competition with private schools, and this market competition maintains the quality of the education as well as providing different cultural settings. In a multicultural society, the problem of what to teach is resolved by the freedom to start new schools that offer education geared to the interests of the students and parents. Competitive schooling not only provides training and knowledge, but also preserves the cultural capital that is part of our diverse heritage.

CHAPTER 3

Land

1. The Meaning of Land

Economic land" consists of everything except human beings and the wealth that they have produced. We will call it "land" for short. Land of economic value consists of natural resources, including underground minerals, metals, and oil; wildlife, including forests; the genetic variety of life; oceans, lakes, and rivers; the atmosphere; the electro-magnetic spectrum (for transmitting radio and television); and the three-dimensional surface area of the earth as sites for living and working.

By definition, an item is economically scarce if there is not enough to provide as much as everyone wants at a zero price. As the first universal proposition of economic theory states, some land is scarce. Of these types of natural resources, the most familiar in every-day life is the surface area of the earth, the space that we live and work in. This land obtains a market value due to its usefulness over time and the scarcity of land of good qualities. This value is called "land rent."

2. Rent

The word "rent" has several meanings.

First, in everyday language the "rent" of an office, house, factory, or shop means a payment for the use of property, which includes the use of land as well as produced wealth, such as buildings, cars, and computers.

In classical economics, "rent" had a special meaning as the amount that one pays solely for the use of land. Thus, when we speak of the "rent of land" or the "rental value of land," in economics we exclude improvements such as buildings and canals.

The term "rent" later became generalized in two directions. First, "rent" came to mean any payment that is more than necessary to put some resource into production. This is called "economic rent." For example, if a baseball star would be earning $50,000 per year in his next best profession, but is earning $600,000 per year playing baseball, the $550,000 difference is called "economic rent" because the player would play ball for just a bit over $50,001.

This meaning of rent became used also to refer to the "economic rents" received by those seeking privileges and transfer payments from the government without really earning it; they are economic rent because the funds do not put any service into production. The attempt to get such loot is called "rent seeking" (better called "transfer seeking") in the literature dealing with this, a body of theory called "public choice."

We can see that land rent is a type of economic rent, since the land is there from nature, so no funds are needed to produce the land. To simplify the language, "rent" here will refer to economic land rent unless otherwise specified. The term "rental" will refer to the actual payments of tenants to landlords, which may differ from the economic rent of the land.

The word "rent" in economics differs from the ordinary usage in another way. In ordinary language, we say someone rents something only when a payment changes hands. You rent a house from someone when you pay the landlord a check every month. But in economics, "rent" exists regardless of whether there are explicit payments. For example, suppose your parents let you live for free in a house they own. If they didn't let you live there, they could "rent" it out for $500 per month. Suppose further more that $200 of this amount is due to the land value, while the other $300 is due to the improvements, such as the building. Then they are losing $200 per month from the land rent which they could have collected. This $200 is rent even though no payments change hands. In effect, you the resident are collecting it, since you would otherwise be paying it if someone else owned it. So the amount of rent that a landowner could get on the market is economic rent whether or not the owner collects it in payments. If you own and work your own farm, the amount that you could have rented the land for is economic rent.

The value of land can be expressed in one of two ways: (1) the amount of rent offered for a fixed term of use, e.g. a week, month, or year; or (2) the transfer price when one obtains title for either a "leasehold" (possession of

land for a fixed term such as 99 years) or a "freehold" (indefinite possession). As stated by Henry George (1879, p. 166), "Rent is also expressed in a selling price. When land is purchased, the payment which is made for the ownership, or right to perpetual use, is rent commuted or capitalized. If I buy land for a small price and hold it until I can sell it for a large price, I have become rich, not by wages for my labor or by interest upon my capital, but by the increase in rent."

In the simplest case, with no taxes, no collection of the rent by a community beyond the title holder, no price appreciation, and no inflation, the sale price tends to equal to the rent divided by the interest rate: $p = r / i$. This is because $r = p * i$ (rent equals the principal or price of land times the interest rate), since the same funds (p) if loaned out at interest rate i would yield the annual amount r. If the money is inflating, then we need to subtract out inflation from the interest rates being paid in order to get the "real" interest rate I. If there is a tax on the land, or the collection of the land rent by a community, then the collection rate is added to the interest rate, since the rent must pay for both the tax and the net yield to the title holder: $r = p * (i + t)$, where t is the tax rate, the percentage of land value being collected, such as 5% of p. Therefore, $p = r / (i + t)$. Hence, as i or t or both increase, the price of land decreases. If rent increases, then of course the price increases.

3. How Does Rent Arise?

The value of land is due to a variety of sources. We can divide land into three types: 1) fixed material resources, 2) renewable resources, and 3) space. Fixed material resources include minerals, oil, metals in the ground, and air. Renewable resources include wildlife, the fertility of the soil, sunlight, and fresh water. The value of material land is due to its scarcity relative to human desires for these items. Oil and minerals are land as long as they are in their natural state; once people apply effort to extracting them or even exploring for them, then the value added is a capital good.

Some space on earth is not scarce, since one may go to the oceans or deserts and freely use all the space one desires. Space obtains value because for a particular use, in a given location, it is scarce; more space for that use cannot be obtained for free. For space, as recognized by Henry George (p. 166), "rent or land value does not arise from the productiveness or

utility of land . . . but upon its capacity as compared with that of land that can be had for nothing."

4. The Determination Of Rent

Suppose a new island arose in the middle of the Atlantic Ocean. It has many thousands of hectares and acres of fertile land. An international agreement allows anyone to settle on the land and claim plots of 100 acres or 40 hectares. To keep our model from being needlessly complicated, there will be only one crop: corn. The unit of output is bushels per acre per time unit. We will fix the time unit so that the best land grows 10 bushels in that amount of time. Also, to simplify at first, we will ignore capital goods and merge them with labor. Later, we will separate out capital goods. For now, there is only the original factors of production, land, and labor.

The foundational principles which were set forth in Chapter 1 can now be applied. Principle #2 states that resources vary in quality. We will let the land be divided into areas which grow 9, 8, 7, 6, 5, 4, 3, 2, 1, and zero bushels, depending on the sunlight, rain, soil, and elevation of various areas.

Now the first family arrives. In our model, there is only one farm worker per family, and all workers have the same ability and put forth the same work effort. Where, then, will the first family settle? Principle #12 states that people economize; they seek to produce a given amount with the least effort, or produce as much as possible with a given amount of effort. With land of different quality, the most will be produced with a given amount of effort on the best, most productive land. Naturally, the first family will settle on the 10-bushel land. Other settlers will also settle on that best land.

Since the best land is available for free, no one can charge a rent for the land that he possesses. So rent is zero. And since we are ignoring capital, all the production goes to wages. Since production on the best land is 10, wages are equal to 10.

Once the 10-bushel land has been settled, newcomers will go to the next best land, where they can grow 9 bushels. Wages on that land are therefore 9. But what about wages in the 10 land?

Suppose one of the owners of a 10-farm wishes to retire, and hires someone to work the farm. He offers a wage of 8. No one from the

9 land will take up the offer, since they can get 9 working on their own farms. If he were to offer 9.5, everyone from the 9 land would want to get hired since that is more than what they can get on their own farms. So competition will set the wage offered at just 9, where someone is indifferent between working in the 10 or the 9 land. Wages everywhere are 9. Note that, as discussed in Chapter 2, this is the principle of how wages are determined: they are set where the best land is available without rent, since any wage offer below this will not be accepted and any above this will have many competitive takers who will drive down the wage.

But what about that extra bushel on the 10 land after the wage of 9 has been paid? Since this is not wages, it must be a return to the other factor, land. The 1 bushel left over is rent. The 10 land has acquired a rent of 1 when cultivation moved to the 9 land.

As discussed in Chapter 2, the boundary where people are settling on the best free land is called the "margin of cultivation." More generally, for land of all uses, the boundary is called the "margin of production." The term "margin" means the edge of consumption or production, where the last unit of a resource or the last item of consumption is being used. The margin of cultivation is the very next acre or hectare of land that would be occupied for farming.

Suppose now that after that one farmer hired on the 10 land, the owner decides to hire a second farmer on the same land, since that would increase output even more. He pays the second one 9 bushels and sees that the yield is now 18.5. The second farmer only added 8.5 bushels to output. The marginal product of labor is 8.5, while the wage is 9, so the second worker is dismissed. As noted in Chapter 2, labor will only be hired in a competitive economy if the marginal product is higher or at least equal to the prevailing wage.

When all the 9-bushel land is taken, the margin of cultivation moves to the 8-bushel land. Wages are now 8. Hence, rent on the 9-bushel land is now 1, and rent on the 10-bushel land has gone up to 2. Now that wages have been reduced to 8, the owner of that 10-bushel farm can hire that extra worker. The marginal product is 8.5, leaving the owner 1.5 bushels as rent. So all the owners of the 10-bushel land hire two workers (those working their own farms hire themselves and another), increasing their rent to 2.5. Land in the 10-bushel area rents for 2.5, since that is

what can be obtained by hiring the optimal number of workers for the maximum possible rent. The population density on the 10 land will now be twice that of the other lands.

We can now derive the determination of rent in general. Rent in a certain location is the highest product of land above what can be produced at the margin of production, where rent is zero, after paying for the factors of production other than land. This is called "differential rent" because the rent at a location arises from the differential or extra product it yields (minus costs for labor and capital goods) compared to land at the margin. All land rent is thus differential rent.

This differential rent is also the marginal product of land, the extra amount of product obtained from using an extra amount of land, keeping other factors constant. Hence if with two workers one gets a tiny additional amount of 10 land, it too will yield a proportional equivalent of 2.5 (18.5 minus wages of 16) per acre too, and the marginal product after subtracting wages will be 2.5 per acre. So the marginal product of land is the same as the differential rent.

More settlers arrive; the 8-bushel land gets filled, and the 7-bushel land is settled. Wages fall to 7. Rent on the 8 land is now 1. Rent on the 9 land rises to 2. But those owners too now hire a second worker, since the product of the second worker on the 9 land is 7.5, which increases the total product to 16.5, which after paying 14 for wages, leaves 2.5 in rent.

In the 10 land, a third worker will have a marginal product of 7 and rent is now 4.5 (3 from the product of the first, plus 1.5 from the product of the second worker). Owners will be indifferent to hiring a third worker; some may and some may not, depending on whim or chance.

When the margin of cultivation moves to 6, the 7-bushel land gets a rent of 1 and the 8-bushel land rent rises to 2 if there is one worker. Keeping the marginal product of the second worker at 1.5 less than the first, second workers are hired on the 8-bushel land, since their marginal product is 6.5, raising the total rent there to 2.5.

Rent on the 9-bushel land rises by one bushel to 3.5 (a third worker's product is 6, just equal to the wage). In the 10 land, total product with three workers is 10 + 8.5 + 7, totaling 25.5, with wages 3*6=18, leaving 7.5 for rent.

Can you see the pattern that develops? As the margin of cultivation moves to ever less productive land, wages go down and rent goes up. The owners of the best land obtain higher and higher incomes due to the increase in population and the decrease in the marginal productivity of land. Incomes thus become more and more unequal. As settlement proceeds further, wages will be driven down to the subsistence level, where workers are just able to survive—a level in fact being earned by many of the poor around the world, including in developed countries.

5. Land Speculation

In the above model, we assume that each farmer gets a lot for actual use. But there is another motive for getting land. Folks will notice that the rent is going up and up on the older lands as the margin moves to newer land that is less productive. So some sharpies will obtain land not just for use but to get the increase in the rent in the future.

Suppose, for example, that the 8-bushel land is being settled. The speculators will try to claim as much of the 8 land as possible, since it is free now but will have a rent when the margin moves to 7. So those wanting some 8 land for farming will find that the 8 land will be all claimed very quickly. When production moves to the 7 land, farmers can either rent land in the 8 region or claim 7 land. They will prefer to claim 7 land in the hopes of getting rent when the margin moves to 6, so much of the 8 land will remain vacant for a while. Eventually, the sites in 8 land get rented, but there is now a rush to occupy 7 land, and then 6 land, and so on, leaving much of it vacant as the margin quickly moves to ever less productive land.

So the effect of excessive land speculation is to move the margin out much more rapidly, reducing wages and increasing rents that much sooner. Excessive land speculation also increases the price of land for current use, since the price reflects the expected future usage. Land speculation is excessive when the land value is subsidized by government, as the infrastructure and services of government raise rents, while financed by taxes on labor and capital rather than having landowners pay back the generated rent.

6. The Effect of Capital Goods

Capital goods will be examined in the next chapter, but let's fill out our model briefly by including them now. Let's start again in the 10-bushel land.

Suppose that farmers were somehow growing the corn with their bare hands, but then someone invents some tools that enable them to double production. These tools, however, only last one month. If half the farmers spend a month making the tools instead of farming, the farmers as a whole are no better off, since only half the workers are now farming, so at twice the crop per worker, the total yield is the same. Hence, the resources needed to make the tools must more than offset the greater productivity from the tools.

Suppose instead that the tools double the monthly output per farmer and last three months. A farmer could make tools one month and use them for three months. The total product would be 60 for the four months, for a monthly average of 15. So the marginal product of the tools, the capital goods, would be 5 per month. Some of the workers might become full-time tool makers. They would trade their tools for 15 bushels of corn every month. They therefore earn a wage of 15, and the farmers also earn 15 per month, after growing 20 bushels and paying 5 to the tool maker every month (for a total of 15 for three months).

We can see that the capital goods have increased productivity by 10, but only 5 is paid to the tool maker, so wages have increased by 50%. The increase in productivity is split between wages and the yield from making capital goods. (We ignore interest rates, which would only slightly affect the calculations). This is because, at the margin (where some workers are indifferent between farming and tool making), wages are equalized, so the return to making capital goods will tend to equal that of farming, assuming the quality of labor is the same.

With the tools doubling productivity in lands of all qualities, we can see that adding capital goods complicates the model but does not alter its essential principles. After the 10-bushel (now 20 bushel) land is used up, the margin will still move to the 9-bushel land (now doubled to 18). The now 20-bushel land will then acquire a rent, while wages will drop in the 18-bushel land as well as in the 20-bushel land from 15 to 13.5 (9 plus 4.5).

Even if we suppose that tool makers do not pay rent (living with farmers, for example), the rent on the 20 bushel land is now 2, having doubled while wages have only increased by 50% (since the other 50% is paid to the tool maker). So the effect of the capital goods in this example is to increase rent in proportion to the increased productivity, while only increasing wages by half the increased productivity. In general, the proportional increase in rent will tend to be higher than that of wages.

Wages will be higher because of the capital goods, but still diminish, along with the yield from making capital goods, as the margin is moved to less productive land. And if the capital goods enable one to use land that previously was unproductive, the margin might be extended to the level where wages are no higher than they were without the capital goods.

So the accumulation of capital goods and technical progress, including a more efficient division of labor, can increase wages, offsetting the effect of the diminishing marginal product of labor, but if the margin of productivity then moves again to less productive land, the benefit of this increased product will again end up going to rent rather than wages or the owners of capital. In studying wages during the first decade of the 2000s, some economists concluded that average wages have remained about the same, while per-capita output rise. One explanation could be that the increase went to higher rent, and indeed, until the real estate boom crashed in 2008, land values were rapidly escalating.

7. Urban Rent

The Austrian economist Friedrich von Wieser (1927 [1967], p. 340), an early theorist of urban rent, stated that "Urban rent is that part of the rental which is paid as a premium for the advantages of the better location." Urban rent arises with the presence of a population and its economic activities.

That people and their collective activities give rise to rent can be seen by looking at any densely populated city. Equally, one can look at those places that have "gone back" to their uninhabited or primitive state. Ghost towns in the American West, for example, which have collapsed through lack of industry or have been over-exploited, "mined out". When the town prospered, land titles had a rental value. Now, with the

disappearance of the population, these land titles have become valueless. Rental value of land will clearly arise and collapse with that of populations and their activity.

Henry George (1879) theorized that the greatest effect on rent was the presence of communities rather than the extension of the margin of production to inferior land. He illustrated this effect with a story about the "unbounded savannah," a field "stretching off in unbroken sameness of grass and flower . . ." (p. 235). Along comes a first immigrant family. Nature is rich with resources, but this single family has to provide for all its needs, so though they have enough to eat, they have little wealth.

Another family comes along, and though the land is the same everywhere, "there is one place that is clearly better for him than any other place, and that is where there is already a settler and he may have a neighbor" (p. 236). The two families may now cooperate to produce wealth previously too difficult for one. (Although if the first family likes solitude or has a lifestyle (such as nudism or loud music) that the second does not like, then the second one might settle just far enough away for privacy but close enough for cooperation when needed.)

As more settlers arrive, they tend to locate near each other. They may form several communities with different values and lifestyles, but there will tend to be one major settlement where many services have become available, and those smaller outer communities also join in to form one greater community. "Labor has now an effectiveness which, in the solitary state, it could not approach" (p. 237). They can cooperate to accomplish large tasks, and can also create a division of labor for specialized work such as teaching. "Satisfactions become possible that in the solitary state were impossible."

Now, says George, go to our first settler and offer him the full value of all his improvements. "He would laugh at you. His land yields no more wheat or potatoes than before, but it does yield far more of all the necessaries and comforts of life . . . The presence of other settlers—the increase of population—has added to the productiveness, in these things, of labor bestowed upon it, and this added productiveness gives it a superiority over land of equal natural quality where there are as yet no settlers" (pp. 238–9).

Let the population continue increasing, and now the town has grown into a great city. The "division of labor becomes extremely minute, wonderfully multiplying efficiency . . . Hither run all roads, hither set all

currents, through all the vast regions round about. Here, if you have anything to sell, is the market; here, if you have anything to buy, is the largest and choices stock" (p. 240). Here are the great libraries, specialists, and center of commerce and government.

"So enormous are the advantages which this land now offers for the application of labor, that instead of one man with a span of horses scratching over acres, you may count in places thousands of workers to the acre, working tier upon tier . . . All these advantages attach to the land, . . . for here is the center of population—the focus of exchanges, the market place and workshop of the highest forms of industry. The productive powers which density of population has attached to this land are equivalent to the multiplication of its original fertility by the hundredfold and the thousandfold. And rent, which measures the difference between this added productivity and that of the least productive land in use, has increased accordingly" (p. 241).

The increasing rent that is generated by an increasing population and the growth of communities thus comes about "not so much from the necessities of increased population compelling the resort to inferior land, as from the increased productiveness which increased population gives to the lands already in use" (p.242).

8. Rent as Surplus, and Why Land is Different from Labor

Sir William Petty (1623–1687), an English economist, was among the first in Europe to examine the nature of rent. Petty regarded the rent of land as a surplus arising after the labor costs of production have been met.

Adam Smith (1723–1790) presented the theory in similar terms. Rent, according to Smith, was a surplus which arose after the basic costs of production had been met. Thus, improvements in the efficiency of production, which reduced costs, raised the surplus income, and subsequently translated into higher rent: "All those improvements in the productive powers of labour, which tend to directly reduce the real price of manufactures, tend indirectly to raise the real rent of land . . . Every increase in the real wealth of society, every increase in the quality of useful labour employed within it, tends indirectly to raise the real rent of land" (1776 [Book I, Chapter XI, Conclusions], pp. 275–6).

A more complete explanation of how rent is measured was developed by David Ricardo (1772–1823). In 1817, he published the *Principles of Political Economy and Taxation,* in which he developed what has since been termed "The Law of Rent." The law states: the rent of land is determined by the excess of its produce over that which the same application of labor and capital goods can secure from the least productive land in use. As we have seen above, this law needs to be qualified to take into account the intensive margin of land, where one lot is being worked with more and more labor. Because more labor is used by the more productive land, lands do not all have the same application of labor and capital goods; land rent will be even higher in the more productive land due to its greater intensity of use, as we have seen in the above model.

9. Generalization to all Land

The agricultural model used above can be generalized to land for all uses. For example, a grocery store in a sparsely settled area will have much less sales volume than one in the center of a large town. One would then expect much of this volume to be distributed to the owner of the site as rent, and indeed, rent in urban centers is much higher than in the rural boondocks. Offices in the center of metropolitan areas pay much higher rent than those in smaller towns. Generally, productivity of any sort generates higher rents relative to lower productivity. The law of rent applies to all land.

An important aspect of productivity is the availability of transportation, both of roads and of vehicles such as trains. This can best be understood by reference to a historical example. During the colonization of Australia and New Zealand, the first settlers to arrive naturally took up the most desirable sites. These sites were nonetheless looked upon by the British Colonial Office as marginal land which commanded little rent. The land was therefore sold for a low price. As further settlers arrived, they occupied less desirable sites. Eventually, the originally occupied sites, which previously had no value, were now prime sites in the center of towns and cities.

Those late arrivals settling on the new margin—remote sites far away from the center of town—faced increasing costs for transport to bring their produce to town. To have occupied sites closer to their markets,

they would have had to bid a higher rent. The sites closer to town have lower transport costs and therefore higher profits per bushel of output, hence are more productive in terms of revenues minus all costs. This extra productivity induces the higher rent closer to town.

The theorist who developed the economics of location, including transportation, was Johann Heinrich von Thünen (1783–1850). In his work *The Isolated State* (1826), he explained that there was a relationship between transportation costs and the rental value of land. A farmer working on the periphery of a market area has the furthest to travel, therefore his land would have a low or zero rent. But suppose the roads leading to this marginal land were improved by the government. The reduction in transport costs results in a higher income per bushel of crop. (Similarly, improvements such as refrigeration further reduce the cost of transporting produce.) As we have seen in the above model, such increases in productivity increases the rent for the land affected. So much of the productivity due to the physical infrastructure—roads, trains, busses, communications—ends up as increased land rent.

Typically, financial firms, such as banks, have sought to locate in the center of town. A close proximity to commerce, government, and customers was needed before the advent of telecommunications. Despite modern communications for both voice and data, including money, financial firms still seek to locate themselves within the center, and rents in financial centers are among the highest. This is because of the extremely high rewards which banks, insurance headquarters, brokerage firms, etc., stand to gain through their location on these sites as opposed to locating on marginal sites.

The same principle applies to the location of commercial office blocks within a city. A firm will wish to locate its offices on a site which has easy access to the services on which it depends, such as transportation (subways, busses) for its customers and workers, and it might also need access to centrally located financial or governmental agencies. There will be a location beyond which the office cannot afford to locate itself—where the costs of getting its staff to work and of not being closer to complementary services becomes too high. Turning now to industrial activities, these too will have a margin beyond which it is unprofitable for them to locate. Some industries will be attracted to "economies (advantages) of density"

as well as by transportation, such as access to a major highway. Thus, the rent of this land too is determined by the increased rewards which the industry stands to gain through its location.

Housing also has a margin. If a family chooses to locate in the countryside, beyond the denser residential developments, it might find that communing costs (including time), longer access to facilities and lack of public amenities make its location economically unattractive. Rent is therefore lower, other things being equal, at the edge of town than nearer to the center, although of course negative factors such as crime and noise and crowding will decrease the desirability and thus the rent in parts of town that are run down, even if near the center. With cars and long-distance communing trains and busses, people can live far from city centers, but still, one would not normally live hundreds of miles and kilometers from a city center; there is some limit to commuting times. And rent in the more desirable neighborhoods near the center of a city will fetch that much more rent than sites yonder.

Hence, each activity has some margin beyond which it will be unprofitable to locate. Generally, a city's economic activity takes place within the margins of the various functions. The corresponding rents are based on the rewards of location for an activity.

Much of land rent is generated by infrastructure and civic services. If landowners do not return the rent to pay for these goods, then they receive an implicit subsidy, and tenant-workers get double billed, having to pay both taxes and higher rent for these public goods.

10. The Supply and Demand for Land and Rent

The quantity of fixed natural resources diminishes with increasing extraction, though the supply of known reserves of the resource increases when new sources are discovered. The supply of renewable resources is variable, making conservation and renewal essential if the supply is not to become extinct.

The surface space of the earth, however, is constant. Land area within any boundary is fixed in two ways: no space can be imported, and new space cannot be manufactured. Hence, for any particular region (given some boundary line), the amount of space is fixed. In a graph where the

quantity of land is measured along the horizontal axis and the rent of land is on the vertical axis, the quantity of land is a vertical line at the amount (acreage, area) determined by the boundary. The supply of land for a particular use, or the supply of lots offered for sale, can vary with price and be upward sloping, but the quantity of surface sites within some area cannot be expanded or contracted.

The demand curve for a particular plot of land is the amount wanted at various prices. The demand curve slopes downwards, as greater amounts of land, like any product, are wanted at lower prices. So the diagonally downward-sloping demand curve cuts through the vertical supply curve at some point, determining the price of land. At the point where the curves intersect, of course, this demand is equal to the marginal product of land, the rent determined by its differential with respect to free land, or its capitalized value as the price of land.

If the demand for land in that area increases (more is wanted at each price), then the price of land will rise. But the supply will not be affected, since it is fixed. Also, if all or part of the rent is collected by a community, the supply of land will still not be affected. If the tax is higher than the rent, then people will no longer want to own land, and the tax will decrease the value of the capital goods tied to the land. So long as the amount of rent collected is not greater than the rent, it has no effect on the demand for the land and thus neither reduces nor increases the rental paid by the tenant. If the landlord is already charging as much rent as possible, the entire collected rent is borne by the landlord, and none of it can be passed onto the tenant.

However, the burden of a tax on land value is only borne by the owner at the time the tax rate is increased. Because the price of land falls, a new owner basically has no burden, because the tax payment replaces what would have been paid as mortgage interest (or the implicit return on his down payment). Since land rent is a pure surplus, after the transition, a tax on the rent or land value is not a burden even to the landowner.

11. Monopoly Power in Land

At a speech given in Edinburgh, Scotland, in July, 1909, Winston Churchill observed "it is quite true that the land monopoly is not the only

monopoly which exists, but it is by far the greatest of monopolies—it is a perpetual monopoly, and it is the mother of all other forms of monopoly" (Churchill, 1909, p. 149).

An "absolute monopoly" means an industry with one seller. But the term "monopoly" has a broader meaning. An "entry" monopoly exists when it is impossible to expand the supply of a good; a new firm can only buy title of an existing supply from the previous title holder.

Since land in a specific location is unique, each plot has a locational micro-monopoly. But there is another sense of the term "monopoly" having to do with entry and exit into an industry. In a competitive industry, firms can enter not just to increase the number of firms but to increase the production of the output. Moreover, the product can also be imported when profits are above normal. Increased supplies reduce the profits in the industry to normal returns. But when the stock of the product is fixed, when the expansion of output is impossible, then this competitive condition does not exist, and in that sense, there is a monopoly of the product among those firms who share in the fixed stock (Foldvary, 1993). The economic rent can persist indefinitely.

The nature of land as a monopoly is furthered by the fact that it keeps indefinitely. Land as a locational site doesn't spoil or rot. When an area is developing and the price of land is rising, the title holder profits simply by owning the site even if it is not rented out. It may be more profitable to avoid building at present and wait until the other sites are developed, when the real estate can then be sold at higher price. Keeping land out of use or in inefficient current use (such as parking lots in a city center) is detrimental to current production. It is not a pure market phenomenon, but a subsidy to the landowner from receiving rent generated by the community. and the absence of the collection of that land rent by the community.

12. Urban Sprawl

Urban sprawl is the difference between the use of urban land and the amount that would be used, for the same output, in a pure free market. It is an excessive urbanization of the countryside surrounding a city relative to that which would occur in a pure market economy. A pure market economy, with land rents collected for public revenue, and no zoning laws restricting the efficient use of land, would result in compact cities

where density gradually decreased to the edges instead of helter-skelter hodge-podge developments.

Sprawl damages the surrounding agricultural and natural land, decreases the efficiency of cities and leads to a considerable waste of infrastructure such as streets, highways, pipes, and lights. Much of the inefficiency consists of longer commuting times, more costlier transportation for agricultural and urban goods, and wasted fuel. Though the loss of good crop land through urban sprawl is a needless waste, the damage goes further. It induces farmers to move further afield, destroying wildlife habitats.

Zoning laws often mandate maximum densities, requiring plots to have minimum sizes. Changes in zoning have been required in planned communities in which homes are clustered together, leaving more room for shared open space (Foldvary, 1994). City laws also often restrict the number of people able to live in a residence if they are not related by family, and restrict the ability to rent out rooms. Laws also prohibit enterprise at home as well as mixed use of land.

Also contributing to urban sprawl is the subsidy of the public works serving the outer edges. The streets, freeways, water and sewer pipes, lighting, security, fire service, parks, schools, and other goods and services are provided at the expense of the taxpayers of the entire city or county, often with the aid of higher levels of governments, so that these are subsidies which the users of suburban land consider free. Not having to pay its cost increases the usage and demand for these goods.

Taxation is a major contributor to urban sprawl. Vertical use of space is penalized by the taxation of buildings, while the horizontal use of space is subsidized. The remedy is the elimination of property taxes on improvements such as buildings and the funding of civic services from rent. In most cities, the high value land in the center is used inefficiently due to the secondary use of land as speculation. Land owners of central sites often have no incentive to redevelop their old buildings and can even leave central sites vacant. New development is displaced outward.

13. Farm Subsidies

David Ricardo showed that if the state artificially raises the value of a product, such as corn, the value of the land that produces the good will rise. Ricardo wrote about the Corn Laws passed in England in the first

decade of the 19th century. He showed that if you tax imported wheat (i.e. "corn") to raise its price to protect the British farmer, the rent of the corn land will go up. Rent will continue to rise until it has effectively wiped out the benefits to the farmer renting that land.

The *Lloyds Bank Economic Bulletin* (1992) reported that of the money spent under the European Economic Community's common agricultural policy, "around half is transferred to the land-owners and the rest is lost in inefficiency. Poor farmers and farm laborers appear to gain little." In Great Britain, it was calculated (Body, p. 209) that the total spent on agricultural subsidies since WWII is equal to the value by which farm lands have risen!

14. Application for Public Revenue

As the factor of production that does not involve human exertion, and due to the fact that land is in fixed supply, land is the efficient and equitable source of revenue for governance and public goods. Chapter 16 on public finance discusses public revenue from land rent. Having analyzed the two factors labor and land, the next chapter discusses the third factor, capital goods.

CHAPTER 4

Capital as Tangible Assets

1. Definition

In Chapter 1, capital goods are defined as goods which have been produced but not yet consumed. Consumption means the using up of economic value, the way fire consumes wood. Production is the creation of economic value. If economic value is still there, such as for a car that still runs and has resale value, then the item remains a capital good.

An "economic investment" is an increase in the stock of capital goods and human capital (education and training). Thus, all production is economic investment. But the term "economic investment" is usually applied to the production of capital goods used by producers.

The economic consumption of a car is its depreciation, not its purchase. Capital goods tend to depreciate, to have their value used up, for two reasons. One is the tendency of most capital goods to wear down unless constantly maintained. The maintenance or repair is a production of new capital goods attached to the old one. The second cause of depreciation is obsolescence, as with an old computer that still runs, but has been superceded by bigger and faster machines.

The economist Nassau Senior is credited with forming "the abstinence theory of capital accumulation" (McConnel, 1980, p. 67). The theory states that land and labor are the primary factors in production; they are used to produce the third factor, capital goods. In order to provide tools, as illustrated by the model in Chapter 3, it is necessary to abstain from the production of immediate consumer goods in order to make the tools instead. Capital goods are produced to be used at a later stage in the production of more wealth.

Capital goods are more than inert matter. Embedded in them is technology that multiplies the productivity of land and labor, like a lever that enables one to lift a much heavier object. Capital goods also enable us to increase our division of labor, specializing in more fields and increasing the productivity in each field. For example, computer programmers and users can specialize in particular types and applications of software, and artists can specialize in particular types of media.

2. The Yield of Capital Goods

Just as labor has a return as wages, and land as rent, capital goods have a return, which we can simply call a capital yield, with the understanding that this "capital" is capital goods. Economists once called this yield "interest," but this confusing usage is now obsolete, since interest earned from financial capital is not necessarily a return on capital goods.

The return on a capital good has two components. The first is that due to labor. The depreciation of capital goods include the using up of the labor value in the capital good and the using up of the value of the natural materials that went into the capital good. If you buy a tool that gets used up—depreciates—in one day and also takes one day to produce, the seller's return is, to him, a day's wage for his effort, which is also equal to the using up of the tool that day, in addition to the raw materials. Each day, he will make another tool that gets used up in one day. Wages will equal the labor portion of the depreciation.

But suppose, on the other extreme, that the tool lasts forever. The maker does not sell it, but loans it out. Since the tool does not depreciate, there is no labor component to the return of the owner. The labor portion was paid when the owner bought the tool. The tool never loses its value, so the labor component is retained as perpetual sales value, which the maker can obtain by its sale. The annual return on the tool is then the second component of a capital yield, the implicit interest. One could sell the tool and buy bonds that pay interest, and so there is an opportunity cost in owning the tool, the foregone interest, and thus the tool yields implicit interest.

Most capital goods neither depreciate immediately, nor last forever, so they will have a capital yield composed partly from depreciation and partly from implicit interest.

The fact that capital goods depreciate implies that a user of these goods needs to maintain or replace them. This capital maintenance is a cost of production even though it is not explicitly paid during some time interval; an accountant will enter an amount for depreciation as a cost. Hence, to calculate the economic profits or net income from an enterprise, depreciation needs to be subtracted from the gross income.

3. Roundabout Production

Productivity can be increased if some production is devoted to indirectly producing a product, such as making the tools that the firm then uses to produce the final product, the crop. This is called "roundaboutness" or "roundabout production." Production becomes even more roundabout if some production is devoted, say, to making the steel and wood that is then used to make the tools for use in farming. As production becomes more roundabout, it takes longer from making the highest level capital good to its final use in the production of consumer goods. This increase in roundaboutness thus lengthens the period of production, the cycle of time needed to make the final good. This concept of greater productivity from greater roundaboutness of capital goods was developed by economists of the Austrian school of thought, especially by Eugen von Böhm-Bawerk.

The founder of the Austrian school, Carl Menger, originated Austrian capital-goods theory with the concept of "goods of higher order." Goods of lowest order are those with a quick turnover. Goods with a longer time of turnover are higher-order goods. Goods of the highest order require many years to mature into profit; the main higher-order capital good is real estate construction.

Menger also recognized that the time structure of capital goods has a relationship to the interest rate. Think of capital goods as a stack of pancakes, the bottom stack being circulating capital such as inventory that lasts only a few days, and the highest stack being real estate construction. When interest rates are lower, then it is more profitable to invest in goods of higher order, and when interest rates are relatively high, the rate of return is quicker and the structure of production flattens.

Suppose a tree takes 50 years to mature, and grows at at 3% per year. If the interest rate is 4%, the tree will not be planted. If the interest rate falls to 2%, then the tree is worth planting. When people save more, the

interest rate falls, and there is more investment in higher-order goods, and that is offset by less consumption (since there is more savings), so all is in balance. Thus the interest is an equilibrium variable that adjusts to equalize savings and borrowing, and thus also savings and investment.

The structure of production, as affected by interest rates, has implications for business cycles, as worked out by F. A. Hayek, another Austrian economist. If money is injected into an economy by monetary policy, artificially reducing the interest rate, the capital structure will deepen as higher-order investment is stimulated. But since intended savings have not changed, consumers have not really reduced their demand for consumer goods, so this extra investment is not economically warranted. The result is inflation and a distortion of relative prices. When interest rates return to their previous level, the effect of the extra money having been dissipated by rising prices, the structure of capital flattens again, and investment in higher-order capital goods, such as buildings, slows and stops, as the recently produced goods become "malinvestments," wasted capital goods.

We saw this happen during the recession that began in December 2007. There was a large inventory of unsold and abandoned houses. Construction stopped, workers in real estate lost their jobs, and the fall in investment brought down the economy. (Business cycles will be covered depth in Chapter 12.)

If the use of capital goods achieves higher levels of return, we must have a means of measuring the increased yield. This takes the form of a rate of return or yield, also called "net productivity," which can be expressed in the form of a percentage per annum. To determine the rate of return of a capital project, first we calculate the costs of the factors we employ to undertake our capital project; then we calculate the total returns we stand to gain from the capital project, using the present value of the future gains. The excess of returns over costs will be our net productivity, and its ratio to costs the rate of return. Only if the total returns add up to more than the original costs do we have a positive net productivity, and only if the rate of return is greater than the prevailing interest rate (the alternative use of the funds) is it worth undertaking the capital project.

The rate of return on an investment in capital goods depends not only on the technical productivity of the roundabout process but on the market structure. For example, if a new process is protected by a patent,

there will be less competition, and so the firm may charge a higher price for the output and get a greater return on the capital good. An entrepreneur creating capital goods needs to determine how much to invest in order to maximize his future return. We have already done some marginal analysis in previous chapters, so we need only apply the same principle regarding costs and benefits. A profit-maximizing firm will increase production up to the point at which marginal revenue equals marginal cost. If the marginal costs are increasing and the marginal revenues are decreasing, any more output and the firm would not be recovering its costs, and any less would result in the firm would not generating possible profits by expanding output.

4. Rental Returns to Capital Goods

Some types of capital goods, such as infrastructure (streets, parks, transit, utilities) yield returns which manifest themselves in higher rentals and site values. An example would be a dam that prevents flooding and greatly reduces insurance costs. The civic goods provided by private communities, such as residential associations, shopping centers, and large resorts often raise both the land value within the community and that in the neighborhood. Since the improvements are capital goods rather than natural resources, this increase in the value of the sites is not land rent, but a capital yield, which takes the form of a rental on land as well as for the buildings. If the improvements were provided by private enterprise, then the generated rentals would be paid to those firms, by contract. If these capital goods are provided by government, then a tax on the value of the sites is partly a payment to government as provider of the civic goods.

As another example, the provision of water systems benefits the affected land-owners. This effect was documented in detail by L. R. East (former Chairman of the State Rivers and Water Supply Commission, Victoria, Australia). He wrote: "the real profits resulting from irrigation development lie not in the sale of water, but in the increases in business activities and in land values resulting from that development." East sites the "spectacular development" of the town of Sheparton, lying within one of the irrigation districts, as evidence that "there are very real benefits received by other sections of the community." Irrigation in Sheparton

ultimately increased the value of land to 100 pounds a foot (1940's value) on its main business street (East, 1945, p. 7). A similar story to that of water supply could be told for the supply of railways, roads, and bridges. The process by which land-owners benefit is as follows: "the carrying out of public works such as roads, railways, and water supply makes possible increased production from the land, or more intensive use of the land, and as practically the whole advantage goes to the owner—as distinct from the worker engaged in production—this advantage is capitalised in increased land values" (p. 27). Some countries have had the vision to tap this rental capital yield in financing capital investments. Public works are thus funded by the subsequent increase in rent and land values. The huge Aswan Dam, which supplies Egypt's Nile Valley, was financed by an increase of 0.5 pounds an acre in the land tax over a very large area which received summer irrigations from the reservoir (p. 24).

The idea of meeting capital charges from a special tax on land values was also adopted by Canada as far back as 1912. A water corporation was established by legislation to supply water in bulk to the municipalities comprising the greater Winnipeg water district. The public works cost approximately 17,000,000 dollars, and from 1912 to 1927 the whole of the revenue required to pay interest and sinking fund was raised by a special levy on land values—exclusive of improvements—of all the lands within the district (p. 26).

The private provision of infrastructure charges the users, such as by tolls, rentals, and fees. The governmental provision of public goods would best copy what private enterprise does, rather than double bill users by taxing them in addition to their paying higher rentals.

CHAPTER 5

Interest and Its Calculation Over Time

1. Time Preference

Interest" is the premium that is paid in order to shift a purchase from the future to the present day. Foundational proposition #14 states that people tend to prefer goods at present to those in the future. This is called "time preference." Given a choice, which would you prefer, $1000 of goods today, or $1000 of goods (adjusted for inflation) thirty years from now?

Most people most of the time would rather have the goods now, for two reasons. First, the human life span is limited, so the sooner one gets something, the more time one will have to enjoy it. Secondly, the future is uncertain, as stated by foundational proposition #15, and we don't know what our situation will be in the future. People also want to buy some products to have them available just in case they need them, even if not for immediate use. People wanting goods now often borrow them, since they don't have the savings to spend to obtain them. When you borrow funds to buy goods with, such as a car or house, instead of waiting until you have saved up enough funds to pay full price, you are shifting the purchase from the future to the present day.

Since future goods are less desired than present ones, they sell at a discount relative to present-day goods. One pays more for present-day goods by paying a premium, which is interest. The rate at which future goods become discounted becomes the natural rate of interest, the interest that would prevail in a pure market. Pure interest based on time preference excludes inflation, risk, and overhead expenses.

We could say, half-jokingly, that interest rates prevent everyone from doing everything at the same time, just as land rent prevents us from wanting to do everything in the same place. The American economist Irving Fisher (1867–1947) built upon the analysis of Menger and Böhm Bawerk in his book *The Theory of Interest* (1930). Fisher explained that greater productivity induces people to shorten their time preference in favor of borrowing more today in order to reap the greater gains, increasing the interest rate. As capital goods accumulate, their increase in productivity becomes reduced, due to diminishing returns, and thus the effect is to lengthen time preferences and reduce interest rates. However, technical progress can offset this by making new types of capital goods, such as computers, more productive.

2. Types of Interest Rates

Part of what is called "interest" in the payment for a loan is not really interest, but overhead costs: the labor and other costs of operations of the lender. Another portion of what is called "interest" is really a risk premium, an amount paid to make up for bad debts. There can be a monopoly premium if there are legal restrictions on lending, reducing the supply of loanable funds. After these premia, if there is no inflation, then the depositors of a bank or the owners of the institution receive the remainder, pure interest.

When the money is inflating, part of what is paid to the saver is not pure interest, but makes up for rising prices. The quoted rate, and what is paid in money, is the "nominal rate" of interest. If we subtract the inflation rate from the nominal rate, the "real" rate of interest. It is the real gain, because we have taken out the inflation, leaving the purchasing ability.

The real rate of interest is used in the capitalizing a flow of funds into a stock, or what one would pay to receive that flow. For example, the flow of rent becomes capitalized into the land value. The flow of funds, whether rent or interest or dividends, is divided by the real interest rate to get the value of the stock, or the price of the asset. This assumes that the annual flow is constant. The tax rate on the value of the stock (e.g. price of land) is added to the real interest rate, so that, for example, $p = r/(i + t)$, price equals rent (or other yield) divided by the sum of the interest and tax rates.

The tax rate t can be converted to the tax rate on r as follows: the amount of tax is t*p, which is divided by r. Since p = r/(i + t), we get t*(r/(i + t))/r. The r cancels out, and we are left with t/(i + t), the tax rate on r. For example, if the tax on land is 20% of the price p and the real interest rate is 5%, then the tax rate on rent is .20 / (.05 + .20) = .20/.25 or 4/5 (80%) of rent.

Capitalization amounts to calculating the present value of a flow of income. At an interest rate of 5%, $100 invested today will in one year be worth $105, or 100*1.05. Equivalently, the present value of $105 is $105/(1.05) = $100. Two years from now, $100 will be worth 100*(1.05)*(1.05), so to calculate the present value, we divide by (1.05)2. In general, the present value of a stream of income is the sum of the incomes for each year, divided by the quantity one plus the interest rate, raised to the power of the number of years into the future.

3. Interest Rates, Investment, and Factors

Suppose you own a firm and have different possibilities for investment. Naturally, you choose those projects with the greatest expected rate of return. But alas, you have no money for investment. No problem, says your financial adviser; we can issue bonds at 8%. The firm will then invest in those projects whose return is greater than 8%.

Interest is usually in the form of money, and it is normally paid on financial capital such as savings accounts in banks, commercial paper (short term borrowing by firms), and bonds. Although the form is money, the substance is goods. By depositing money, you abstain from buying the goods that the money could have bought; when you get money interest, it is a claim on the current stock of goods.

A capital good that does not depreciate can be loaned out indefinitely. The good yields an interest, since funds equal in market value to that good could have been invested as financial capital. But it is an error to call all returns to capital goods interest. As noted in Chapter 4, capital goods depreciate, and the depreciation is a payment for the factors that produced the good, rather than interest.

Since the three factors of production are land, labor, and capital goods, and their returns are rent, wages, and a capital yield, which factor does interest belong to? Interest can be a return to any factor, depending

on who does the borrowing. If someone borrows money and buys land, paying the interest on the loan from the rent of the land, then the interest received by the lender is actually rent from that land; hence, some of the interest earned by money in savings accounts which is loaned to landowners is rent. In effect, the lender of the money is the recipient of the rent rather than the nominal owner.

Similarly, if a worker borrows money for his education and pays the interest entirely from his wages, then the return to the lender consists of wages; some of the wages are earned by the worker and are in effect earned by the lender in return for investing in labor improvement, enabling the worker to have improved his skills sooner rather than later. Finally, if the borrowed funds are used to buy capital goods, the return to the lender is part of the yield on those capital goods.

If the loan is for consumption, then the interest paid by the consumer comes from his income from wages, rent, and capital yields, so the interest constitutes those returns earned in part by the lender. For example, if a landowner borrows money for a vacation and then pays it back from his rental income, the recipient of the interest income is getting some of that rent, because that is where the income originated. If all land rent is taxed, then a landowner cannot pay interest on loans from rent, and interest must then come from wages and capital yields. Investment in either better labor or more capital goods would be needed to yield interest. Loans for consumption then reduce the net returns from labor and capital goods of the borrowers, since some of the returns go to the lender. In effect, borrowing to consume now rather than tomorrow reduces your future net returns from your factors, reducing one's future consumption. Likewise, lending to consume tomorrow rather than today increases your future income and consumption.

4. Interest and Money, Usury and Illusions

Is a high interest rate exploitative? "Usury" is the name for an exploitative rate of return on loans. Sometimes what is seen as a high rate of "interest" for loans is really a risk premium. But exploitation can occur when there is some monopoly leaving borrowers desperate for goods. There can be restrictions on credit and banking, so that the supply of loans is artificially

reduced. There can also be laws making it difficult for some people to borrow money; for example, too-liberal bankruptcy laws make personal loans riskier, driving up the risk premium. These exploitative rates on loans are often premiums on monopoly and artificial risks, rather than pure interest due to time preference alone. When pure interest is due to short time preferences, a high desire to consume today rather than tomorrow, it is not exploitative, so it cannot be properly called "usury."

Some people wish to abolish interest because they think it is caused by a money monopoly, since it is an amount paid for the use of money. But, as we have seen, the origin of interest has nothing to do with money, but with time preference. The same interest could also be paid as goods. Money is the medium of loans and borrowings, but goods are the substance.

Interest rates can be analyzed using supply and demand. "Loanable funds" consists of money available to lend. The supply comes from savings and the creation of money. The demand comes from borrowers. The supply curve slopes up, and there is more savings with higher interest rates, and the demand curve slopes down, as more is borrowed with lower rates of interest.

In the short run, interest rates can be affected by changes in the money supply. Suppose the government increases the supply of money. The banks then have more reserves (funds) to loan out. It is as though savings had increased; the supply of loanable funds has gone up. An increase in voluntary savings means time preferences have shifted towards less consumption today and more tomorrow. An increase in the money supply and thus of loanable funds makes interest rates go down, just as they do when savings go up. But real time preferences have not changed. So the added investment is not economically warranted. Too many capital goods, especially of longer duration, will be produced, such as shopping centers and office buildings that stay half empty. The extra money pushes prices up and so the money supply relative to prices goes down to where it was before. Interest rates then go back up to their natural level. So changes in the money supply can make interest rates change in the short run, but not in the long run. In the long run, the natural rate of interest, caused by time preference, will prevail, since an increase in the money supply just raises prices.

Pure interest cannot be abolished, just as rent cannot be. If the government prohibits the payment of interest, then a borrower is in reality receiving the interest—the benefit of present-day use of resources—if he does not need to pay interest, just as tenants of land receive the land rent if they do not need to pay it to others. If banks share in the profits from enterprises that borrow funds rather than directly charging interest (a practice in Islamic banking), then this bank share is implicitly interest whether one wishes to label it so or not.

Another illusion is the notion that interest creates inflation because more money is needed to pay back a loan, that if you borrow $100 and pay back $110, we need $10 more in cash, creating inflation. But if the loan is for current consumption, the interest is paid by the borrower's reduced future consumption; he must in the future consume less in order to also pay back the loan. No new money is needed. One person is paying interest and another receiving it; so the net demand for cash is the same. If the loan is for investment, the interest is normally paid from the increase in productivity and output.

We see, then, that pure interest does not exploit any factor of production. What does cause economic problems is the distortion of natural interest rates and the increase of premiums by government interventions, whether directly, in the form of interest-rate controls, or indirectly by money creation, banking monopolies, the government's money monopoly, and restrictions on credit, which force desperate people to borrow at high rates in order to survive. Poor people with bad credit, for example, may have to borrow money at high rates from loan sharks; this is not pure interest but a premium for high risk and money and credit monopolies (the restriction of credit due to loose bankruptcy laws and financial restrictions). Also, small businesses are often forced to borrow funds because regulations make it too expensive for them to issue shares of stock.

In a pure market economy, with no restrictions on honest and peaceful money, banking, or lending, interest rates simply ration goods over time between future and present-day uses, enabling those who most urgently want to consume or produce today to do so at a price that reflects preferences over time.

CHAPTER 6

Consumption, Demand and Supply

1. What Happens to Income?

What happens to income? It can either be used for consumption, to buy goods and services, or the income can be saved. Savings are usually put in a bank or other financial institution, which loans it out to borrowers who spend. One can directly invest one's income for capital goods or for human capital. If your investment is direct, such as buying a car, you are getting an implicit return. If you instead had to borrow a car, you would be paying the owner a periodic fee; so if you own your own car, you are paying the yield to yourself, just like if you own land, it has a rent regardless of whether any explicit or money payments are being made.

Another way of spending income is to waste it. Waste involves the use of resources in different ways than the owners would have wanted, and without benefit to them. If a vandal destroys your car, for example, this is not only your own loss, but a social loss as well, since this resource has been wasted. If the government builds a road in a forest that benefits only a few individuals (say, the owners of a lumber company), then the wealth of the original owners has been wasted, used in ways they would not desire and do not benefit from. This road is not a true investment, since its yield is less than what investors would get in a free market. Government subsidies are usually wasteful, and costs imposed by government restrictions and taxes are also a social waste of resources, since they not only extract funds that people would want to spend in different ways, the interventions can also make resources such as labor idle, i.e. unemployed.

2. Consumers' Utility

What, then, determines how income is consumed? Let's go back to the foundational premises of economics in Chapter 1. Proposition #8 states that human beings have ends, or goals, desires, and needs. Proposition #11 states that human values, and thus these ends, are subjective. Proposition #9 states that people are able to rank their ends into those of greater and lesser importance.

The theory of economic values was pioneered by the founder of the Austrian school of economic thought, Carl Menger (1871). He defined value as "the importance that individual goods or quantities of goods attain for us because we are conscious of being dependent on command of them for the satisfaction of our needs" (p. 115). Values are thus subjective, since each individual perceives the importance of goods from his own needs or desires, from his own feelings and thoughts. Value originates in the human mind rather than in things: "Value is thus nothing inherent in goods" (p. 120). We know from experience that what pleases one person will be detested by another person.

Goods have utility, or usefulness and importance, because we value them. As Menger stated it, "Utility is the capacity of a thing to serve for the satisfaction of human needs" (p. 119). In ranking our subjective ends or desires, we also rank the desirability or utility of the various goods that have the capacity to satisfy these desires.

Menger then notes that "these differences in the importance of different satisfactions can be observed not only with the satisfaction of needs of different kinds but also with the more or less complete satisfaction of one and the same need" (pp. 123–4). For example, you not only desire food, but a different amounts for different reasons, such as for survival, health, or enjoyment. The highest rank might be for survival, then for health, and then for enjoyment, or perhaps some folks would rank enjoyment greater than health, as we often do some of the time. Hence, we rank food-value in terms of the decreasing importance of various amounts. In general, for any good, different amounts satisfy different types of needs or desires of different importance, so that as we consume ever more of it, ever decreasing desires are satisfied, until the amount is reached when, at some moment, "a more complete satisfaction of that particular need is

a matter of indifference" (p. 125). Even more consumption would then become a burden.

So the utility derived from some good diminishes as you obtain more and more of it. Maybe the first small increments give increasing utility, but eventually the utility of increasing amounts diminishes. This is called "diminishing marginal utility."

The marginal utility, the value you place on an extra amount of a good, depends on the subjective importance or value of only that extra amount, regardless of the value of the previous amounts. So what you would be willing to pay for that amount depends only on that marginal or extra utility obtained, not on the utility of the previous amounts. And here is the kicker: when you buy a certain amount of a good, if all the units of the goods (like individual oranges in a bag) are priced identically, then the price you are willing to pay per unit depends on the utility of the last, the marginal, the least important unit. Otherwise, you would not buy that extra or last unit!

For example, if apples and oranges are each priced at 20 cents each, but you prefer oranges to apples, you would rather buy one orange than one apple. You value the second orange (at that moment) less than the first, but still greater than the apple. But now, having two oranges, you would rather have one apple than a third orange. If oranges are priced cheaply, you'll then take a fourth or fifth one. You don't want an extra apple because an extra orange is more important at the time. But if that day oranges are expensive, then you might only take two, plus one apple, since the first couple of oranges are more important and valued than the next few. So the price you are willing to pay for one orange or one apple depends on the value to you of that last apple or orange taken, rather than the first.

As Menger put it, "The value of a particular good . . . is thus for [an individual] equal to the importance of the least important of the satisfactions assured by the whole quantity available" (p. 139). Marginal utility explains why water is so cheap where it is in large supply, even though it is an important commodity overall. When lots of water is available, the last unit you buy will be put to a relatively unimportant use, like washing your car. The marginal value of water, then, is low, even though some of the water (the first amounts you would buy if it were scarcer) is very important.

The consumer optimum is the amount of goods for which the marginal utility of each good relative to the price is equal for all goods. If the marginal utility of one good, relative to the price, is higher than for other goods, the consumer is better off buying more of that good, until its marginal utility has diminished so that its MU/P is equal to that of all other goods. In that case, the last dollar or pound spent on all goods give the same amount of marginal utility, and you can't do any better.

3. The Law of Demand

To "demand" in economics does not mean to boldly and angrily insist on getting something, but simply to both want something and be able to pay for it. A demand for a product means the quantities that consumers are willing and able to buy at various prices. Because of diminishing marginal utility, people buy more of a good if the price is less, since they will be willing to satisfy less important desires if the price is lower.

The price of a good is not just the amount one has to pay for something, but also the foregone opportunities. If you spend $100 for a coat, you have lost the opportunity to spend that amount on something else. The real cost is therefore the foregone opportunities, not the $100, since that $100 will be spent on something. At a lower price, the foregone opportunities are fewer or less valuable. This concept of cost being a foregone opportunity is called "opportunity cost," another concept pioneered by the Austrian school, namely by Friedrich von Wieser.

Land has an opportunity cost for an individual, but for an economy as a whole, the land is already there, so there is no opportunity cost in its use overall (as opposed to use for a particular purpose). Labor does have an opportunity cost because the alternative is leisure. Thus, for any good, at different prices there are different opportunity costs, and the lower the price, the lower these costs, relative to the subjective value of another unit of a good. Hence, there is a relationship between different prices of a good and the amount one is willing to buy: the lower the price, the more one is willing to forego other opportunities and buy another unit of the good.

When the price of a good falls, the quantity purchased can increase for two reasons. One is the substitution effect, in which people substitute the cheaper goods for the now more expensive other goods. The second

reason is the income effect, since lower prices provide a greater purchasing power, so people buy more stuff.

A demand function consists of the various quantities of a good demanded for each price of a good. We can mathematically describe this function as $Q = f(P)$; quantity bought is a function of price. The actual relationship for various values of the function can be written in a schedule or table, or plotted in a graph. Such a graph has price along the vertical axis, the opposite from the normal mathematical convention of having the independent variable on the horizontal, thanks to Alfred Marshall (1842–1924), the economist who pioneered the analysis of supply and demand. Quantity is then on the horizontal axis, and we then draw points showing the quantities demanded for each of various possible prices. These points are then jointed together to form a "demand curve." You can search for "demand curve" in the Internet for graphic examples.

Since a greater quantity is demanded at a lower price, a demand curve slopes down, diagonally to the right. This is called the "law of demand," a fundamental principle of economics. As we have seen, the law of demand is derived from the diminishing marginal utility for increasing amounts of consumer goods. Demand curves can also be horizontal (where the price is always the same) or vertical (where the quantity is fixed regardless of price, such as for urgently needed medicine). The law of demand states that the quantity demanded does not increase with higher prices.

The quantity of goods demanded can be of two types. For flows, goods in continuous production and consumption, the quantity is an amount per unit of time. For example, the demand for oranges is a certain amount per week or month. The quantity axis is then Q/T, quantity per time interval.

The other type of good is a stock, a certain quantity at one moment of time. For example, the demand for a certain rare stamp is that of a stock. No more of these stamps are being produced; there is no flow of product. Instead, collectors bid already-existing stamps from one another, competing for a fixed stock. At any moment in time, there will be bids and offers for stamps, and an exchange takes place when a bid and offer match in price. The demand for the stamps are therefore the number of bids at any particular moment in time at some price. At a higher price, fewer collectors are willing to bid for a stamp. So the demand curve for stocks of goods also slopes down, but the quantity axis is for bids at a

certain moment. The market for land is such a market, since people bid for existing parcels, buying from a previous owner rather than for newly manufactured or imported land.

The demand for consumer goods depends of many variables, such as subjective preferences (or tastes), income and wealth, and alternative goods and prices. The demand curve for a good for an entire economy is simply the addition of all the demand curves of the individuals.

A "change in demand" is a shift of the entire demand curve, or a change of the quantities at all the prices. If there is only a change in price, we move along the demand curve or schedule to the quantity demanded at that price. When an economist talks about an "increase in demand," the meaning is that there has been a shift in the demand curve towards greater quantities at all prices.

4. Elasticity or Responsiveness

An entrepreneur wishes to know how much more of his product will be bought if he lowers or increases the price. What he wants to know is called the "price elasticity of demand." We can blame Alfred Marshall for this confusing terminology, just as he is to blame for switching the demand-curve axes. A better term for elasticity would be "responsiveness." The price elasticity is the responsiveness of the quantity demanded to a small change in the price of a good. It is measured as the proportionate (or percentage) change in quantity divided by the proportionate change in price. The technical calculation is as follows: (change in quantity/ quantity)/ (change in price/price). By convention economists often use the absolute value of the ratio.

Demand is said to be elastic if the price elasticity (quantity responsiveness) is high. This means a small change in price leads to a relatively large change in quantity demanded. Demand is called elastic if its value is greater than one, which means the proportionate change in quantity is greater than in price.

The opposite is an inelastic demand, if the quantity responsiveness is low, less than 1. If it is exactly 1, then it has "unitary elasticity" or is "unit elastic." These elasticities are important in applied economics, since enterprises want to know the effects of price changes, and they experiment to see what the quantity responsiveness is.

There are two other elasticities relevant to demand. One is (again confusingly called) the "income elasticity of demand," which would be better called the "quantity responsiveness to income." Here, we measure the change in quantity demanded when there is a change in income. This is calculated as the percentage change in quantity divided by a percentage change in income—this time keeping the positive or negative sign. If we expect incomes to rise, an enterprise wants to know how the demand curve for his product will shift.

A third type is the "cross elasticity of demand," which means, in English, the responsiveness of the quantity demanded of one good when the price of another good changes. It is calculated as the percentage change in quantity for one good divided by the percentage change in price for another good. This cross-price elasticity is important in being able to tell which goods are substitutes and which are complements.

A good is a substitute for another if one can switch from one to the other without much loss in utility. For example, if you don't much care whether your pencil is colored yellow or blue on the outside, they are close substitutes. We can measure this with the cross-price elasticity. If it is positive, the goods are substitutes, because it means an increase in the price of one good (blue pencils) induce an increase in quantity demanded in another good (yellow pencils). A big increase in the price of green apples would likely lead to people buying less of them and more of its substitute, red apples.

If the cross price elasticity is negative, the goods are complements. A complement is a good that is used together with another one. For example, your left shoe is a complement to your right one. Suppose they were sold separately, and the price of right shoes went way up. You would not only buy fewer right shoes, but also fewer left ones, since you want both together. The cross price elasticity would be negative, since an increase in the price of left shoes leads to a decrease in quantity bought of right shoes. If the cross price elasticity is zero, then there is no relationship between the two goods.

5. Consumer Demand and the Factors of Production

In our model in Chapter 3 on rent, we had only one crop, corn, and the rent and wages were calculated in terms of corn. But in a complex economy with many goods, which goods determine the value of wages and rent?

We again turn to Menger. We have seen that the values of consumer goods are subjective. As you recall from Chapter 4, capital goods are called by Menger "goods of higher order" than consumer goods. As Menger put it, "the value of goods of higher order is always and without exception determined by the prospective value of the goods of lower order in whose production they serve" (p. 150). In other words, if consumers value corn highly relative to other goods, then the goods of higher order used to produce those of lower order will have value. Land and labor used in growing corn have value because corn has value. The value of corn is not due to the costs of growing it; on the contrary, the resources used in growing it have value because the product has value. If you worked all day making mud pies, your labor would have a value of nil, if no one would want that product (unless your mother bought some).

As Menger recognized, the determination of the factors land and labor from the expected value of their final products is consistent with the law of rent, as analyzed in Chapter 3: "The existence of the special characteristics that land and the services of land . . . exhibit is by no means denied. In any country, land is usually available only in quantities that cannot be easily increased; it is fixed as to situation; and it has an extraordinary variety of grades" (p. 169). Menger's theory deepens our insight into why land and labor have value at all. It was because the corn had value that the differential rent appeared in the first place and that the wages paid in corn had any value. But it is not the value of corn alone that would determine the rent of a particular lot of land.

As Menger wrote (p. 169), a factor of production will also have a greater value if complementary factors have a smaller value. Suppose that we need irrigation in order to use a particular area of land. If this water is cheap, then land rent will be higher than if the irrigation is more expensive.

6. Cost-Influenced Choice

Building upon the theory of subjective values, James Buchanan (1969), a pioneer of public choice theory which uses economics to analyze government, also developed the theory of choice. If you are choosing among goods with your own money, your choice will be "cost influenced," since

each choice has an opportunity cost of foregone alternatives. Recall in the beginning of the chapter that we had three categories of spending: consumption, investment, and waste. Consumption is a cost-influenced choice. Some government spending is equivalent to cost-influenced consumption and investment, since the taxpayers would, if given a direct choice over their funds, want to buy some of the things government buys, such as perhaps highways and wildlife conservation. But some of government spending will not have been so chosen.

An individual may choose to buy an item he later regrets having bought. In retrospect, it was a waste. But since he did not know this beforehand, the spending is wasted. The key is whether the choice is cost-influenced. If at the moment of choice, you bear the cost and make a free choice, then the item bought is in the category of consumption or investment. The relevant utility of an item is that which is made at the very moment of choice. The fact that you might later have buyer's remorse is irrelevant so far as the consumption is concerned. It's too late. The spending is now a sunk cost. Economics looks forwards to the future. Prices are always based on the expected utilities, not on the actual utility once a choice has been made. The past is important in influencing the future. A bad product will not be bought much in the future, if the word gets around. The past guides our future, but our expectations about the future determine our utilities and therefore the prices of goods.

CHAPTER 7

Making Products and Profits

1. The Three Sectors of An Economy

Having analyzed consumption, we turn now to production. We already have seen how production uses three factors, land, labor, and capital goods. We can now examine the second dimension of an economy, its three sectors: household, firms, and government. The third dimension was described in Chapter 5, the categories of expenditure as consumption and investment.

Households consist of individuals, families, and other living-together arrangements, such as room mates and communes. Firms are the organizations which engage in the production of wealth. Government is the agency that has authority and power over the rules that firms and households are legally required to abide by in their consumption and production.

All wealth is owned by households. They either own firms personally or own shares in corporations. Households who are citizens are the ultimate owners of the land of a country and of its government's wealth. As owners, households rent their factors to firms, hiring themselves to firms as workers, loaning firms their capital goods, and renting land to firms. Firms may, of course, nominally (in name) own resources, but the ultimate owners are households.

Households obtain an income from loaning the factors to firms, and they use this income to invest in capital goods (for firms or government) and for consumption. Households are thus the sector that consumes the wealth produced by the firms.

Firms can be generally divided into four types: 1) families; 2) for-profit firms; 3) non-profit organizations; 4) government enterprises. Non-profit organizations include churches and foundations. The motivation of

for-profit firms is normally to maximize profits, but the ultimate goal is the maximization of utility, which can also involve achieving a large size of firm or getting prestige.

For-profit firms in turn have various possible forms of organization. The simplest is a single proprietorship, where an individual owns the firm. Next in complexity is a partnership, owned by several persons. A family is one possible partnership; it engages in household production, such as gardening. A marriage is also a partnership.

Most complex are corporations, firms whose owners have shares of stock and elect a board of directors to operate the firm. Typically, the board hires a president or manager. A corporation may take the form of a cooperative, in which a shareholder is also a member, and each member has one vote regardless of shares. Another form is a non-stock corporation, such as condominium housing, in which the owners are members and have various types of voting rights, but where ownership consists of having title to units, like apartments, which are tied into the membership of the firm.

In many countries, corporations have legally limited liability, which means that the ordinary shareholders may not be liable for debts of the company beyond the value of their shares. The board, though, is usually liable for the debt beyond their shares in the company. A disadvantage of corporations in the U.S. is that corporate income is taxed twice, once when the corporation earns it, and secondly when it distributes dividends to the shareholders, which is taxed again as personal income.

An advantage of corporations with shares of stock is that the shares trade on active exchange markets, so an investment in the corporation is highly "liquid," and also divisible. This enables a corporation to raise large amounts of money by issuing stock.

A problem with corporations is that the management may seek to promote its own well-being rather than that of the share holders. This is resolved by means of profit-sharing methods of payment as well as managers competing for their positions and the possibility of a takeover if a firm becomes too fat with management benefits.

Besides stocks and units of ownership, companies can also issue debt in the form of bonds. These pay interest for a certain period of time, after which the bond "matures" and the company takes them back and pays back the principal.

The agents who organize the factors of production are called the entrepreneurs. Often they are owners of firms or executives with a great deal of control, but sometimes they can be sales persons or anyone that is able to organize production. Entrepreneurs are the drivers of a market economy; they actively seek out opportunities to bring resources to a more productive use, such as by developing new products or new methods, or bringing goods to markets they were not previously available in. If you have a vegetable garden in your back yard, you too are an entrepreneur, since you decide what to grow, how to grow it, and who gets the produce.

Government plays three different roles. First, some of its agencies act as firms. Governments run enterprises such as the post office, railroads, and street maintenance. In this role, government is a firm hiring factors and producing wealth. So part of government is also in the category of firms. Governments also issue bonds, but not shares of stock.

Secondly, government enacts rules that must be followed by the three sectors (including itself). These rules are of two types: 1) rules creating markets, and 2) rules that intervene or interfere in markets. The distinction follows the universal ethic developed in Chapter 1. Rules which coerce households and firms, which impose costs or restrictions, i.e. any rule other than prohibiting and penalizing coercive harm to others, are interventions. Subsidies are also interventions. It is not claimed here that interventions are bad; those who advocate them claim that the benefits exceed the cost.

Other rules create and maintain markets. For example, uniform and constant rules protecting property rights and enforcing contracts enable markets to operate efficiently. Laws setting up copyrights make a market in literature and art more efficient, preventing publishers from having to write a contract with each buyer regarding copying of the books. Laws setting up patents also simplify transaction costs, also preventing firms from having to write complex contracts with each user of the firm to avoid copying, and making the enforcement easier. Some people feel that copyrights and patents create monopolies and are thus inefficient, while others are of the view that patents and copyrights are general contracts between the sellers and buyers of new products, without which the new products might not be created or written.

The third function of government is redistribution. Governments typically take much of the wealth from the first and second distributions

(see Chapter 5) and forcefully redistributes it to others. Hence, much consumption is done by recipients of redistributed income who have not earned it by supplying factors to firms.

There is another agent in the economy, the thief. A thief also performs forced redistribution, taking wealth and income by force. Their redistribution is not legally sanctioned; they impose their own independent rules and redistribution.

There are two circular flows in the economy among the sectors, for goods and money. Goods, produced by firms (including government firms), flow to households and to government. In turn, households supply factor resources to firms and government.

The second flow, money, circulates in the opposite direction, since a consumer obtains goods in exchange for money. Firms pay money to households and to government for the use of factors. Households and government then use the money to pay the firms (including government) for goods, or lose some money to thieves, or pay taxes to government.

2. The Production Possibilities Curve

When we disaggregate wealth into various products, the question arises as to how much of each is to be produced. We begin the analysis with a simple economy with only two products, bread and lettuce. The economy could produce only bread or only lettuce or some of each. Suppose that only bread is being produced, for a total of 100 loaves. Now we want to produce ten heads of lettuce. If the economy was fully employed in making the 100 loaves, some resources must be taken away from bread making to make the lettuce.

To analyze the economics of the trade-off between the two products, we again begin with the foundational principles. Proposition #1 states that some natural resources are scarce. So there is a finite amount of land and labor available in an economy, which is why only so much bread or lettuce can be produced, and not enough to satisfy everyone if they were free. Proposition #2 states that resources vary in quality, and proposition #4 tells us that different amounts of inputs will produce different amounts of outputs. Applying this, we see that some land is more productively used to grow wheat for bread, while other land is

more productively used to grow lettuce; and likewise some labor has been trained to grow wheat and bake bread while other labor has been educated in growing lettuce. Then capital goods, of course, have been made specializing in one or the other.

When we use a bit less land, labor, and capital goods to produce bread and use it to make lettuce, how do we do this? Proposition #12 states that people economize. So we economize by giving up that wheat-growing land that is most suitable for growing lettuce, as well as labor, etc., that was trained to grow nice lettuce. Likewise, if we had been growing only lettuce, the most productive resources would be used to grow the first increment of wheat and to bake bread. The next most productive resources will be used to produce the next increments of the products.

If we then plot all combinations of bread and lettuce that can be produced, we get a curve. We have bread on one axis and lettuce on the other axis, and for each amount of bread, there is a certain maximum amount of lettuce that can be grown. This is the "production possibilities curve," PPC.

This curve demonstrates several principles. First is the principle of efficiency. Production is efficient if the total product is on the PPC. If the total output lies within the PPC, then one or more of the products can be increased without decreasing the other; hence, society is not producing efficiently. Efficient production means that the production of any product cannot be increased without reducing the amount of any other product. The existence of waste, of course, implies that society is inside the PPC relative to the desires of those obtaining the income of the first two distributions.

A second principle is opportunity cost. To produce more bread, we must produce less lettuce; the opportunity cost of producing more bread is less lettuce.

The third principle is the "law of increasing cost." Economizing persons use the most productive resources first. Increased amounts of the good will require less productive resources, resources which may be used more productively in other uses. So the law of increasing cost states that as we increase the production of one good, the opportunity cost of foregone production of other goods tends to increase. Therefore, the shape of

the PPC is "bowed out," like a rainbow. Economists say it is "concave to the origin," but you can just think of it as bowed out unless you want to impress your friends. The curve is bowed because at first, giving up a little wheat gets you a lot of that first amount of lettuce, and vice versa.

The effect of better technology is to push out the whole PPC outwards, so that the same amounts of inputs yields a greater amount of output. The effect of accumulating more capital goods is also to push out the PPC, since land and labor become more productive. There is therefore a trade-off between consuming now and consuming tomorrow. We can consume less today and invest in new capital goods in order to consume more tomorrow.

3. The Production Function

Since a firm, in abstract, is an organization inputting resources and outputting products, we can describe it as a production function, or a product as a function of inputs. A function is a relationship between a dependent variable and some independent variables. The inputs are the independent variables, and output is dependent. The relationship can be concisely stated as $Q/T = f(N,L,K)$, where Q/T is output per time interval, N is the number and quality of workers, L is the amount and quality of land, and K represents capital goods. Note that these are all physical inputs and output; there is no financial capital such as money in the function.

The methods of production, the technology and rules (including government regulations, the role of luck, and the goals of the firm owners) are included in the functional variable f.

We can see then that if the amount of a resource such as N, labor, is varied, output will vary. The marginal product of labor is in fact the change in output Q/T caused by a change in labor, N.

In order to maximize profits, costs must be minimized, and this implies that the amount of each input will be determined by its marginal product, proportional to its cost. In the cost-minimizing combination of inputs, the marginal productivity of a dollar's worth of all inputs must be the same. If the marginal product of one input, divided by price, is lower than that of another, then costs can be reduced by switching, if possible, to the inputs with the higher relative productivity.

4. The Theory of Exchange

Some people have the idea that agriculture or manufacturing is "productive," but trading, buying goods in one place and selling in another, or exchanging one good for another among two persons, is not productive, but just moves things around. But Menger showed that this is not so.

Menger showed how in an exchange of goods, the goods have unequal rather than equal subjective value. They may have an equal market value, but the subjective values must be different, otherwise the trade would not take place. Trade only takes place if person A has some good that is of less value to him than some good that B has.

Using Menger's (1871, p. 183) example, suppose A has horses and B has cows. Because of diminishing marginal utility, each extra horse has less and less value to A, and so with cows for B. Suppose the first cow or horse has a value of 50 to A and B, and that each extra one has a value of 10 less than the previous. Then if A has 5 horses, the fifth is only worth 10 to him. But it would be worth 50 to B. Same with cows. So they trade. A now has 4 horses and 1 cow. He lost 10 of value by giving up the horse, and gained 50 by getting the cow, for a net gain of 40. Likewise, B has a net gain of 40 from getting a horse.

As Menger states (p. 184), "each of the two traders obtained an economic gain from this first exchange equivalent to the gain that would accrue to him if his wealth had been increased by a good whose value to him is equal to 40 . . . Trade is therefore no less productive than industrial or agricultural activity." Economic exchange contributes to consumers' utility and thus an increase in the subjective value of their wealth just as effectively as the physical increase of more goods. As Menger stated, "the end of economy is not the physical augmentation of goods but always the fullest satisfaction of human needs" (p. 190).

Both will continue to exchange as long as the marginal utility of the other's good is greater than that of the goods they have. The next horse or cow has a marginal utility of 20 to the owner and 40 to the other, so they exchange, each increasing utility by 20. After that, the marginal utilities are 30 for both, so they stop trading. Note that once the goods are of equal marginal value, trade comes to a halt. Trade went on because of unequal rather than equal subjective marginal values.

An important principle of exchange is that trade will continue until the economic gains, the gains from trade, are exhausted, until the marginal values of what one has is equal or greater than what others have.

5. Supply Curves

We derived demand curves in Chapter 5, and now we will derive supply curves, the supply of goods offered by firms.

A market supply is the quantities of goods that a producer is willing to produce and sell at particular prices. Like demand, the quantity is a function of (dependent on) price, but the quantity axis is horizontal and the price is on the vertical axis, by the convention set by Alfred Marshall.

Like demand schedules or curves, a supply can either be a flow, or quantity produced during some time interval, or a stock, a certain amount of goods at one particular moment in time.

The amount of goods competitive firms are willing to supply at various prices depends on the costs of production. Costs in the short run are either fixed or variable, fixed costs being those which cannot be changed during that time interval. In the long run, all costs can be changed; indeed, the "long run" is defined for any particular firm as that time interval at which all costs are variable.

Average costs are simply total costs divided by the number of units, while marginal cost is the cost of producing one more unit (or tiny amount of product). A firm obtains a maximum profit when total revenues are greater than total costs and its marginal cost just equals its marginal revenue, since any extra unit would cost more than it gets in revenues. If the long-term average cost is greater than the price, then the firm will shut down, unless the owners enjoy taking losses.

While demand curves slope down, supply curves slope up for produced goods, while, as noted in Chapter 3 on land, the supply curve for land as space is vertical, as is the supply curve for goods no longer in production, like rare coins and stamps, or old art.

Supply curves, for goods being produced, slope up diagonally due to the universal propositions about physical resources. Inputs are scarce, and some locations are more productively devoted to one use than another, relative to the values placed on them by consumers. So to increase the

amount of bread, as we saw, land use must be drawn away from the production of lettuce, which may not be as productive for bread as the use devoted to bread. So the relative price of the new inputs is higher. Some locations, such as city centers, generate more output per unit of area than others, so the cost per unit of output is lower. If the output units sell at one price, then a higher price is need to engage the less productive locations. Greater quantities are produced only if the price fetched in the market is higher, and the supply curve slopes up.

Short-run marginal costs increase as the marginal product of variable factors declines, since some of the inputs such as land are fixed during that time. However, over the long run, this tendency of input prices to increase can be offset by a change in f, the technology and method of production. At greater amounts of production, in some industries, more efficient ways of producing can be achieved. For example, it is more expensive to make a few cars than to mass-produce them in huge factories. There can be physical reasons for long-run economies of scale; the volume of facilities such as pipes and buildings increase at a greater rate than their surfaces, so the per-unit volume costs can go down with larger structures.

This is called "economies of scale." Unit or average costs decrease with greater production. So an industry with such economies of scale will have a downward-sloping long-run average cost curve until economies from greater scale are exhausted, after which the firm has constant returns to scale. With an even larger scale, there can be diseconomies or increasing average costs of managing an ever larger firm, as management gets more and more complicated and eats up a bigger portion of costs, so at a very large scale, the long-run average cost would go up again.

It is also possible for the supply curve to be horizontal, if the costs of inputs are the same at all levels of production and there are no economies of scale. This is called a "constant cost" industry.

If the average cost of firm output first decreases and then increases, in a U-shaped curve, then the marginal cost curve must cross it at its lowest point, the marginal costs first pulling down and then pulling up the average.

Just as we distinguish a shift in a demand curve from a movement along a demand curve, we distinguish a shift in the supply curve (a change in the quantity supplied at all prices) from a movement along a supply curve (a change in quantity supplied as the price of a product changes).

A decrease in government regulations or improvement in technology, for example, would shift a supply curve out, making it possible to produce more for any particular price.

6. Price Equilibrium

Now that we have the two sides of the market, demand and supply, we can join the two curves in one graph. As the demand curve slopes down, it will typically intersect the supply curve. The point where the curves intersect determines the market price and quantity at that time.

It is quite possible that the curves will not intersect at all. You offer a poem on the market for only a dollar, but the demand curve starts at 25 cents for one poem and slopes down to 100 copies of the poem demanded if it is free. But your supply curve began at $1 for the first poem; you refuse to sell it for any less, so the curves do not intersect. The quantity exchanged in that case is zero.

If the curves do intersect, then economists call the price and quantity an "equilibrium." An equilibrium at a moment in time is a situation in which the gains from trade have become exhausted. If there were a shortage, gains from trade would be possible as sellers increased the price to buyers willing to pay more to get the goods. If there is a glut or surplus of goods, gains from trade can be made as sellers lower the price to get rid of the stuff. Either would be called a disequilibrium, since gains from trading can still be made. At equilibrium, trade halts, because gains from trade have become exhausted.

But wait a minute! If trade stops, then how can there be a market equilibrium price? Is this a paradox? The answer is that at each moment, markets are moving towards equilibrium, but never actually reach it (or if they do in limited situations, trade stops). People eat; they are in equilibrium and stop eating. Then they get hungry again, in disequilibrium, and go to market. They bid for food, while sellers make offers. The market price attains an equilibrium price as gains from trade are exhausted, but then hungry new buyers in disequilibrium keep coming afterwards, making new bids.

Hence, a market price constantly equilibrates or matches demands and supplies, and does not grind to a halt. This equilibrating clears the

market—sellers are matched by buyers at the current price—but new buyers and sellers are always dynamically making bids and offers. Just because the price is stable does not mean equilibrium (in each moment) has been achieved. It just means that bids and offers are somewhat constant during some time interval. The price is always subject to change if there is some change in the amounts and flows of bids or offers. Of course if we consider a time range greater than a moment, then, looking back in time, we observe that often there is some narrow price range in which exchanges have taken place, and we can consider that to be a equilibrium during that duration.

7. Profit

The term "profit" is used in different ways by accountants and economists. To an accountant, profit is the difference between explicit or money revenues and money costs. An economist subtracts from this profit the implicit costs of a firm. Suppose a farmer owns his own land, which he could rent for $10,000 per year. An accountant says his profit was $50,000 that year. But the farmer could have rented his land out for $10,000; by not doing so, he lost $10,000 in potential revenue. This is an opportunity cost of using the land himself. Since the land rent is a cost regardless of to whom it is paid, an economist subtracts it as an implicit cost, not paid in money but a cost of using that factor nevertheless. So the economic profit is reduced to $40,000. But wait: the farmer's own wages must be subtracted too. If he could have earned $30,000 working for someone else, that too is an implicit cost, a wage. So we have $10,000 left. But what about the capital goods? He could have hired them out for $6,000 that year. Subtract this implicit yield on capital goods, and we are left with a $4,000 economic profit. This is also called an entrepreneurial profit.

But since wealth is divided into a first distribution of wages, rent, and capital yields, where does this economic profit fit in? Since entrepreneurs are also workers, this entrepreneurial profit is really a type of wage.

But it is a special type of wage. Foundational proposition #15 states that the future is uncertain. Entrepreneurs are innovators, but they can't be sure whether they will earn a profit from trying or organize factors in what they expect to be better ways. Uncertainty cannot be insured

against, unlike ordinary risk. Normal risks such as fires occur with some regularity in a larger-number environment, and insurance companies can measure how much loss there has been in an average year and provide insurance against it. But, as economist Frank Knight pointed out, uncertainty does not have probability distributions. New products are unique in time and circumstance, we cannot know what the probability of success is. So entrepreneurs and their fellow investors take a chance, and if they are right, their reward is entrepreneurial or economic profits. If they are wrong, they take losses.

Profits and losses are important signals in a market economy. Consistently high profits in an industry indicate that more resources can be devoted to this product. Losses indicate that too much production has taken place in that industry. Taxes on profits skew these signals, reducing the potential investment and entrepreneurship in an economy, reducing output, efficiency, and employment.

We see then, that profits induce firms to produce, innovate, and employ factors. How this is done by the economy as a whole will be the subject of the next chapter.

CHAPTER 8

Structuring an Organization

1. How Markets Work

In order to understand the concept of a market, it is helpful to look at its evolution.

Trade may have begun with informal agreements among neighbors, such as the basket-weaver offering his handiwork to the farmer next door, in exchange for some of the farmer's fresh vegetables. But, in time, trade became carried out on a more community-wide basis—in particular gathering spots called markets. Today, we use the term "market economy" to refer to a society where people meet their needs through voluntary agreements of exchange. A "market" is not a particular place or group of people, but the process of voluntary production, exchange, distribution, and consumption.

Initially, trade was carried on through barter—the direct exchange of one good for another. However, for large-scale trading, barter is a cumbersome business: any farmer, for example, who wants to swap his vegetables for some tools, needs not only to find someone willing to trade in tools, but to trade tools specifically for vegetables. Trade through barter requires what is called "a double coincidence of wants."

The story of how traders solved this problem has been related by many authors, Menger (1871, p. 258) having analyzed it in detail. As related by Menger, although barter limited the exchanges of traders, "there were elements in their situation that everywhere led men inevitably, without the need for a special agreement or even government compulsion, to a state of affairs in which this difficulty was completely overcome. Traders realized that they could exchange their specialized products for commodities which had a greater marketability, which they could then trade for what they wanted to consume." Cattle, for example, were readily saleable in many areas.

And so, in any particular market area, certain goods, such as cattle, cocoa beans, gold, or wampum, became intermediate goods with exchange value, and the use of such goods for payments became a social custom. Menger emphasizes the importance of custom, since "the actual performance of exchange operations of this kind presupposes a knowledge of their interest on the part of economizing individuals" (p. 261).

Money eliminated the need for the more cumbersome system of bartering, but whether one is trading with money or through barter, the act of voluntary exchange is what creates a market. As defined in Chapter 1, a "market" is the totality of voluntary economic acts in some context. The term market denotes anything from an exchange between two individuals on a street corner, to the more elaborate trading on Wall Street. A market process cannot be perceived as detached from the people who create it.

Classical economists noted that a free market will allocate scarce resources effectively without the need for any central direction. The output produced is generally the amount that people wanted to buy, without severe shortages or surpluses. How is it that production manages to adjust to the continually changing wants of consumers?

Markets are able to allocate resources through the price system. Prices serve as a signal to a consumer about the relative costs of goods, which they compare to their subjective valuations of these goods. Prices also indicate the costs and revenues to producers. Consumers and producers react to prices in their buying and selling decisions, determining the types and quantities of products.

If more of one item—a hammer for instance—is demanded, hammers become sold out. There is a temporary shortage. This will cause the price of a hammer to rise, eliminating the shortage. But then since producers are making a greater profit, this stimulates them to produce more hammers. As more hammers are supplied, the price of hammers will go down again, though perhaps not to the previous level, since it may cost more to draw resources from other uses in order to make more hammers. And so the market ends up with more hammers, perhaps at a bit higher price, equilibrating the desires of consumers with the costs of producers.

Similarly, imagine that too much of a certain commodity, such as coffee, has been supplied. Through competition among sellers to get rid

of their coffee, the price will be pushed down. But at the lower price there is less profit, so producers will reduce their supply, thereby reducing the glut of coffee.

So we can see that through an enormous system of trial and error the price system will ultimately balance supply with demand—fluctuating prices will ensure that a glut or shortage of an item or service does not persist.

Adam Smith, in his classic text *The Wealth of Nations,* praised the workings of the price system. He demonstrated how through the price system, the an individual's pursuit of his own interest contributes to the well-being of others. Thus, from the pursuit of individual interest, society is led, by an invisible hand, to the common good.

Smith showed that "self-interest" in the course of history had led to the specialization and division of labour. The exchange—through trade in markets—which naturally followed specialization, was responsible for the world's progress. As such it should be allowed to progress unhindered by government intervention.

2. Competition

The achievement of community welfare through the pursuit of individual interest presumes a freely competitive market. Competition is rivalry among producers and consumers, bidding against one another for goods or sales. It is only through competition that more producers enter a market when profits are high, increasing supply and reducing the price. As such, the economic society envisaged by Smith was to be devoid of economic privileges, which hinder competition.

Competition has been criticized as being chaotic, but it is in fact an orderly process, a spontaneous rather than planned order that follows ethical rules, namely that of not harming others. Far from being destructive, rivalrous firms competing for scarce resources determine the best use of the resources by their bidding for them. Without this competition, we could not perform economic calculations, because in a complex economy there would be no way of knowing the relative scarcity of resources relative to consumer demand.

If competition is prevented from operating in any way, whether by organised groups, criminals, or legislation, the result is usually higher

prices or a lower quality of goods and services, as well as the social waste of inefficiency.

Competition tends to eliminate profits other than normal returns to the factors of production. The firms that maintain the lowest costs of production will earn a greater share of profits. In an effort to keep costs of production down to a minimum every avenue of innovation will be explored. Indeed, the process of competition is essential to efficiency—it provides for cost saving innovations and induces firms to adapt to change.

One obvious barrier to competition is tariffs and quotas imposed at national borders. Another barrier is a governmental monopoly such as the postal service, or the domination of a service such as education by making the governmental service gratis, while those who prefer private services have to pay the price. Licenses are another barrier; they restrict entry into an industry to certain specified persons. The alleged reason is to assure some standard of competency, but the result is often restricted entry even for qualified practitioners.

3. Market Structure

The market structure of an industry consists of the number of firms and their relative size. Different market structures induce different types of outcomes. Economics categorizes four basic types of market structure.

Atomistic Competition

In economics there are two different meanings of "competition." One is rivalry, in which a greater share for one implies less for others. The second meaning of competition is "the degree of absence of pricing power." Pricing power is the ability to set a price. This is also referred to as market power. When economists refer to competition in market structure, they usually mean the second type, the relative absence of pricing power.

The most competitive structure, the one with the least pricing power, is the structure in which there are many, thousands and millions, of firms, none of which is large enough to affect the price of the product, and in which there are no barriers to the entry and exit of firms into that industry. This structure is called "atomistic" competition. For analysis,

the hypothetical extreme case of an infinite number of firms and costless entry is called "perfect" competition, meaning a complete absence of pricing power. Many products in atomistic competition have basically the same outcome as with perfect competition.

In atomistic competition, a firm is so small relative to the rest of the industry that it must sell its output at the price set by industry supply and demand. The product is homogenous or uniform, so that if one firm tried to sell at a price a bit above the market, no one would buy it, since others are selling the same stuff for less. The firms are price takers, selling at the market price.

If firms in atomistic competition make economic profits (as discussed in Chapter 7), then new firms will enter the industry to obtain some profits. As the industry expands, the industry supply curve shifts out and hence price will be driven down the demand curve. Economic profits are therefore a short-run situation; in atomistic competition, the long-run equilibrium economic profit is zero, and the firms only make normal accounting profits that provide normal returns to the factors of production.

But what if one farm has superior land. Wheat farming may be an industry with atomistic competition, but would not the farms with better land will have more profit? They will have lower average costs, but the better land has a higher rent, so after subtracting out the rent from the accounting profit, the net economic profit is still zero. The "producer surplus" goes to rent.

But wait a minute! Each firm in atomistic competition has no control over the price, yet the industry price can move up and down as industry supply and demand curves shift. So how can industry prices change if no firm can change the price?

Here's how it works. Suppose we have a million wheat farmers. One farmer wants to sell his wheat. He calls a wholesale dealer. The wholesale dealer might be overstocked with wheat, so he calls his broker at the commodity exchange, where the wholesaler can buy or sell all he wants at the quoted price. In the commodity market, such as the Chicago Board of Trade, there are thousands of buy and sell bids being entered every minute. The price is set second-by-second by the brokers who match the bids to buy and offers to sell. No single bidder can dictate a price, but each bidder has a small influence by increasing the bid in one or the other

direction. So the price is set in the commodity exchanges, where the broker adjusts the price to match buys and sells. The same concept applies in financial markets for stocks and bonds.

In atomistic competition, firms produce at the lowest possible cost, which is the industry's minimum average cost. Since profit is maximized at the quantity where marginal revenue equals marginal cost, and marginal revenue equals price, the price equals the marginal cost. This price is socially efficient, because the price, being on the demand curve, reflects the marginal willingness to buy. So the marginal willingness to buy equals the marginal cost, and we cannot get more efficient than that. If price is greater than marginal cost, we are better off producing the next unit. If marginal cost is greater than price, the next unit has a loss. So when price equals marginal cost, we are in economic heaven for that good.

Monopoly

The other extreme of market structure is monopoly. A monopoly exists either when there is only a single seller of a particular good or service, or as a different type of monopoly, when there is no entry into the industry for the expansion of product. We can call the first type a number-monopoly or absolute monopoly, and the second an entry-monopoly. Economists today usually mean absolute monopoly when they use the term "monopoly," although classical economists also referred to entry-monopoly.

If there is only one firm in an industry, the firm's demand curve is that of the industry, so it has the ability to set either the price or quantity of output. Note that in a market economically it cannot do both, since if it sets a price, market demand will determine the output sold at that price. A government monopoly, operating outside a market, can of course dictate both price and quantity, forcing people to consume and pay a price.

A profit-maximizing absolute monopoly will set its price at the level where its marginal cost equals its marginal revenue. Since its demand curve slopes down, its marginal revenue curve slopes down too, and even steeper. This is because each extra unit of output is not only sold at a lower price, but all previous units are also sold at that lower price. The quantity that maximizes profit is that for which the marginal cost and marginal revenue curves intersect. At that quantity, the monopoly price is the point

on the demand curve. The price minus the average total cost at that quantity is the economic profit.

Thus the monopoly can limit the quantity to maintain a higher price than would prevail in a competitive industry. Unless there are offsetting benefits, this reduction in quantity creates a loss to society, called a "deadweight loss."

A monopoly can also practice price discrimination, the practice of charging different prices to different types of users. Firms price discriminate when they have discounts to older people or children. In Eastern Europe, some hotels and expensive restaurants charge a higher price to tourists than to the locals. Price discrimination increases sales by adjusting the price to the elasticity of demand, with the demand of wealthier customers more inelastic, or less responsive to price increases, i.e. they continue to buy even at the higher price. Price discrimination both increases the profit and decreases the deadweight loss, since the firm can sell more at lower prices.

A monopoly having no current competition may face potential competition if its economic profits are consistently high, so it may be induced to keep its price below the short-run profit-maximizing level.

In entry-monopoly, even where there are many firms, they can earn economic profits, since other firms cannot enter to expand the output. An example of entry monopoly is taxi cabs in New York City; one needs a government permit to enter the field, but the number of permits is fixed, so to enter, one needs to buy a permit from a previous owner. Land, being fixed in supply, works the same way.

Types of absolute monopoly include locational monopoly, natural monopoly, new-product monopoly, and government-protected monopoly.

A locational monopoly is the only firm in some immediate market area, such as the only drug store in a small town. Such firms can have monopolistic profits, but these profits are limited by competition from farther-away firms and from the personal relationships that may develop in a small town.

A new-product monopoly is a temporary restriction against competition for creators of new literature (copyrights) and for new inventions (patents). A government-protected monopoly is a legal barrier to entry not warranted by new products.

Patents are a controversial form of monopoly. Some argue that patents are necessary to protect the investment in research of the inventor and

stimulate new inventions; others that they give the inventor too much of a monopolistic privilege. Those who favor patents argue that they are a method of economizing on contracts. It would be costly for the maker of a new firm to make a contract with each buyer not to copy the item, especially to enforce such a contract. Patents, like copyrights, simplify the contract with a notice on the product that it is patented and a conventional number of years that the contract applies for. Patents also enable the inventor to register his invention and check to see that it is really new.

New-product monopolies may charge higher prices than they would in a competitive market structure, but these new products might not be brought to market otherwise, so the net result can be beneficial. This is not so for government-protected monopolies. Here, the public might pay a higher price and gets less output for no good economic reason.

A naturally monopoly occurs when there are economies of large scale. There is a large fixed cost and a small marginal cost, and it is not profitable for other firms to enter. Examples include municipal utilities such as piped water; a second firm would duplicate the pipes of the first, and bringing in water by truck is much more expensive.

One way government has dealt with natural monopolies is to set a price, usually at the average cost (including a normal accounting profit). A problem with this method is that the firm has little incentive to control costs, other than from government oversight and hearings, which may not be effective in controlling costs. The commissioners regulating a monopoly may have come from that industry and may in fact be working to benefit its owners rather than the public.

In many cases, the government runs the industry directly. In some cases, government enterprises such as trains and subways provide good service, though often not, but it is again difficult to control the costs, and the incentive of workers, often in a union, may be to increase their own benefits and power, which increases costs.

A third option is to periodically open control of the natural monopoly to a competitive bid. The industry itself is a monopoly, but bidding to own it for a while can be highly competitive. The highest bidder then runs the outfit for a certain period of time, charging what it pleases. But the government keeps the fee paid by the bidder. If the bidding is competitive, this fee represents the monopoly profit, which is now paid to

the government. The company then has the incentive to minimize costs during its operations. This option is also not perfect, because the firm still charges the public a monopoly price, but it avoids the social waste of artificially high costs that may compensate for that.

A fourth option exists when the firm cannot make a profit only from the user charges, and the service is a territorial collective good that increases land rent. Suppose there is a subway in a city, which is a natural monopoly. Even maximizing profits, the revenues would not cover the costs. But there is a second profit in the increased land value and rent, which when added to the fares would make the operation profitable. The service can therefore be funded by a combination of rent and user charges, especially when the charges are based on the congestion of the service, charging more when it is crowded to compensate society for the crowding the users impose and to even out the usage.

Monopolistic Competition

Whereas with atomistic competition, there is a uniform product, with monopolistic competition, there are many firms, but there is product differentiation: each firm produces a different version of the product, such as a different style, brand, or location. Each firm also has a mild monopoly on its variant, the products being close substitutes. There are also no barriers to the entry of new firms and increased production. An examples of product differentiation is book publishing. There are hundreds of publishers producing similar books.

Because of this mild monopoly, each firm faces a downward-sloping demand curve, and has some control over price. These firms will then set prices where marginal costs equal marginal revenues. But over the long run, because of competition, there will tend to be no economic profits, firms operating where the price equals the average cost. But since this cost curve is tangent to the demand curve (coming down to touch it and then bouncing off), the firms are not operating at the minimum possible average cost.

Although the price is above marginal cost, and above minimum cost, product variety is valued by many people. Firms compete by advertising, which is not wasteful, since ads provide some information, and ads can

act as a signal that the company has a quality product, otherwise it would not have repeat sales. Advertising also pays for newspapers, magazines, radio, television, and web sites.

It is unrealistic to expect markets where products are or can be differentiated to behave like those where they are uniform. The market has in fact produced generic brands as well that have less fancy labels for lower prices. Firms try to influence consumers, but in the end, consumers choices are voluntary. There is nothing imperfect about an outcome that is the best one can have given the conditions of the products.

Oligopoly

An oligopoly, an industry with few sellers, includes firms with a homogenous product such as oil and with a product differentiation such as cars.. There are often few firms in an industry due to economies of scale, which induce firms to become large. A firm in an oligopoly is very much affected by the action of any of its competitors, but exactly how it responds depends on circumstances. A type of game can be played by one firm lowering price and the others responding, and like chess, the oligopoly game has no one exact sequence of plays.

Oligopolists can collude to create an industry monopoly among them, either secretly or openly as a cartel. But there will be a great temptation to cheat on the agreement, since a firm that lowers its price just a bit will be able to sell much more product. If one or more of the firms sell below the cartel price, then eventually that price cannot be maintained, and the oligopoly will fall apart. Also, new firms may enter the industry to take advantage of the cartel price, and as the supply curve shifts outward, the increased product must be sold at a lower price. The cartel must either lower its price or fall apart. There can be an equilibrium in which no firm has any gain in reducing its price, each firm maximizing profits given the market shares of the others; this is referred to as a "Nash equilibrium," after the mathematician John Nash.

An economic "game" consists of interaction among players, rules, and payoffs for each player. A game can be cooperative or rival. Game outcomes can be positive sum, negative sum, or zero sum. A game can be one-time or repeated.

Game theory explains decision making with incomplete information. It involves strategies. "Strategic" means each player takes into account the likely reactions of other players.

A dominant strategy is one that is the best for a player, regardless of the actions of the others.

The most famous game is called the prisoners' dilemma. The story is that there are two thieves who get caught. They are partners in crime, but they don't care about each other. They are questioned in separate rooms. If one confesses and the other does not, the confessor goes free and the other 20 years in prison. If they both confess they get 8 years. If neither confesses, they get a lesser charge, and only 1 year in prison. The incentive is for both to confess. Confessing is the dominant strategy, even though they would both have been best off not confessing.

4. Government and Competition

The degree of oligopoly or industry concentration can be measured using the Herfindahl index. You first calculate the fraction of the industry that each firm has. Then square each fraction. Finally, add up the squares. The result is a concentration index between zero and one, one being a monopoly and a number close to zero being atomistic.

Governments have reacted to oligopolies and collusion with "anti-trust laws," breaking large firms (formerly called "trusts") into small ones. But the success of this policy has been questioned, since large firm size or high concentration does not necessarily imply that the market can be improved by intervention. Some firms become large because they provide superior goods and services—breaking them up would punish market success. Large firms can also capture the benefits of research, and focusing only on industry within a country overlooks the fact that we live in a global economy, and in a market economy, there can be plenty of rivalrous competition. Moreover, government agents do not have the knowledge needed to know just how much competition is optimal, and they are unable to know the unintended consequences of meddling in the market.

While trying to break up some industry oligopolies, governments sometimes deliberately create them with price controls and restrictions on entry. For example, some city governments limit the number of taxi cab

firms and cars. Before 1978, the airline industry in the U.S. was prevented from competing in prices, and the entry of new firms was restricted. The industry therefore engaged in non-price competition, such as offering more luxurious service or more frequent flights on half-empty airplanes. Many consumers prefer lower prices to such high-cost services.

As noted, competition is not just the existence of many firms, but of rivalry among them. In a free market, rivalry for the consumer's marginal dollar or pound will ultimately win out against inefficiencies, since in a global economy any excess profits or costs are like bait to the hungry wolves stalking the woods for profit opportunities. The best that government can normally do is to take down the barriers but prevent the consumer sheep from being fleeced, with stiff laws against fraud, and easy access to the courts, making the loser of a lawsuit pay all legal costs.

Though competition is rivalrous, at the same time, firms have an incentive to cooperate where their interests are mutual. Firms create industry associations to provide them with research, information, and camaraderie. Hence, competition and cooperation are complements rather than opposites in a market economy. It is difficult enough to provide a sound legal basis for market processes without trying to improve outcomes when the cure may well be worse than the alleged disease.

The next chapter will take a closer look at the outcomes of our current economies and analyze them to see whether it is indeed the market or intervention that is the foundational cause.

CHAPTER 9
Social Ramifications

1. Inequality

All large economies have unequal distributions of income. We first examine how to measure inequality, and then inquire as to why it is ubiquitous, existing in all cultures and economic systems, and the morality of equality.

A simplistic way to measure inequality is to count the percentage of the people with the top or bottom x% of income, such as saying that the top 10% own 50%. But to really measure the extent of inequality, we need to take account of the entire distribution, not just one part of it.

One widely-known way to depict inequality is the "Lorenz curve." This is a square with a diagonal drawn from the bottom left to the top right. The horizontal axis measures the cumulative percentage of population, with zero at the left and 100% at the right corners of the square, from the poorest (on the left) to the richest. For example, the point 3/4 of the way from the left indicates the lowest 75% of the population. The vertical axis plots the phenomenon being measured, such as income or wealth, with zero at the bottom and 100% at the top. If the distribution has complete equality, the Lorenz curve coincides with the diagonal line. The more unequal the distribution, the closer the curve to the bottom and right sides of the square. A totally unequal distribution would have one person owning everything and the rest nothing, which would be a curve along the bottom and right sides of the square.

The Lorenz curve can be used to generate a number that measures inequality, called the "Gini coefficient." This is calculated by measuring the area between the diagonal line and the Lorenz curve, and dividing it by the area of the triangle formed by the diagonal and the sides of the square. The greater the ratio, the greater the inequality. The Gini coefficient (G) can also be measured directly from the distribution. Let n be the number

of persons or units in the distribution, y be the average income or wealth of the distribution, and y1 be the highest income (or wealth), y2 the second highest, etc., then $G = 1 + 1/n - (2/(n2 * y)) * (y1 + 2\,y2 + \cdots + nyn)$.

An easier way to calculate inequality is the "inequality index" developed by the author. First we measure the concentration of the distribution. Suppose the distribution is (50, 30, 20) for three persons. The Herfindahl concentration index is calculated by first computing the fraction of the total held by each person. This would be: (5/10, 3/10, 2/10). Then square each fraction: (25/100, 9/100, 4/100). Finally, add up the squares: the total in this case is $(25 + 9 + 4)/100 = 38/100$, or .38. This is the concentration index C. To get the inequality index I, simply multiply C by N: $.38 * 3 = 1.14$. Perfect equality has an index of 1, so the greater the inequality, the greater the index. For example, for three persons, if one had all and the other two had nothing, the index I equals 3.

To understand why inequalities exist, we can divide incomes by the returns to factors, as wages, rent, and capital yields (along with interest income, which originates in any of the three factors). Obviously, if the ownership of land and capital goods is unequal, then this explains a major source of income inequality. But then we can go deeper and inquire as to why assets are unequally owned.

Property is obtained by two methods: voluntarily and by force. Some wealth and a great amount of land has been obtained through conquest, which accounts for many great fortunes, and thus much inequality in places such as Latin America, where estates date back to the Spanish conquest. In some cases, monopoly power has been a source of income which is inherently unequal, and government-provided privileges and subsidies have been a key source of unequal income.

But a great deal of wealth has also been accumulated through voluntary means, by entrepreneurs and their lucky heirs. The profits of entrepreneurs are a return to their labor. Wages are unequal due to differing abilities, education, goals, effort, and just plain luck. Discrimination is also a cause of inequality, although in a prosperous market, it is a minor cause, since there are ample opportunities, including self-employment, and the market penalizes discrimination with higher costs and lower profits. Normally, effort plus talent are get their just rewards in a pure market economy. But when interventions create unemployment, employers can be more choosy, and are then able to discriminate against those they dislike.

Some social scientists have argued against income inequality, saying that a poor person values an extra dollar more than a rich person. But this is not universally so, and such subjective values cannot be measured or compared. Even if this is true for most people, it does not morally justify stealing wealth from the rich, and taxing the rich to benefit the poor would also create disincentives to create wealth, leading to less production and employment.

A sounder argument against inequality is that the rich can give their children a better head start, with better education and other opportunities. In a pure market economy, however, all have an opportunity to obtain an education and employ their labor, even though the wealthy have better opportunities. But again, the universal ethic does not permit stealing out of envy, and economically, such theft would reduce opportunities by reducing enterprise.

Much of the inequality that exists in the world today is not due to better talent or luck, but from privileges and plunder gained through government coercion, especially the taking of land from previous inhabitants. Taxes on wages deepen this cause by reducing the ability of workers to save income and accumulate capital. As Henry George (1883, p. 9) stated, "at the bottom of every social problem we will find a social wrong." When the benefits of land and other natural resources are shared equally, a major cause of inequality is eliminated without resort to the forceful taking of legitimately earned wealth.

Taxes can be characterized by how they relate to inequality. A tax is called "progressive" if the tax rate is higher for higher amounts of the thing being taxed, whether income or property. A "regressive" tax has a higher rate for lower income or property. A flat or proportionate tax has the same rate for all levels of income or property. An income tax may appear to be progressive, with higher rates on higher incomes, but actually not be if high-income earners are able to avoid the tax with deductions and exemptions. Also, a tax or fee may appear to be proportionate, but actually be progressive. For example, if an assessment is paid by land owners, equal to the amount of their land rent, then the tax rate is proportional, with the same rate for all sites, but if the ownership of land is concentrated among the wealthy, then in effect the assessment is progressive.

Taxes are often imposed not just to raise revenue, but to redistribute income to something allegedly more just. But as Henry George (1883, p. 83)

noted, "As to what is the just distribution of wealth there can be no dispute. It is that which gives wealth to him who makes it, and secures wealth to him who saves it." A just distribution is the outcome of a just process. And the universal ethic (described in Chapter 1) prescribes that the just process in the distribution of income is the full retention of wages and capital yields by those who produce wealth, and the sharing of the yield of land by members of a community. Whatever inequalities arise from this process stems ultimately from unequal talent, effort, and luck. Those inheriting past wealth, or talent, or finding luck, may not deserve them more than others, but justice is not about fairness in the fortunes of fate but the deeper equality of each person being able to freely pursue one's own life.

2. Poverty: The Great Enigma of Economics

We began this book with the question of prosperity and social justice, and the realization that our ideals must be founded on sound principles. The first eight chapters have laid out the principles of micro-economics, the theoretical foundation of economics. We will now apply them to analyze the issues of poverty, economic inequality, unemployment, and other social problems.

The greatest problem in economics is the explanation for and solution of poverty. The French economists of the 1700s, whose school of thought is called "Physiocracy" (the rule of natural law), examined the poverty of French peasantry and proposed free trade and a tax only on "net product," which we now recognize as rent. Adam Smith also called for unrestricted, free exchange as the way to create the "wealth of nations." David Ricardo analyzed free trade further and also noted the suitability of land as a source of revenue. Henry George, therefore, did not really break new ground in his policy that called for free trade and a tax only on land rent, but he was foremost in focusing on the problem of poverty in the midst of progress, on just why advancing wealth and technology did not eradicate poverty.

George starts his classic *Progress and Poverty* with a statement of the puzzle:

"At the beginning of this marvelous era it was natural to expect, and it was expected, that laborsaving inventions would lighten the toil and improve the condition of the laborer; that the enormous increase in the

power of producing wealth would make real poverty a thing of the past" (p. 3). Not only could one expect material prosperity for all, but also, in lifting everyone from want and economic anxiety, the moral uplifting of social life:

"And out of these bounteous material conditions he would have seen arising, as necessary sequences, moral conditions realizing the golden age of which mankind have always dreamed.

Youth no longer stunted and starved; age no longer harried by avarice . . . discord turned into harmony!" (pp. 4–5). This is not to say we would have utopia, but that the core of our economic problems would disappear, making other social problems, such as raising our children as sympathetic members of society, much easier to resolve.

Such were the hopes of enlightened people at the dawn of the industrial age. George wrote 100 years after the beginning of the industrial revolution, and it is now another 100 years later. The idea of progress has remained popular, but, as George wrote, "Now, however, we are coming into collision with facts which there can be no mistaking. From all parts of the civilized world come complaints of industrial depression; of labor condemned to involuntary idleness; of capital massed and wasting; of pecuniary distress among businessmen; of want and suffering and anxiety among the working classes" (pp. 5–6).

That these words apply today just as much as they did 100 years ago is a sad testimony to the failure of our institutions to have implemented the remedies for these problems. This in part is due to the economic ignorance of the public, which is exploited by authorities. Armed with economic knowledge, the public will demand that the remedies be applied immediately and no longer be fooled by fallacies and economic quackery.

For George then points out that the existence of similar problems throughout the world cannot come from local and unique causes. These problem occur under many cultures, historical settings, and types of government. "Evidently," he wrote, "beneath all such things as these, we must infer a common cause" (p. 6). Moreover, a study of new territories versus old ones and the history of economic development demonstrate that while people may start out in equality, as an economy develops, as wealth grows, so too does poverty. Social difficulties are "engendered by progress itself . . . The promised land flies before us like the mirage" (p. 8).

George, of course, does not deny that on an absolute scale, the condition of many workers has improved with technical progress. But the lowest classes still lie in the muck of poverty—witness the homeless in the U.S.A. today, and the slums of the inner cities. "Those who are above the point of separation are elevated, but those who are below are crushed down" (p. 9). So, George poses the problem:

"This association of poverty with progress is the great enigma of our times. It is the central fact from which spring industrial, social, and political difficulties that perplex the world, and with which statesmanship and philanthropy and education grapple in vain. From it come the clouds that overhang the future of the most progressive and self-reliant nations. It is the riddle which the Sphinx of Fate puts to our civilization and which not to answer is to be destroyed. So long as all the increased wealth which modern progress brings goes but to build up great fortunes, to increase luxury and make sharper the contrast between the House of Have and the House of Want, progress is not real and cannot be permanent" (p. 10).

As George (1883, p. 81) stated, "For every social wrong there must be a remedy. But the remedy must be nothing less than the abolition of the social wrong."

But instead of eliminating the cause, many political interests have come forth with superficial treatments of the effects. Quack economic medicine can be recognized by noting that it only treats effects, ignoring the causes. Even eminent economists such as J. M. Keynes have misdiagnozed the cause as inherent instability in markets, requiring government intervention, when in fact it is interference in the natural economy that has caused the turbulence in the first place. Quack remedies include the ideas that "labor and capital" are in fundamental conflict, that automation and technical progress is responsible for unemployment, that profits or interest are evils, that more and more money can solve a "lack of demand," and that the government has to pump up investment and demand. Such ideas, recognized George, bring the masses under the power of "charlatans and demagogues" (p. 11).

But economics can tell us the answer, by examining the causes and their consequences. This we have done. We have the philosophical foundation in Chapter 1 and the three resources or factors of production in Chapters 2 through 4. There we saw that the margin of production

determines the wage level, and that the wealth produced after the payment of wages and the return to capital constitutes rent. We saw how as the margin extends to poorer lands, wages decrease while rent increases, and how rent also vastly increases as communities and commerce arises. We saw how land speculation accelerates these trends by using up land faster than warranted by current use, decreasing wages by moving the margin of production to less productive land. We saw how the use of more capital goods increases wages, but increases rent also, and how wages declined back again if the capital goods extend the margin once again to less productive lands. Competition can lead to efficient production, but cannot by itself raise wages which have been driven down to an unproductive margin.

We have seen also how the taxation of wages makes workers poorer and reduces employment, while the taxation of production, trade, and enterprise reduces the amount of production from what would otherwise take place, reducing employment and wealth. Workers at the bottom, or those without work, and thus squeezed in the double pincers of costs imposed by government and, over the long run, ever-increasing rent. The diversion of the rent to the holders of land title is not a part of a pure market, since much intervention consists of taxing labor to provide public works that benefit especially the owners of land, thus transferring income from workers to landowners.

But government restrictions and costs are imposed, in part to treat the effects of social problems, in part due to special interests that lobby and pressure for privileged benefits with the excuse of benefiting the public, and in part due to ignorance about the causes and effects. Social problems thus provide manure for ever more taxes and restrictions. Hence, the original and ultimate cause is as described by George (p. 282): "The reason why, in spite of the increase of productive power, wages constantly tend to a minimum which will give but a bare living [for those at the lowest end of the wage scale], is that, with increase in productive power, rent tends to even greater increase, thus producing a constant tendency to the forcing down of wages."

"As land increases in value, poverty deepens and pauperism appears. In the new settlements, where land is cheap, you will find no beggars, and the inequalities in condition are very slight. In the great cities, where land

is so valuable that it is measured by the foot, you will find the extremes of poverty and of luxury" (p. 288).

Suppose a great island arose in the ocean, on which people could cultivate plants that would give them twice the wage in the old country. What would happen to wages in the old country. As workers leave the old country, rent rapidly decreases as marginal land is abandoned for the new land, and thus wages rapidly rise. With good free land available, rents in the old country must fall and wages must rise.

Now imagine, as George did (p. 294) that a small village grows into a large city. Will wages be higher, will the return on capital be greater? No. What, then, will be higher? "Rent; the value of land. Go, get yourself a piece of ground, and hold possession." And when the city is built, we will find luxurious mansions armed with guards and alarms to protect themselves from the thieves and muggers that fester in the slums.

Though poverty in the midst of progress has puzzled even economists who should know better, "so simple and clear is this truth, that to see it fully once is always to recognize it" (p. 295). George describes those pictures made up of a labyrinth of lines, which you can keep staring at but can't make out—"until once the attention is called to the fact that these things make up a face or a figure." One follower of George looked at just such a picture, and then realized it showed a cat. Once he saw the cat, the figure was clear and obvious. Adherents of George's thought henceforth have said that when one grasps the central idea of the cause of poverty, you have "seen the cat!" The jumble of economic activity become clear, as springing from a central principle: "The great cause of inequality in the distribution of wealth is inequality in the ownership of land." Land has been the foundation of great fortunes, as said the Brahmins ages ago (George, p. 296): "To whomsoever the soil at any time belongs, to him belong the fruits of it. White parasols and elephants mad with pride are the flowers of a grant of land."

Governments throughout the world react to poverty mostly by treating the symptoms rather than the underlying causes. Old-age insurance, medical aid, food coupons, government housing, and welfare payments may prevent the poor from suffering, but they do not cure the problem. They instead perpetuate poverty, first by funding these programs with taxes on production, reducing employment and wages, and secondly, by

making it costly for the poor to escape poverty, since they both lose benefits and get taxed if they obtain employment. Minimum wages compound the problem by preventing the least able from working legally. Laws prohibiting drugs create large profit opportunities to those in the slums but the resulting violence and crime makes it even less desirable for business to employ people there.

Only a policy that removes the cause of poverty will cure it permanently. The ultimate antipoverty program is the desire of each person to better his or her own condition. All that is really needed is to avoid putting barriers in the way of self-improvement.

George (1883, p. 78) put the matter very succinctly: "There is in nature no reason for poverty." Remove the restrictions and tax barriers to employment and collect revenue from site rents instead. The elimination of taxation will raise the wage at the margin, and the collection of the land rent will, by eliminating the speculative and consumptive ownership of less productively used land, move the margin toward more productive land. With this remedy, the poverty problem will rapidly melt away as communities regenerate themselves spontaneously.

3. Unemployment

A person is "unemployed" if she or he is willing and able to work at the prevailing wage rate, but is unable to find employment. As George (1879, p. 5) put it, the unemployed are "condemned to involuntary idleness." Since an alternative to being employed by others is to become self-employed, the presence of unemployment implies also that it is difficult or infeasible to start one's own enterprise as well.

But why should this be? George (1879, p. 270) called it a "strange and unnatural spectacle" that willing workers "cannot find employment." In his book Social Problems (p. 8), George illustrates how odd this really is:

"Give us but a market," say manufacturers, "and we will supply goods without end." "Give us but work!" cry idle men.

Is it possible that the number of jobs might be limited? But people can create their own jobs, so this must imply that all desires have been filled. But the 10th foundational proposition of economics, as stated in Chapter 1, states that human desires tend to be unlimited. There cannot

be too much production in general, especially so when "people suffer for the lack of things that labor produces" (George, 1879, p. 270). Hence, there is always a demand for labor to fulfill these desires. The other original factor of production being land, labor can always be applied to land, unless something is blocking it—"somewhere there is an obstacle which prevents labor from producing the things that laborers want" (p. 271).

One possible block is a shortage of land, of natural resources. If the economy an area of the earth, such as a desert, is mainly geared to using the scanty vegetation, such as for grazing, then too many users will lead to poverty, but not necessarily to unemployment, unless the land is closed off to additional users. Here, then, is the key to unemployment. The problem is not an overall lack of natural resources. If that were so, the prices of basic commodities such as metals and grains would be going up as they became scarcer. But this has not happened. Even with today's massive population, the earth has plenty of resources available. We are not running out of oil, metals, agricultural land, or living space.

Some unemployment is short term and an unavoidable part of a dynamic economy in which enterprises change and workers search for better opportunities. This "frictional" unemployment consists of the time needed for job search, such as getting information and interviewing. This "natural unemployment" is not part of the jobless problem.

Another type of unemployment is called "structural." As some industries shrink or move to better locations, workers are left stranded. To be employed at their accustomed wages, they need to move or to become retrained. Again, this is natural in a dynamic economy, and in a free market, the prospering industries can absorb this structural unemployment, which again is temporary. "Full employment" is thus defined not as 100% employment, which is not even desirable, but as the situation in which the only unemployment is short-term frictional or structural.

"Cyclical" unemployment occurs during the depression phase of a business cycle, when many workers are laid off and can't easily find work because the entire economy is depressed. The remedy is to eliminate such cycles, as is discussed in the chapter on business cycles.

The unemployment problem as such is not the frictional, structural, or even cyclical types (which is primarily a cycle problem), but the chronic

joblessness in which workers cannot find jobs even during more prosperous times, and often remain unemployed for years. Why would labor be continuously unemployed?

If the problem is not a lack of land, or natural resources, it must be that some barriers have been erected between land and labor. Something is blocking the ability of labor to freely access natural resources to produce wealth. Who has done this? The only institution with the power to prevent labor from mixing with land is government.

The barriers erected by government include the prohibitions, restrictions, and taxes imposed on labor and enterprise, as well as barriers indirectly caused by government policy. As discussed in Chapter 2 on labor, high taxes create a wedge between the cost of labor to an employer (including the self-employed) and the net earnings of an employee. By making labor more expensive, this tax wedge reduces employment. Taxes on enterprise reduce profits and sales, and therefore reduce employment. Policy which induces land speculation (such as public works that increase land rent while not collecting that rent to finance the works) can raise the price of land above that warranted by present uses, thus reducing current enterprise and employment.

The situation becomes even worse when restrictions are imposed, such as minimum wage laws. Labor that would have been employed at the below-minimum rate now is unemployable. Unemployment insurance and welfare enable the unemployed to eat, but make it even more difficult to find employment, since switching from welfare to employment typically brings little extra reward for the lowest-paid workers.

Other restrictions, preventing people from working in certain occupations, days and hours, and other costs, such as the costs of filling out forms and making it difficult to fire workers, result in even less demand for labor and the attempt to substitute machinery for workers.

Employment is like a race around a track. If we put up high hurtles that the runners must leap over, the most able and determined can jump them, but the others cannot. The least able and motivated become unemployed. But if we remove all the hurtles, then all runners can go through, even the weak ones. Let the walls come tumbling down—eliminate the barriers of taxes and restrictions, and all willing and able to work shall be employed.

Unemployment is not only a personal and social tragedy, but wasteful of labor. As George (1883, p. 76) put it, "this enormous waste of productive power is due, not to defects in the laws of nature, but to social maladjustments which deny to labor access to the natural opportunities of labor and rob the laborer of his just reward."

Index

The Foundations of Economic Theory

Fred Foldvary

Fred E. Foldvary teaches economics at Santa Clara University, California. His books include Soul of Liberty, Beyond Neoclassical Economics, Public Goods and Private Communities, Dictionary of Free-Market Economics, The Half-Life of Policy Rationales, and The Depression of 2008. He writes a weekly column for The Progress Report and The Free Liberal, and is an associate editor of Econ Journal Watch and a member of the editorial board of the American Journal of Economics and Sociology.

BUSINESS EXPERT PRESS

cognella academic publishing
www.cognella.com 800-200-3908

ISBN 978-1-63157-204-3