Economics of Everyday Life

Christopher Bruce, PhD
Professor Emeritus
Department of Economics
University of Calgary

TABLE OF CONTENTS

A. INTRODUCTION - ABOUT THESE NOTES

MODULE 1. THE GREAT DEBATE

I often encounter people whose deep interest in public life has left them wishing that they had a better understanding of economics, but who feel that they do not have time to register in a formal course. These notes were written for them. On the one hand, most of the concepts that you would expect to find in a first year university course in Economics will be found here, plus many others. On the other hand, each of these concepts is explained in a simple, straight-forward way, allowing the reader to master them in a relatively short time.

Furthermore, the presentation is divided into modules, each of which is short enough that it should be possible to read and fully appreciate it in fifteen or twenty minutes – the time it takes you to commute to work by train or bus, or to enjoy your morning coffee break. Nevertheless, these notes would also be appropriate for an introductory course at a junior college or technical school; or as supplementary material for an introductory course at the university level.

Although the presentation is structured around what I call the *great debate* between market-based economies and planned economies, the notes do not lose sight of their broader purpose, which is to provide an understanding of all of the concepts normally included in introductory courses.

1. Markets Versus Planning

Every day, your news feed is filled with headlines like: Why are house prices so high? How should we deal with air and water pollution? Are the salaries of CEOs too high? What is the effect of allowing additional immigrants? Should the government provide a universal health care scheme?

All of these headlines concern issues that are the subject of economic analysis. And underlying most of them is the great debate about whether economic decisions should be made by *markets* or by *planning*.

By "markets," economists mean the interaction of producers and consumers, in which producers offer to sell goods at specified qualities and prices and consumers offer to purchase those goods also at specified prices. Sales in each market occur when buyers and sellers agree on a price/quality combination; that is, when *supply* equals *demand*. These notes will use the term *market-based economy* to refer to an economy that relies on markets to answer most of the questions raised above (and others like them).[1]

[1] In public debate, these economies are usually referred to as *capitalist*, meaning that capital – the equipment, factories, and land used to produce goods – is owned privately. However, as the focus of these notes will be on the decision-

By "planning," economists mean an economy in which decisions about the production and distribution of goods and services are made by government officials known as planners. Hence, a *planned economy* is one in which the questions raised above are answered by technocrats and politicians working in a planning office.[2]

These notes will investigate the relative advantages and disadvantages of these two economic systems, not to determine whether one is superior to the other, but to think about what the optimal mix of markets and planning might be; because even the most capitalist of countries, like the United States, use government agencies to produce education, roads, and policing; and even the most highly planned economies, like Cuba, allow private agencies to provide restaurants, taxi services, and hairdressing.

2. Microeconomics versus Macroeconomics

Economists divide their discipline into two broad categories: macroeconomics and microeconomics. Although these notes will mostly be about microeconomics, it will be useful to distinguish briefly between micro and macro.

Macroeconomics analyses questions about aggregate, economy-wide issues such as unemployment, inflation, taxation, international trade, and economic growth. It is macroeconomic issues that tend to dominate discussions in the financial pages of our newspapers. Why is unemployment so high? Would a reduction in corporate taxes increase economic growth? Is a trade deficit a bad thing? Should governments seek to balance their budgets? Etc.

Microeconomics focuses on the actions of individual components of the economy, like households, workers, and businesses. An often-quoted definition of microeconomics is:

> Microeconomics is the study of the allocation of ***scarce*** resources among ***competing*** ends.

What this means is that, as a society, we have available to us a huge variety of resources – land, air, water, trees, animals, minerals, people, time. But we do not have such large quantities of these resources that we can afford to use them wastefully. Relative to the size of our "demand" for goods and services, the amount of resources available to us is *scarce*. Somehow, we have to decide how these resources should be allocated among the alternative uses to which they could be put. For example:

---

making mechanisms by which goods are produced and distributed, and not on the ownership of the means of production, I prefer the term market-based economy.

[2] In public debate, these economies are often referred to as *communist*, meaning that capital is owned by the government. Again, however, I prefer the term planned economy, reflecting that the focus of these notes is on decision-making, not on ownership.

- Should land be used for housing or for agriculture and, if for agriculture, should we produce beef, wheat, or grapes?
- Why are the wages paid to some occupations higher than those paid to others? What causes wages of men to be higher than those of women, for example, or of whites to be higher than of people of colour?
- Should we harvest as many whales as we can? A small number of whales? Or are we going to leave them to roam free?
- How are we going to produce electrical power? With hydroelectric dams? Coal or oil-fired plants? Wind driven generators?
- And once we have decided what to produce and how to produce it, how will we decide who should receive society's output? Who gets a new car? A trip to Hawaii for Christmas? Who gets their newspaper or mail delivered to their door and who has to go to a corner box (or smartphone) to get it?

Microeconomics might be thought of as "the economics of everyday life." Whereas macroeconomics is concerned with issues that seem remote from us, like government deficits and the interest rate policies of the central bank, microeconomics is concerned with issues that affect us personally.

3. Three Fundamental Types of Microeconomic Questions

Economists often think of there being three broad types of microeconomic issues to be resolved:

What? What is to be produced? Given that we cannot have as many units as we could possibly want of everything we could possibly want, we have to decide which products are going to be produced in the greatest quantities. Do we want more red t-shirts if that means we will have fewer blue t-shirts? Do we want more bank machines if that means fewer TV sets? More tomatoes if that means fewer ski trips?

How? How are goods to be produced? For example, is bread going to be baked in wood-fired ovens at local bakeries or in gas-fired ovens at large manufacturing plants? Is wheat going to be harvested by hand or by combines? Are students going to be taught in small tutorials or in large lecture halls?

For whom? Once society has decided what is to be produced, some system must be devised for determining who is going to get those goods.

4. The Structure of These Notes

As we live in an economy in which the three questions raised above are answered by buying and selling in markets, most of these notes will discuss how a *market economy* works (or fails to work).

First, however, it will be useful to investigate how three other types of "economies" – individuals, communes, and *central planning* - answer these questions. In each case, the primary focus of the analysis will be the question: does the economic system allow society to use its (scarce) resources in the manner that provides the greatest benefits to its citizens? That will be the focus of Part B.

Parts C through F will then investigate how a market economy works and ask whether it can answer the three fundamental economic questions better than do the alternatives. Finally, in Part G, the notes recognise that even though we often think of ourselves as living in a market, or capitalist, economy, over a third of the economic decisions in our country are made by government agencies – we live in a *mixed economy*. In that Part, therefore, the tools that were used in parts C through F will be applied to the understanding of how government agencies operate.

B. NON-MARKET ECONOMIES

MODULE 2. INDIVIDUAL DECISION-MAKING - THE ROBINSON CRUSOE ECONOMY

Imagine that Robinson Crusoe has been abandoned on an uninhabited island with some basic supplies for survival – for example, a machete, a fish net, and a flint (for starting a fire). There are two types of food on the island: coconuts (up tall trees) and fish (in a stream). Like every other "society" Crusoe has to answer the first two of the fundamental microeconomic questions: what to produce and how to produce it? (He doesn't have to decide "for whom".)

1. What?

Each morning, Crusoe goes out to collect coconuts and catch fish. The amount of time he spends on each will determine *what* he will be able to consume. How will he allocate his time?

Economists assume that Crusoe answers this question using the following approach. When he sets out for the day, he calculates how many fish he can catch in the first hour and how many coconuts he can collect. Assume that he estimates he can catch one fish or collect three coconuts. Next, he asks himself: which of these outcomes would give him greater pleasure? If he prefers one fish to three coconuts, he spends his first hour fishing.

He repeats this process for every hour that is available to him. Note that, in following this process it is quite likely that at some point, he will decide to use the "next" hour collecting coconuts, even if he chose catching fish for the first few hours. The reason for this is that, as individuals consume more and more of any one good, the less is the pleasure they will get from "one more," incremental unit of that good. So, even though the first fish may give Crusoe more pleasure than the first three coconuts, it is quite possible that, say, the fourth fish may not give him more pleasure than the first coconut. So, after he has fished for three hours, (and caught three fish, and picked no coconuts), he may then switch to collecting coconuts in the fourth hour, etc.

Put another way: when Crusoe spends the next hour catching fish, the benefit of that activity is the pleasure he will receive from the number of fish he can catch, and the "cost" is the pleasure foregone from the number of coconuts he could have collected. So, if one fish would give him ten "units" of pleasure whereas three coconuts would give him six units of pleasure, the extra, or *incremental*, benefit of that hour spent catching fish exceeds the incremental cost. [Note: economists often refer to "marginal" costs and benefits instead of "incremental." I will use "incremental" as it sounds less technical.]

The rule that will allow Crusoe to maximise his *total* net benefits is that, at the beginning of each hour, he should compare the incremental benefit against the incremental cost of fishing (which is measured as the value of the coconuts that are foregone). If the incremental benefit exceeds the incremental cost, he should use the hour for fishing. And he should continue adding more hours to fishing as long as the incremental benefit from the next hour exceeds the incremental cost. When

he reaches a point at which the incremental cost exceeds the incremental benefit, he should switch to his next best activity, collecting coconuts.

This may all seem a bit esoteric or convoluted, but it reveals two basic points that apply to *all* economies. First, the cost of any activity is the amount of pleasure that would have been obtained from consumption of the next best set of goods (the goods that have to be foregone). Economists refer to this as the *opportunity cost*. [Opportunity cost will be discussed in more detail in Module 3 as it is one of the most important concepts in economics.]

Second, in order for any society to obtain the maximum amount of pleasure from the resources available to it, it should ask, with respect to each possible activity: does an extra unit of this activity provide more pleasure than could be obtained from the "next best" activity? If so, it should increase that activity; if not, it should decrease it. (In Crusoe's case, he asks: does the number of coconuts I can harvest in the next hour give me more pleasure than would the number of fish I could catch?)

MODULE 3. APPLICATION - OPPORTUNITY COST

The concept of opportunity cost that was introduced in Module 2 is so important when thinking about economic decisions that it will be worthwhile to consider some examples of how it applies in real world decisions.

1. Non-economists often assume that rich people will retire earlier than poor people "because" they can afford to. But, if "rich" means high annual income and "poor" means low annual income, then the opportunity cost for rich people to retire is higher than it is for poor people. The rich person might forego earnings of $150,000 per year, whereas the poor person only $30,000. Accordingly, people in occupations such as lawyer and doctor often retire much later than most people; and very rich people, like Warren Buffet, often continue working well into their 80s and 90s.

2. Assume that Ms A is thinking of starting a restaurant. It will cost $3 million to build the restaurant, and the building will last for 10 years - a cost of $300,000 per year. If she expects to serve about 30,000 customers each year (say, 100 per day for 300 days), the building costs are about $10 per customer. Assume also that the cost of labour and materials (mostly food and heating bills) is about $20 per customer. In that case, in order to break even over the ten years, she will have to charge $30 per customer. But what happens if there is a downturn and she finds that she can't charge more than $25? Should she shut down, because she is not covering her costs? The answer is 'no' *if* there is no alternative use to which her building could be put as, in that case, the opportunity cost of the building is zero, and the "economic" cost of each meal is just $20. (If she keeps operating, she will make $5 more per customer than if she had shut down.)

3. Although we tend to think of the cost of education in terms of tuition fees and the costs of textbooks, the most important cost is the opportunity cost of foregone income. For example, if you could have earned $50,000 a year if you hadn't gone to university, and you can earn $10,000 a year while you are at university (perhaps in a summer job), then five years of university has an opportunity cost of $200,000 (= 5 x ($50,000 - $10,000)), plus the cost of tuition and books, which would probably be about another $50,000 ($10,000 per year). So, if a university education allows you to increase your annual income by $10,000 per year, it will take you 25 years to recover your investment.

4. Imagine that you decide to use a $500,000 inheritance to start up a business. At the end of the year, your accountant tells you that your revenues were $350,000 and your costs (for materials, rent, salaries of employees, etc.) were $280,000, for an *accounting* profit of $70,000. Before you can determine whether this venture was *economically* profitable you have to deduct your opportunity costs. One of those is the interest foregone on your inheritance: for example, if you could have invested it at 5%, you would have earned $25,000. A second opportunity cost is the income you have foregone because you didn't work at your "next best" occupation. If that would have paid you $60,000, then adding that to foregone interest income gives you a total opportunity cost of $85,000. When you include those costs in the calculation of the profitability of your

business, you find that you have actually lost $15,000 – that is, you have earned $15,000 less than if you had stayed at your next best job and had invested your inheritance in the bank. [Note: When opportunity costs are included in the calculation of profit, we call the resulting profit level ***economic profit***[3]. In the case described here, your economic "profit" was *minus* $15,000.]

5. One of the reasons that the money price of houses on the edge of a city are lower than the prices of comparable houses near the city centre is that the true cost of any house must include the cost of commuting to and from your job. Imagine, for example, that you would be willing to pay $20 per day to avoid having to sit on a train for sixty minutes each way to work, and that a ticket for such a commute costs you $10 return. In that case, the daily economic cost of living sixty minutes away from your job, instead of within easy walking distance, is $30 ($10 for the train ticket plus $20 opportunity cost of your time). If you work 200 days per year, then you would be willing to pay $6,000 per year more for a house near your work than one far away. In twenty years, that could account for a $120,000 difference in house prices.

[3] "Economic profit" has been put in bold letters to encourage you to remember it, as it will prove to be very important later in these notes.

MODULE 4. COMMUNAL DECISION-MAKING

When non-economists think about what an ideal economy would look like, they often imagine a communal society, in which everyone works together cooperatively. This Module briefly considers how one such society operated.

Castaway 2000

When Crusoe wanted to decide "what" to consume, he just had to ask himself, which do I prefer, one fish or three coconuts? And he didn't have to answer the "for whom" question because he was the only consumer.

But how would a society of, say, 35 people make decisions about the multitude of "what" questions that faced them: what they were going to have for dinner tonight, or how much time they would spend building housing, or how many sheep they would raise? And once they had decided those questions, how would they decide how to divide their goods among themselves?

An interesting example of the problems that are encountered in societies of this size was observed in the BBC television series *Castaway 2000*. In that series, approximately 35 people were placed on a remote Scottish island, along with farm animals, seeds, and the supplies to build houses, barns, and greenhouses. There was no competition among participants, as in the U.S. television series Survivor. Rather, the goal was to find out whether a group of strangers could work cooperatively to live together under strenuous conditions.

Many of the problems that were discussed in the program arose from the difficulties that the castaways encountered making communal decisions (that is, decisions that affected everyone). Three examples include:

1. With their limited resources, it was impractical for each individual or family to cook for themselves. Cooking was done communally. This created two problems.

First, it was not clear how to allocate the kitchen duties among the members of the community. At first blush it might appear that those duties should be given to those who were the best cooks. But what if those individuals preferred to be outside, tending the garden or herding the cows? (Perhaps the best cooks liked the cooking aspect, but not the preparation and clean up.) Wouldn't they try to "hide" their culinary skills? And what if the second-best cook was terrible at carpentry but the best cook was a very good carpenter? Wouldn't it, in that case, be preferable to have the second-best cook in the kitchen and the best cook out building houses? On Castaway 2000, it appears that the community gave up trying to solve this problem. Instead, at first, they just required that each adult take a turn working in the kitchen. But this caused serious divisions among the residents as some of them pointed out, correctly, that they (a) were not very good cooks and (b) were far more valuable to the community if they worked as carpenters and labourers. Eventually, the community

resolved this issue but only at very great cost, both in terms of the amount of time required and in terms of the amount of ill feeling that was generated.

Second, because cooking was done communally, it was difficult to tailor menus to individual preferences. In the end, basically everyone had to eat the same foods, like it or not.

2. The participants agreed, in advance, to limit the number of pets on the island, (because pets had to be fed from the group's limited supplies). However, one of the dogs became pregnant and delivered nine puppies. Of course, some of the residents fell in love with the puppies and wanted to keep them whereas others wanted to live up to the original agreement. Again, this created lengthy debates within the group, in which those who wished to keep the puppies had to convince the others that they (the "pro-puppy" group) would obtain sufficient pleasure from the puppies to offset the loss of pleasure the group would suffer due to reductions in the amount of food available for human consumption. Ultimately, a small number of puppies was kept on the island and the rest were sent to the residents of a nearby island.

3. Some of the issues concerning food were resolved by allocating a limited amount of supplies to individuals and to families (some of the participants were family groups) for their personal use. But this raised additional problems. There were accusations that some participants were making unreasonable requests for "special" rations; and that some participants were "hoarding" food. These disputes became sufficiently rancorous that one family left.

In each of these three cases, the underlying problem was that it was difficult for the members of a large group to identify the true preferences of each of the other members of that group. As a result,

a. some people tried to take advantage of the lack of information to bias outcomes in their favour (for example, by claiming that they had a strong preference to do something, or a strong preference to avoid doing something);
b. even when individuals were not trying to bias the outcome, others *suspected* that they were trying to do so (and had insufficient information to be able to determine whether their suspicions were correct or not);
c. the community devoted a considerable amount of time negotiating over allocations of goods and responsibilities; and
d. the community often "gave up" trying to find the best solutions - choosing instead to rely on simple formulas, such as requiring that everyone share kitchen duties equally.

These problems will become infinitely larger when a society is composed of hundreds of millions of people, or even, in the case of China, billions.

MODULE 5. APPLICATION - FREE TRADE

One of the issues that arose on Castaway 2000 was that there were often situations in which the person who was most productive in one activity, say cooking, was also the most productive in another, say building houses. How would you allocate duties between a productive person and an unproductive one?

The answer, economists have realized, is that you should allocate duties according to the parties' *comparative* productivities. For example, assume that individual A could produce either 100 units of food per hour or 50 units of carpentry, whereas person B could produce either 60 units of food or 40 units of carpentry. Assume also that they each have eight hours available.

The comparative advantage rule says that, as A produces 66 percent more per hour than B (100/60) in cooking, but only 25 percent more (50/40) in carpentry, A should specialise in cooking and B in carpentry.

For example, assume that each of them was initially working four hours in cooking and four hours in carpentry. If A transfers one of his hours from carpentry to cooking, he will produce 100 more units of food and 50 fewer units of carpentry. Assume that, at the same time, B transfers 1.25 hours from cooking to carpentry, thereby producing 75 fewer units of food and 50 more units of carpentry. In total, there has been an increase in food and no reduction in carpentry. So, a mutually beneficial trade is possible. For example, if A gave 80 units of food to B in return for 50 units of carpentry, the net effect is that A would gain 20 units of food (100 – 80) and B would gain 5 (80 – 75); while both would end up with the same amount of carpentry as they had to begin with.

What we see is that, even though A is more productive in both activities (A has an *absolute* advantage in both), it would not be optimal to have A do both carpentry and cooking. Both parties can gain if A specialises in cooking and trades for carpentry; and B specialises in carpentry and trades for cooking.

This is the argument for free trade between countries. Even though one country may be more productive at producing just about everything, it will pay that country to specialise in the production of the goods in which it has a comparative advantage, and trade for the goods in which it is *relatively* less productive.

MODULE 6. CENTRAL PLANNING

If the Castaway 2000 example is multiplied by a million, we have Canada; by a hundred million, we have the U.S.S.R.; and by 300 million, we have China. With this many people, try to imagine how society might resolve the questions of what, how, and for whom.

Recognising that 100 million people couldn't get together after dinner to decide these questions, as did the Castaway participants, many "communal" (i.e. communist) states appoint a *central planning agency* to determine how to "allocate scarce resources among competing ends." [Note: it is not just communist countries that have central planning; any country that produces goods and services using government agencies, is "planning" the allocation of resources, rather than allowing the market to make that allocation.]

Imagine that the goal of the central planning agency was to allocate resources in such a way as to maximise the total welfare of all of the residents of the country. How might it attempt to answer each of the what, how, and for whom questions, and what kinds of problems could it expect to encounter?

1. What?

There are two fundamental issues that the central planning agency has to face when deciding what to produce. First, it has to identify what consumers want. Second, it has to determine whether the plans it has created have been put into place.

1.1. What do people want?

How does a central planner decide:

: whether to make cars, running shoes, TV sets, loaves of bread?
: whether bread is to be brown, white, French, German?
: what the quality of goods is to be?

Possible answers:

: Produce those goods that have ***traditionally*** been consumed, or, perhaps ***what was produced "last year"***

: Allocate resources to those goods that are ***necessary***

: Set ***prices*** for the goods that the government produced and then allow consumers to purchase whatever they wanted at those prices

: The planner could determine what people want based on ***polls/surveys/ elections/referenda/complaints***

Why might these "solutions" not be efficient?

: tradition/what was produced "last year": If what was produced traditionally, or last year, was not what people wanted, or if it is not what they want now because their tastes have changed, the planner will produce the "wrong" thing. That would be inefficient if there was something else which could have been produced at the same cost, yet which would have been preferred. That is, producing the traditional set of goods is inefficient if there is a second set which everybody would have preferred.

: necessity: "Necessity" does not provide a useful criterion as there are extremely few products that are truly "necessary" - oxygen and some minimum amount of water, food, and heat. But in a society as wealthy as ours, most of the allocation decisions do not involve minimum amounts. We are not deciding whether people should have a minimal set of clothes; we are deciding whether they should have one, two, three, or four new pairs of shoes each year and whether those shoes should be crocks, runners, pumps, or high heels.

: prices: Setting prices and allowing consumers to buy whatever they want at those prices does not provide a useful allocation mechanism in a truly centrally planned economy as it is the planners who would set the prices. How would they do that?

: polls/surveys/elections/referenda/respond to complaints: There are many problems with relying on what people tell you they want in surveys. Most importantly, it is usually difficult to give them appropriate information about what the costs of alternative goods will be; and individual consumers often don't know what alternatives are available to them. (The difficulties of using polls, surveys, etc. will be discussed in detail in Module 47 of these notes.)

1.2 How do planners know what has been done?

Assume that the planning agency has solved all of the questions that were set out above. Now what does it do? It makes up a plan, which it sends out to all of its producers, telling them what to produce. The producers write back, saying "yes, we have done exactly what you asked (indeed, we have been even more efficient than anticipated and produced more than you asked)." How does the central planner in, say, Beijing know that the baker on a back street in Lijiang has actually produced what was in the central plan? or that the quality was what the planner had specified?

This became a major problem in 1950s China and Russia. When the planning agency in Beijing instructed farm cooperatives to increase their rice production by, say, 25 percent and to deliver that increase to Beijing or Guangzhou, the cooperatives wrote back to say that that is what they had done – even though they had done no such thing. They got away with it because the record-keeping system was so poor. And when the Russian planning agency instructed steel mills to produce additional quantities of steel, at specified quality levels, the mills responded by providing the quantities required, but at lower quality (because quality was easier to disguise than was quantity).

So, if it is difficult for central planners to determine what *citizens* want, how can they resolve the question of "what to produce?"

1. They could give the people what the central *planners* want – the employees in the central planning office may just produce the bakery goods (bagels, chocolate cakes, rye bread) that they like.

2. Give the people what the planners think they *should* want: The Chinese government did not ask what kinds of clothing its citizens wanted, it just told them that they were all going to wear the same clothes.

3. Give the people whatever is easiest to make (including low quality and limited selection). With respect to "limited selection," the office of bakery goods in the central planning office of the Yugoslav government gave up trying to determine which bakery goods people wanted. Instead, it just produced one kind of product on Monday (say, sourdough bread), another kind on Tuesday (chocolate cakes), another on Wednesday, etc.

4. Give the people what politicians and special interest groups want. Probably the easiest response from the central planning agency is to give in to groups that have the greatest power or the most persistent lobbying.

2. How?

Assume that the planners have decided how much of a particular product, say loaves of bread, to make. How do they decide *how* that product is to be made? For example, how do they determine whether bread is to be made in local bakeries, which use a lot of labour relative to the amount of capital (machinery) and energy (e.g. electricity); or whether it is to be made in factories, which use a lot of capital and energy relative to labour?

2.1 Opportunity cost problem

If the planners are attempting to produce efficiently, they will need to know what the effect of switching from one technology to another will be on output in other industries. When you switch from making bread in local bakeries to producing it in factories, you will probably need more steel and electrical energy. What is the *opportunity cost* of that steel and energy? That is, where does the steel come from? Is it taken from production of cars, or TV sets, or railways, or ships? Or is *extra* steel to be produced? If so, where are the workers going to come from to mine the extra iron ore and work in the steel mills?

The *opportunity cost* of producing bread in a factory relative to local bakeries depends on the answers to questions like these. Most importantly, that "cost" is ultimately measured, not in

terms of units of steel and electricity, but in terms of the other *consumer products* which have to be foregone when production is transferred from bakeries to factories.

The *benefit* of using factories would depend on answering the question of "what would happen to the workers who previously worked in the bakeries?" Would they now work in banks, or airlines, or universities? And, if they did so, how much would they produce?

Moreover, even if the planner knew what had to be sacrificed in order to build the bread factory and what the workers released from the bakeries were now producing, how would the planner know whether that constituted a net gain? For example, assume that the new factory produces 10,000 more loaves of bread than the old bakeries; that the steel to build the new factory comes from the automobile industry, where it would have been used to build 1,000 cars; and that the 100 workers who were laid off from the local bakeries and automobile factories all became elementary school teachers. In that case, the net gain from mechanizing bakeries would equal the increase in the number of loaves of bread, plus the increase in the amount of elementary school "production," minus the decrease in the number of cars produced. How would the planner determine whether the "exchange" of 1,000 cars for 10,000 loaves of bread and 100 elementary school teachers represented a net gain to society?

2.1 Incentive problem

How do planners monitor workers to make sure that they are working as hard as they could; making the highest quality of products that they can; and using the best techniques available?

[Note: This is one of the fundamental problems that face governments in mixed economies (like ours) as well. How do we ensure that university professors, or park wardens, or (government) highway construction crews work as efficiently as possible?]

3. For whom?

Planners face four related problems when deciding to whom the available goods are to be given:

1. Identifying the optimal distribution of income. Which families should receive the largest/smallest shares of goods and products?

2. Determining which families will obtain the greatest benefit from a particular set of products. For example, how does the planning agency decide which families will receive the largest allocations of gasoline, or apartment space, or winter clothing?

3.Selecting the mix of products and services for each household. For example, if the planning agency has decided that each adult is going to receive 1,500 calories per day, how does it select the mix of foods that are to make up those calories?

4. Meeting incentive criteria. The manner in which income is distributed can have a significant effect on individuals' willingness to work hard. If income is distributed according to some measure of "need" rather than a measure of "contribution to society," the incentive for individuals to contribute to society (e.g. by working harder) may be reduced.

4. Summary

Problems with Central Planning

1. Difficult to identify consumers' *relative preferences* for different goods - e.g. how many TVs is an individual *willing* to give up to get an extra room in her apartment? a holiday? a dress?

2. Cannot easily convey to consumers what the *opportunity costs* of their choices are - e.g. if I want an extra TV, what do I have to give up? A room in my apartment? a holiday? a dress?

3. Difficult to set *incentives* for planners, government officials, and workers to do what society wants.

Another problem with central planning, that falls outside the purview of economic analysis, is that the centralization of power that results makes it easier for totalitarian governments to arise – see Stalin and Mao.

Advantages of Central Planning

1. The government can *distribute income* and goods in any way it wants, as long as it is willing to accept the consequences (in terms of incentives).

2. Another “benefit,” again outside economics, is that central planning, with its ethos of all members of society coming together to decide how resources are to be used, is more attractive to many people than is capitalism, with its focus on individuals and self-interest. Christian and Muslim scholars, for example, seem to prefer cooperation over competition.

C. MARKET ECONOMIES - SUPPLY AND DEMAND

MODULE 7. INTRODUCTION - THE MARKET-BASED ECONOMY

1.The Structure of the Analysis

Part B of these notes described how various non-market economies allocated scarce resources, beginning with a simple Robinson Crusoe economy; then working through communal societies, like Castaway 2000, to full-blown centrally planned economies, like China.

Parts C through F will investigate how *market*-based economies, like ours, allocate resources.

Part C – Supply and Demand: The modules in Part C develop a very simple model of the market economy based on the concepts of supply and demand. Despite its simplicity, it will be seen that this model can be used to answer many interesting questions concerning factors such as the impact of price ceilings and floors and the causes of the decline in agricultural sector.

Part D - Competitive Industries: Part D shows that the supply and demand model is actually a simplified description of what economists call a competitive industry; that is, an industry in which there is a large number of relatively small firms. It is the competitive model that underlies the arguments made by most advocates for market economies. This model will be applied to questions concerning the determination of insurance premiums, causes of high house prices, and the impact of minimum wages.

Part E – Non-Competitive Industries: Market economies often deviate from the competitive ideal in ways that prevent markets from allocating resources efficiently. One source of inefficiency, discussed in Part E, arises when industries are non-competitive; that is, when the number of firms in an industry is relatively small. Three types of non-competitive structure will be considered: monopoly, monopolistic competition, and oligopoly. Some of the issues discussed in this section include patents and copyright, franchise arrangements, and cartels.

Part F – Market Shortcomings: The modules in this part discuss an array of additional sources of market inefficiency, including externalities (leading, for example, to pollution), the principal-agent problem, and racial and sexual discrimination.

2. Characteristics of a Market

In a market-based system, we generally think of there being two types of actors (or agents): producers and consumers. In a market, two basic factors interact to allocate scare resources:

- ***prices*** - act as ***signals***
- ***profits*** - act as ***motivators***

How does this system answer the three basic questions?

How?

If consumers would like to purchase more goods and services than are available at current prices, producers will begin to run out of goods to sell. Seeing the opportunity to increase their profits, they will increase their prices and offer to sell more goods.

What?

The inputs that are most productive, in terms of contributing to the creation of valued outputs, will be in greatest demand. Unless these goods are in great abundance, their prices will be relatively high. As a result, the greater is the value of the output which an input can create in industry X, the more will other industries have to pay for it in order to attract it away from X.

Thus, producers are encouraged to seek ways to use inputs which are not very productive in other industries; and to avoid the use of inputs which are productive elsewhere.

For Whom?

One of the prices in a market economy is the price of labour - the wage rate. As workers are "inputs" into the production process, there will be a tendency to pay the highest wages to those individuals who possess skills which are in short supply and which allow them to produce outputs which are valued by society will be paid the highest wages.

MODULE 8. DEMAND

There is an old joke that you can make a parrot into an economist by teaching it to say "supply and demand." The truth behind the joke is that the concepts of supply and demand are very simple (you will be pleased to hear) and that the interaction between them underlies a very large part of economic behaviour.

This module and the one that follows, describe the basic concepts of supply and demand. It will then analyse how they interact to determine the allocation of scarce resources – that is, how supply and demand interact to answer the questions of what, how, and for whom.

1. Determinants of Demand

"Demand" measures the amounts of goods and services that consumers are *willing* to purchase - as opposed to the amounts that they *actually* purchase. Some of the factors that will determine the demand for a particular product include:

: tastes and preferences, including
    : is the good "necessary?"
    : psychological and sociological factors (including fashion and fads)
    : weather
: population size (e.g. demand for houses increases as population increases)
: average household income
: income distribution
: prices of "other" goods and services
    : complements (demand for chips might increase with demand for fish)
    : substitutes (demand for beef might decrease with demand for chicken)
: price of the good/service itself
    : includes any opportunity cost

2. The Demand Curve

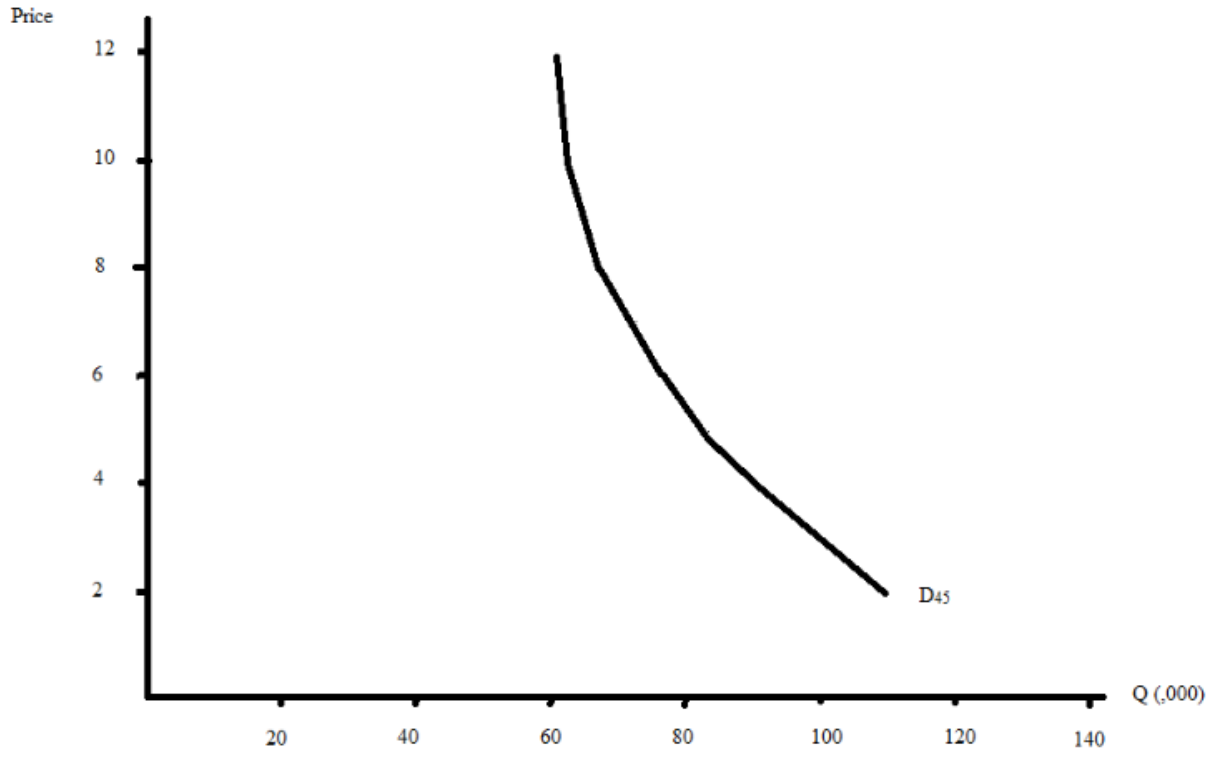
The demand curve plots the relationship between the *quantity* of a product that is demanded and the price that is charged for that product, all else being held constant. Price in this case is the "causal" or "independent" variable and quantity demanded is the "dependent" variable. That is, changes in price are assumed to "cause" changes in demand.

Table 1, below, provides hypothetical data for the number of pizzas demanded in a week in a small city, Q; and Figure 1 plots the demand curve, D_{45}, implied by Table 1. In this case, it plots the quantities demanded at various prices when average income is held constant at $45,000 per year.

Table 1 Demand for Pizzas

Demand for Pizzas (Average Income $45,000)	
Price/Unit (Independent Variable)	Quantity Demanded/Week (Dependent Variable)
$2.00	110,000
4.00	90,000
5.00	83,000
6.00	77,000
7.00	73,000
8.00	67,000
10.00	62,000
12.00	60,000

Figure 1 Demand Curve for Pizzas



3. Changes in Demand

A change in *quantity* demanded

In Table 1, if income is \$45,000 and price is \$12.00, then quantity demanded, Q_d, is 60,000; and if price falls to \$10.00, Q_d increases to 62,000.

An increase in Q_d that results from a fall in price, holding everything else constant, is called a ***change in quantity demanded***. (Here, the "everything else" that is being held constant is income.) It is associated with a ***movement along*** the demand curve. (When price changes, the demand curve remains unchanged, by definition.)

A change in demand

An increase in Q_d that results from a change in a variable *other* than price is called a *change in demand*. It is associated with a *shift of* the demand curve.

Table 2 presents hypothetical data for the quantity of pizzas demanded at two income levels, \$45,000 and \$55,000. These data have also been plotted in Figure 2, providing two demand curves, D_{45} and D_{55}. (Although it has been assumed here that, at each possible price, consumers are willing to purchase more pizzas when average income is \$55,000 than when it is \$45,000, there are cases in which quantity demanded decreases when income increases – e.g. demand for cheap cuts of meat.)

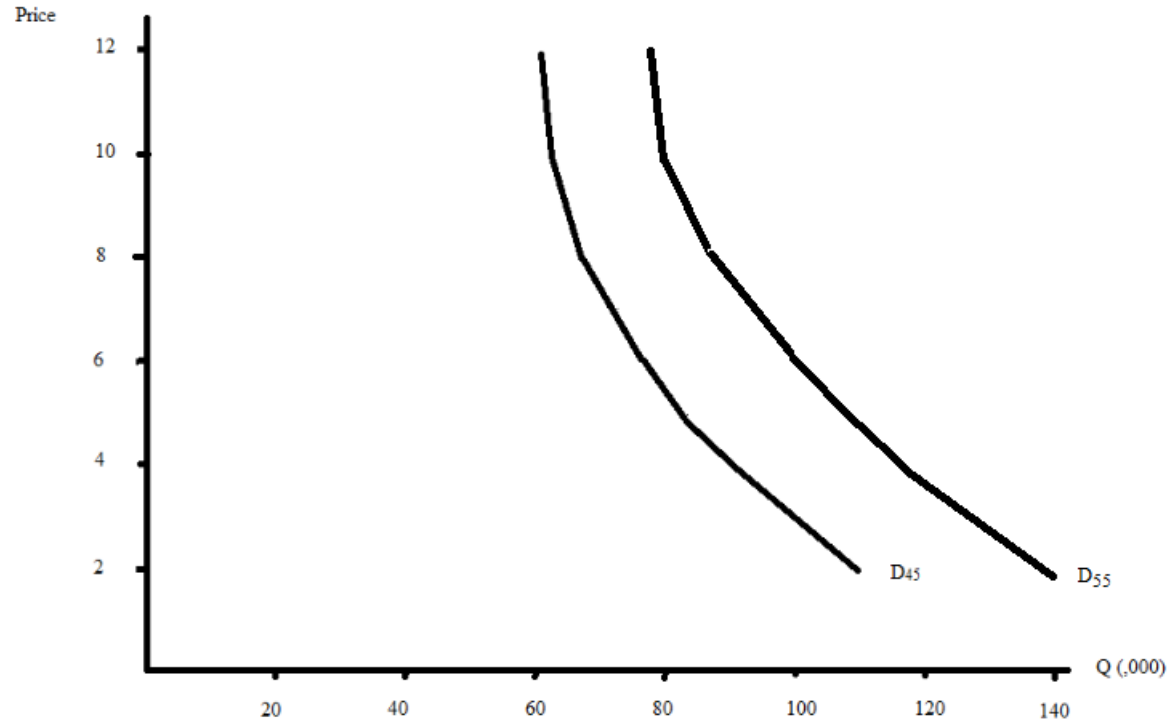
Table 2 Effect of Income on Demand

Effect of an Increase in Average Income on the Demand for Pizzas		
Price/Unit	Quantity Demanded/Week at:	
	I=\$45,000	I=\$55,000
\$2.00	110,000	140,000
4.00	90,000	116,000
5.00	83,000	107,000
6.00	77,000	100,000
7.00	73,000	92,000
8.00	67,000	87,000
10.00	62,000	80,000
12.00	60,000	77,000

With Q_d being higher at every price, the demand curve may be thought of as having been *shifted* to the right. Alternatively, the change in quantity demanded shown in Figure 2 may be thought of as having resulted from a *vertical* shift of the demand curve: at every quantity demanded, consumers are now willing to pay a higher price. For example, in Table 2 it is seen that if price is

$6.00 and average income is $45,000, the quantity of pizzas demanded is 77,000. If income increases to $55,000, 77,000 pizzas will still be demanded if price rises to $12.00.

Figure 2. Shift in Demand



4. Price Elasticity

It is often useful to have a measure of how responsive a dependent variable will be to changes in an independent (or causal) variable. For example, by how much will the harvest of fruit change in response to a change in rainfall? Or by how much will attendance at football games change in response to a change in the percentage of games won by the home team?

When a percentage change in the dependent variable is divided by the percentage change in a causal variable, we get a measure known as *elasticity*. For example, if a ten percent change in annual rainfall results in a 20 percent increase in the harvest of grapes, the "rainfall elasticity" of grape production is 2.0 (= 20/10). And if a ten percent increase in games won results in a 5 percent increase in attendance, the "win elasticity" of attendance is 0.5 (= 5/10).

Economists often apply the concept of elasticity to demand (and, as will be shown later, to supply). For example, in Table 1, the 25 percent increase in price from $4.00 to $5.00 lead to a 7.8 percent decrease in quantity demanded, from 90,000 to 83,000. Hence, in that region of the demand curve the *price elasticity of demand* was -0.31 (=-7.8/25); although, for convenience, it is the absolute

value of price elasticity that is usually reported – here 0.31. When the price elasticity of demand is less than 1, it is said that demand is price *inelastic* – the quantity demanded is not very responsive to changes in the price – and when price elasticity is greater than one, demand is price *elastic*.

Price elasticity of demand is very important to producers because, if demand is price inelastic, an increase in the price will lead to an increase in the total revenue (price times quantity) collected by producers. For example, that the price elasticity of demand between $4.00 and $5.00 in Table 1 is 0.31 (less than 1), implies that total revenue at $4.00 ($4 x 90,000 = $360,000) will be less than total revenue at $5.00 (= $5 x 83,000 = $413,000).

Note also that Tables 1 and 2 could be used to calculate the *income elasticity of demand* (the percentage change in quantity demanded divided by the percentage change in income).

5. Determinants of Price Elasticity of Demand

The general rule is that demand will be more "price elastic" the more available are substitutes. For example, if goods A and B are close substitutes for one another (A is a donut from Tim Horton's while B is a donut from Dunkin Donuts), when the price of A increases, we would expect that consumers would switch to B. So, a small increase in the price of A might produce a very large change in quantity demanded of that product. Conversely, if there are no close substitutes, it is possible that consumers might make only a small change in purchases when price increased.

There are three primary determinants of "substitutability."

- ***"Need"*** If the product is necessary, like water, there could be a substantial change in price before there would be a significant change in quantity demanded.
- ***Time*** The longer is the time period being considered, the easier it will be to substitute one product for another. For example, when oil prices rose dramatically in the early 1970s, people were stuck with gas guzzling cars. But, as those cars wore out, consumers were able to replace them with new, efficient cars. Thus, although the demand for gasoline was, at first, price inelastic, that elasticity rose over time.
- ***Narrowness of definition*** It is likely that the price elasticity of demand for cars in general will be lower than it will be for individual brand names of cars. It is easier to substitute a Chevrolet for a Ford than it is to substitute public transit for private automobile use.

MODULE 9. SUPPLY

"Supply" measures the maximum amounts of goods and services that producers are *willing* to sell - as opposed to the amounts that they *actually* sell.

1. Determinants of Supply

It is usually assumed that firms will be more willing to supply goods, the higher are the profits they can make. In turn, profits increase as revenue increases and as costs decrease. Thus, firms can be assumed to be willing to supply a greater amount of output, the higher is the price offered for that product (as price determines revenue) and the lower is the cost of production.

1.1 Cost of Production

Opportunity Cost
The cost of using a particular input in the production of a good or service is the value of the output which that input could produce if it was used for a different purpose. For example, although we may think of the cost of a wooden table as being the money price of the wood and labour that enters into it, the true cost is the value of the goods and services that could have been produced by those inputs in their "next best" uses.

Normally, in a market economy, these opportunity costs are signaled by the prices charged for inputs. For example, if there is an increase in demand for wooden houses, that would increase the opportunity cost of using wood in the production of furniture. This change in opportunities would be signaled to the furniture manufacturer through an increase in the prices of wood and carpenters. (Wages are just the "price" of work.) When economists talk about the "costs" that firms consider when determining how much they will supply, they are talking about those firms' opportunity costs. [A number of examples of opportunity cost were provided in Module 3 of these notes. Another example, concerning "economic profit" will be discussed in Module 15.]

Technology
Technological change usually acts to reduce the costs of production. Sometimes that change is *exogenous*; that is, it arises by "chance" even though there is no particular shortage of inputs. For example, the race to put a man on the moon has produced many cost-saving devices, such as communications satellites and Teflon coatings.

Most technological change is *endogenous*, however – it arises in response to cost increases brought about by shortages of inputs. For example, many energy-saving inventions have been made in response to the increases in the costs of petroleum that arose from the establishment of OPEC.

Time
When the price of an input increases, firms can often minimise the impact on costs by switching to techniques that do not use that input, or use it only sparingly. For example, when the price of oil

rises, firms may switch from oil-fired furnaces to gas- or coal- or electricity-fired; and taxi companies may switch from gasoline engines to electric or natural gas engines.

Often, these changes cannot be made easily in the short run. But as old equipment wears out, it may become relatively inexpensive to introduce new techniques. If you have a perfectly good electric heating system in your house, you are not likely to run out and buy a natural gas furnace the day after electricity prices rise (even if it is clear that they will stay high). But, when the electric heating system needs replacing, or you are renovating the house for other reasons, you might at that time consider installing a natural gas furnace. Thus, whereas an increase in input costs may increase the cost of output significantly in the short run, in the long run, costs may increase only slightly.

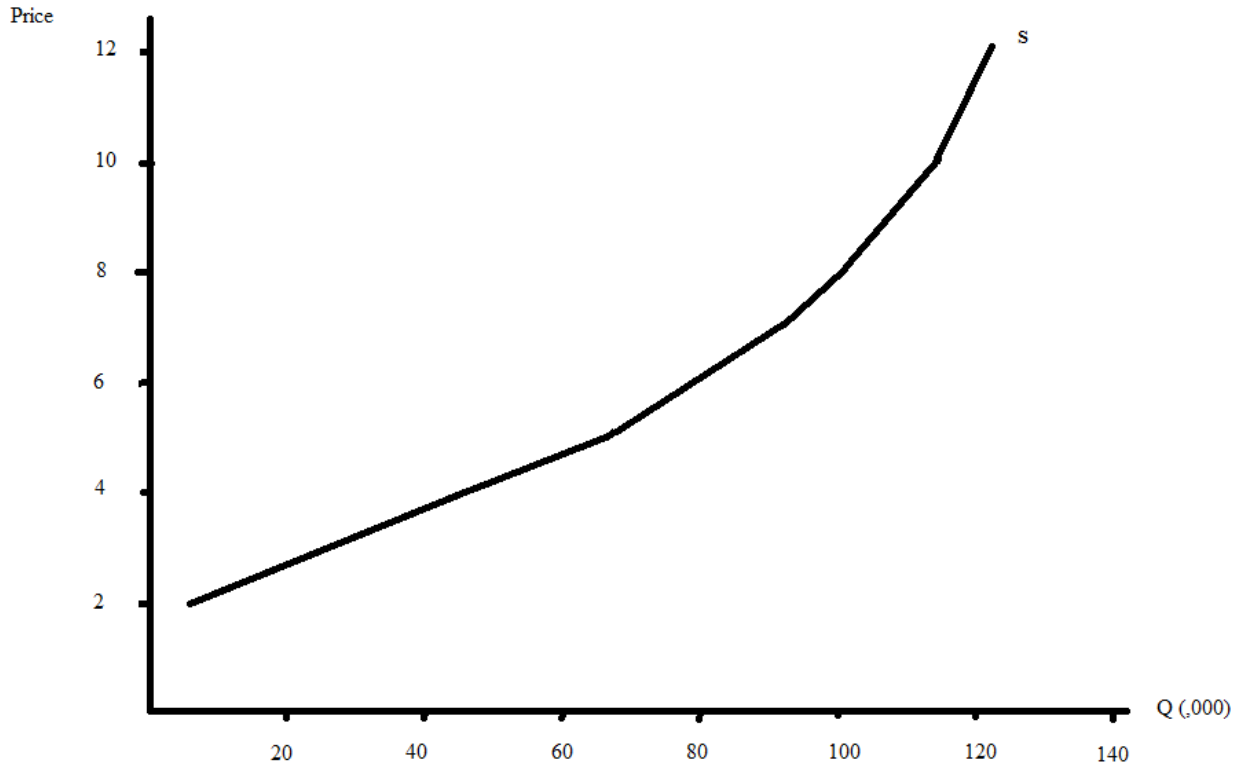
2. The Supply Curve

Like the demand curve, the supply curve represents the quantities that firms are *willing* to supply at various prices, all else being held constant. One possible supply curve for the pizza market discussed in Module 8 is described in Table 3 and Figure 3, below.

Table 3 The Supply Curve

Supply of Pizzas	
Price/Unit	Quantity Supplied
$2.00	5,000
4.00	46,000
5.00	66,000
6.00	77,000
7.00	92,000
8.00	100,000
10.00	115,000
12.00	122,000

Figure 3. Supply Curve



Shifts in the supply curve arise when one of the factors that affects supply, other than price, changes. For example, if there is an increase in the price of pizza flour, firms will require an increase in price in order to remain profitable at any particular quantity and we would expect to see an upward shift in the supply curve that was depicted in Figure 3.

3. Elasticity of Supply

Just as the price elasticity of demand is determined by dividing the percentage increase in quantity demanded by the percentage increase in price; the price elasticity of supply is determined by dividing the percentage increase in the quantity supplied by the percentage increase in price. Unlike the price elasticity of demand, however, the price elasticity of supply is normally assumed to be positive. (An increase in price is normally associated with an increase in quantity supplied.)

4. Slope of the Supply Curve

Upward-sloping supply curve

In Figure 3, the supply curve has been depicted as having an upward slope – firms in the pizza "industry" will only supply more pizzas if they are offered a higher price. Supply curves are upward sloping like this if the cost of production increases as output increases.

For example, if there is an increase in demand for pizzas, we would expect individual firms to increase supply by increasing the average size of their pizza parlours, and by building additional parlours. Also, additional firms may be drawn into the industry. The effect is that there will be an increase in demand for land (and buildings) to be used for pizza parlours. That land will have to be "bid" away from other users – other fast food producers, gas stations, furniture stores, apartment buildings, etc.

The first businesses that pizza parlours will bid land away from are the businesses that produce the least valuable products and, therefore, can bid the lowest prices for land. So, initially, pizza parlours will get land at a relatively low price and, therefore, will be able to produce pizzas at a low price.

But, as the pizza industry grows, pizza parlours will have to bid land away from businesses that are more and more profitable, so the price they will have to pay for land may rise as the industry expands. Similar effects may apply when pizza parlours attempt to hire more labour or buy more flour and cheese for their products.

In this case, the result is that the greater is the supply of pizzas, the higher will be the cost of producing pizzas and the price pizza parlours will have to charge for their products will rise – an upward-sloping supply curve.

Downward-sloping supply curve

It is, however, possible to have a downward-sloping supply curve. Assume, for example, that at first the pizza industry is quite small and pizza firms have to import products like pepperoni and mozzarella from Italy. The cost of these products would be very high. As the industry expands, however, local producers of these products will become established nearer to the firms using them, thereby reducing the costs of production, and driving down the price at which pizza parlours are willing to sell their product.

Also, as production of inputs like specialty cheeses increases, the firms producing those products will be able to build larger production facilities. As those facilities might experience *economies of scale* – for example, larger facilities may be able to use cost-saving techniques of production, like assembly lines - the cost of production may, again, fall as output increases, and the supply curve will, again be downward-sloping.

MODULE 10. PRICE DETERMINATION – "THE PARROT SPEAKS"

Module 8 joked that you could make a parrot into an economist by teaching it to say "supply and demand." To see how supply and demand work to allocate scarce resources among competing ends, this module first describes how they act to set prices.

1. Determination of an Equilibrium Price

Table 4, on the following page, reproduces the supply and demand information from the preceding two modules, for the case in which average income was assumed to be $45,000. That information has also been plotted in Figure 4.

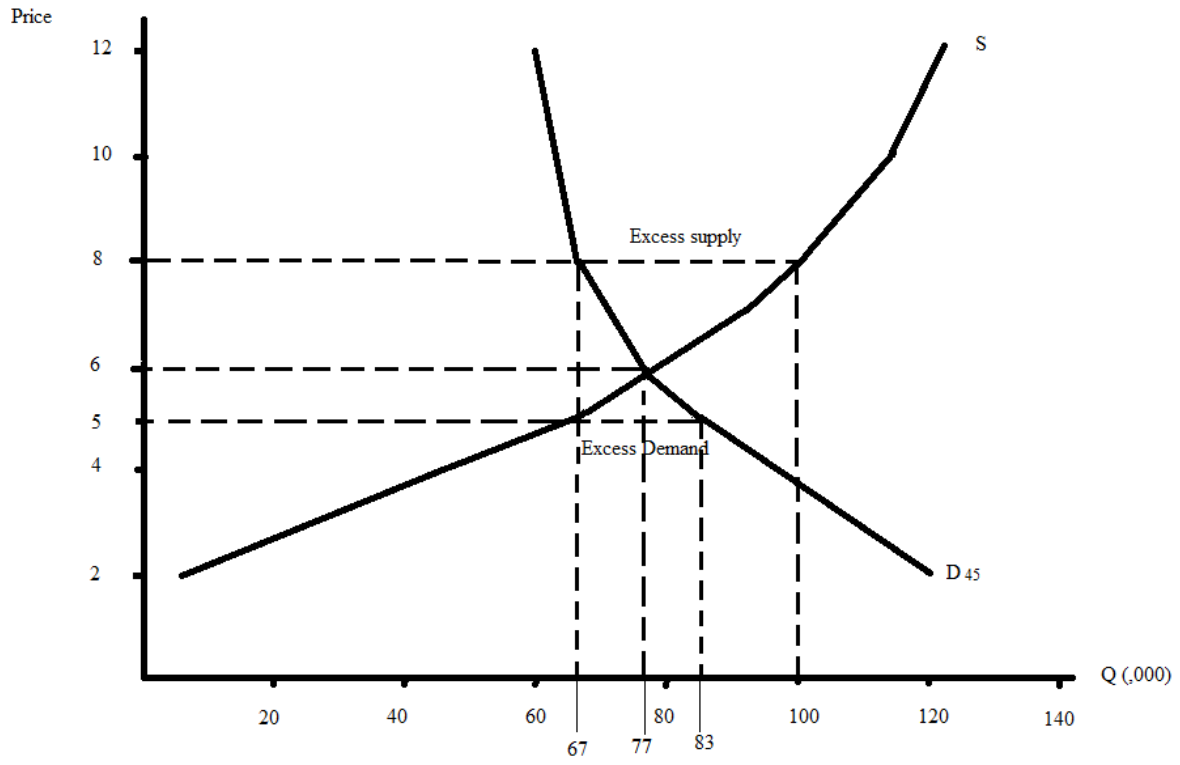
Assume, initially, that the market price is $8.00. What the demand curve indicates is that, at that price (and at an average income of $25,000), consumers would *like* to purchase 67,000 pizzas. And the supply curve indicates that, at $8.00 per pizza, pizza firms are *willing* to supply as many as 100,000 pizzas. There will be an **excess supply** of 33,000 (= 100,000 – 67,000) pizzas. That is, at a price of $8.00, the number of pizzas that firms are willing to supply will exceed the number that consumers are willing to purchase.

This excess supply cannot be maintained. The reason is that, as the price, $8.00, is higher than the cost of production, $5.00, it can be expected that there will be at least one firm that will realise that it can sell its excess supply by reducing its price slightly. A firm that offers to sell pizzas at $7.00, for example, will find that it will be able to sell 73,000 pizzas (as that is the quantity demanded at $7.00). Of course, most other firms will soon follow suit, in order to protect their market. But some firms will not find it profitable to sell pizzas at $7.00 and others will only find it profitable to sell fewer than before. So, supply will fall, in this case to 92,000.

But there is still excess supply, now of 19,000 (= 92,000 – 73,000). Again, at least one firm will reduce its price, to try and sell its excess pizzas. Assume that it reduces its price to $5.00 and that all other firms follow. Now firms will only be willing to supply 66,000 pizzas whereas consumers will want to buy 83,000. Pizza restaurants will find that if they continue to charge $5.00, they will run out of pizzas – there will be **excess demand,** of 17,000 pizzas (= 83,000 – 66,000).

Now, some firms will realise that they will still be able to sell their pizzas even if they raise their prices. Assume that they charge $6.00 per pizza and that all other firms follow suit. At that price, the number of pizzas that firms will be willing to supply equals the number that consumers wish to purchase, 77,000. At that price, we say that the **market is in equilibrium**. Because firms are able to sell all of the pizzas they want to sell, they will have no incentive to lower their prices. And because consumers can purchase all the pizzas they want, they will have no incentive to offer higher prices. That is, there will be no pressure from either "side" of the market (supply or demand) to change the price. The resulting price, which you can see occurs at the intersection of the supply and demand curves, is called the **equilibrium price**. (The market is said to "clear" at this price.)

Figure 4. Determination of the Equilibrium Price



2. Response to a Change in Demand

What happens now if demand increases? For example, what would happen if there was an increase in average income, causing consumers to wish to purchase more pizzas at each price level? Figure 5 uses the information contained in Module 8 to add the new demand curve, D_{55}, to the curves in Figure 4.

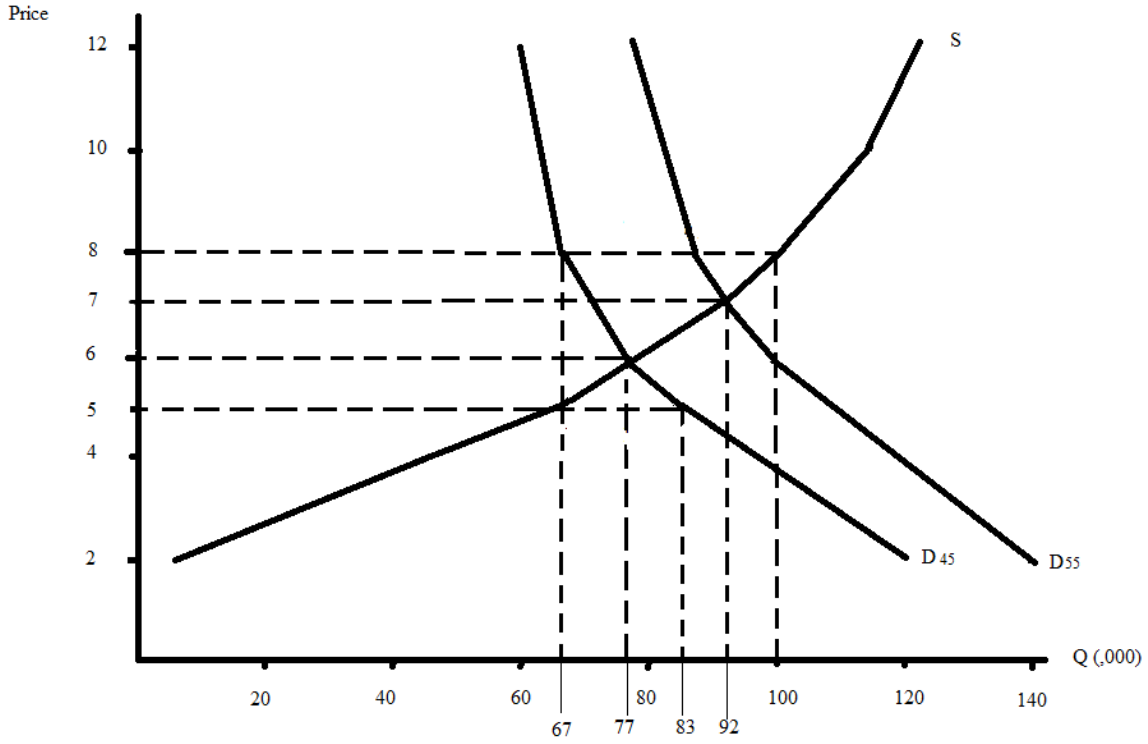
What will probably happen, initially, is that when their incomes increase, consumers will try to buy more pizzas – here, 100,000 instead of the "old' equilibrium quantity of 77,000 - at the existing price of $6.00. This means that there will be excess demand of 23,000 pizzas (= 100,000 – 77,000). But firms are not willing to supply that many pizzas at $6.00. First, if they *can* purchase additional inputs, the supply curve tells us that those additional inputs will cost $8.00, not $6.00. So, if firms sell additional pizzas at $6.00, they will lose money on them. Second, in the short term, pizza parlours may simply be *unable* to supply more pizzas because that requires increasing the size and number of pizza parlours, which could take a year or more.

What we expect will happen is that (i) pizza parlours will see the need (and the opportunity) to increase their prices and to supply more pizzas, and (ii) new firms will enter the market.

Consumers, at the same time, will be willing to pay those higher prices (in return for the offer of additional pizzas).

The new equilibrium will occur at $7.00 and 92,000 pizzas.

Figure 5. Effect of a Shift in Demand



3. Summary: The Net Effect of an Increase in Demand

When demand for pizzas increased, what happened to supply was the following:

(i) At the initial price ($6.00) consumers expressed a willingness to purchase more pizzas – more than firms were willing to supply at that price – so there was excess demand

(ii) At the same time, however, consumers also revealed that they were willing to pay a higher price for pizza.

(iii) Firms responded by raising the price.

(iv) At the new price, they were willing to supply more pizzas; and consumers bought more (but less than they had demanded at $6.00).

(v) In short, the increase in demand brought about an increase in supply. Consumers expressed a desire to consume more pizzas and producers satisfied that desire.

Who got the additional pizzas?

(i) When the price of pizza increases, it is those consumers who had the strongest preference for pizza who would be most willing to pay the new price. So, the first group to "succeed" is those who place the highest value on pizzas.

(ii) Also, when prices increase, it will often be the consumers with the highest incomes who will be willing to maintain, or increase, their consumption. So, the second group to "succeed" will be those with the highest incomes. If we ignore those who inherited their wealth, the people with the highest incomes will usually be those who have spent the most time and effort obtaining job skills (either by acquiring formal education or by investing in "on-the-job" training) and working to use those skills in paid employment. Whether this is how society wants to allocate resources (here, pizza) is a more complicated and contentious issue than can be resolved here. However, Module 12, Price Ceilings, will contrast this method of dividing up scarce resources with other techniques.

(iii) Finally, it has to be recognised that an increase in overall demand for pizza may arise even though demand by a sub-set of consumers has not increased. The subsequent increase in price will leave those consumers worse off. When talking about pizza, this may not seem like an important issue. But if the product is a necessity, like housing, or a staple food, like rice, flour, or eggs, it can produce a political outcry that is difficult for the government to ignore. One possible government response to such an outcry will be discussed in Module 12.

MODULE 11. APPLICATION - PRICES IN THE HEADLINES

We often see headlines like "Price surge reduces ability to buy housing (or gasoline, or running shoes, or whatever.)" The implication is that because the price of a product has increased (surged), consumers *will not* be able to afford as much of that product as they had before the price increase.

This statement is probably correct *if* the source of the price increase was a reduction in supply. But if the increase in price arose from an increase in demand (that is, a shift upward or leftward in the demand curve) the statement is, in most cases, wrong.

As was seen in Figure 5, although an increase in demand will induce an increase in the equilibrium price, it will also *increase quantity* supplied and consumers *will* be able to afford as much as, or more than, before. What has happened is not that the higher prices have meant that consumers can purchase *less*; it is that consumers have offered higher prices to producers in order to induce them to produce *more*.

In the example in Module 10, when consumers decided that they wanted to use some of their increased income (from $25,000 to $35,000) to buy more pizzas, they signaled that desire by revealing their willingness to pay higher prices, and suppliers responded by increasing the availability of pizzas. Consumers were able to buy *more* when prices increased, not less.

A caveat: Assume that an increase in demand for housing has arisen because there has been an increase in the number of people moving to a city. The supply and demand model predicts that the price of housing *and* the number of houses built will both increase, so housing firms will have provided the increased supply that consumers want. But, of course, some citizens who would have been able to afford houses before the influx of migrants are now unable to do so at the higher prices. So, for those individuals, the increase in price of housing has meant that they cannot afford as much housing as they did before.

In the latter case, the headline this module started with was *incorrect* in aggregate – the increase in price was associated with an increase in the total number of housing units available. But, for some individuals, the headline was *correct* – the increase in demand, followed by an increase in price, has meant that some people can afford fewer housing units than before. One possible response from governments is to impose "rent ceilings." (See Module 12: Price Ceilings.)]

There are two important lessons from the discussion here. First: maintain a healthy skepticism concerning newspaper reporting about economic factors – they often have an overly simplistic view of the world. Second: even if you start with a sound understanding of economic principles, you often find that the "real world" is a lot more complicated than you would have hoped.

MODULE 12. APPLICATION -

ALLOCATIVE MECHANISMS IN THE FACE OF PRICE CEILINGS

1. Price Ceilings

In some cases, even though a situation of excess demand has created upward pressure on prices, the price has become "stuck" at a level below equilibrium. This level is known as a *price ceiling*. There are many reasons that such a ceiling might arise

(i) Government regulation of the price of necessities: When prices increase, some consumers have difficulty paying for goods. If those goods are necessities – like housing, food, and cooking and heating fuel – and if the affected groups are large enough, pressure may be placed on the government to "do something." One such response is to place an upper limit on the price of certain defined goods. A common form of price ceiling in developed countries is that on rents; an even more common form, in developing countries, is a ceiling on the price of staple foods, like rice.

(ii) A publicly-announced price that proves to be too low and cannot easily be retracted: For example, a promoter may underestimate the demand for a concert and set a price at which demand exceeds supply. By the time it becomes clear that there will be excess demand, it may be too late to set a new price.

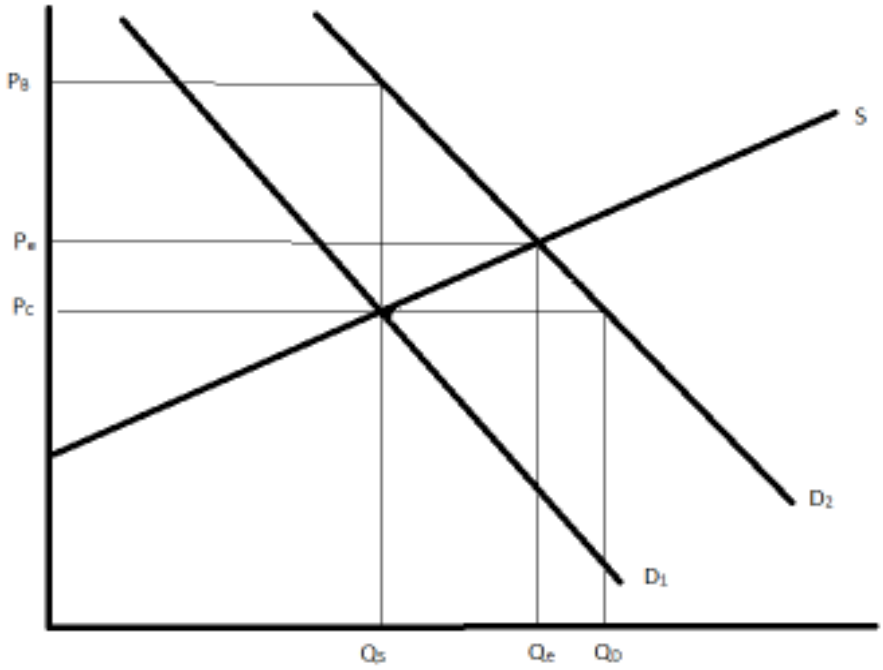
(iii) The price of one good may be set below its equilibrium level in order to encourage customers to buy a related product: For example, a football team may set the price of a "finals" ticket below the market level in order to encourage fans to purchase season tickets for the following season. And a rock group may set its concert tickets below the equilibrium level to encourage consumers to buy the group's songs on the internet.

The effect on supply and demand of a price ceiling can be seen using Figure 6. Assume that initially, with demand curve D_1, the equilibrium was at (Q_S ,P_C). Although an increase of demand to D_2 would normally have driven prices and quantity to (Q_e ,P_e), prices have been fixed at the initial price, P_C, (where P_C is a ceiling price). At that price, firms will only be willing to produce Q_S units, while consumers will want to buy Q_D. Two sources of inefficiency will arise.

First, consumers will receive fewer units of this product than they would have had the ceiling not been imposed – that is, Q_S will be less than Q_e. One of the main advantages of a market economy, that firms will respond to an increase in demand by increasing supply, has been thwarted.

Second, because the amount supplied at (Q_S, P_C) is less than the amount demanded, Q_D, some mechanism will have to be developed to allocate the scarce resource among competing users. Almost invariably these mechanisms will introduce further inefficiencies.

Figure 6. Effect of a Price Ceiling



2. Alternative Allocation Mechanisms

a) Queuing If the price of a good, say running shoes, has been fixed, when demand increases sellers might allocate the excess demand by requiring that buyers of running shoes queue (line up) for them, selling shoes at the fixed price on a first-come-first-served basis. In this case, what are the factors which determine who gets the scarce resource?

- strength of preferences – Those who have the strongest preferences for running shoes will be willing to queue for the longest times, *ceteris paribus* (everything else being equal).

- opportunity cost of time – Those who have the lowest value of time will be willing to queue the longest, *ceteris paribus*. So, everything else being equal, a queuing system will reward people for not working – the very young, the retired, and the unemployed.

b) Sell to regular customers only Sometimes retailers will give preference to their regular customers. In this system, one of the customer characteristics being rewarded is "length of time lived in the same neighbourhood" (i.e. people who have shopped at the same stores for a long time).

c) Discrimination If retailers are not able to raise their prices and they have less product to sell than customers demand, there is no cost to them of discriminating among customers. They might, for example, refuse to sell to young people, or blacks, or university students, or Seventh Day Adventists. The price controls would not have "made" retailers discriminatory, but the controls would have significantly reduced the cost to retailers of discriminating.

d) Reduce quality If producers are unable to obtain higher prices for a given quality of goods, they might respond to increased demand by reducing quality for the same price. Landlords, for example, might make fewer repairs, allowing their rental units to deteriorate. Thus, although renters have signaled, through the shift of the demand curve, a willingness pay higher prices to obtain more rental units of a given quality, what they will actually get is the same number of rental units at the same price but a lower quality.

e) "Side" deals: When a ceiling is placed on the price of rice, grocery stores might offer to sell rice at the regulated price *if* consumers purchase another over-priced product at the same time. When it is rents that are regulated, landlords often charge tenants exorbitant fees for their apartments' keys.

And if the price of new cars is regulated, but not the price of used cars, a car dealer might offer a "purchase-repurchase agreement," in which the buyer purchases a new car at the regulated price in return for a promise that he or she will sell the car back to the dealer at a set price a few months later. The dealer then sells the car on the unregulated used car market for an even higher price. (For example, the dealer might sell the new car for $20,000 then buy it back two months later for $22,000 and re-sell it for $28,000.)

f) Black markets: Producers of a price-regulated good might *sell directly on to the black market*. This occurs, for example, in markets for illegal goods, such as drugs. The black market supply curve is shifted upwards by the increase in cost created by:

(i) the reduction in scale of production (e.g. legal marijuana could be produced using techniques of mass production on large farms. Illegal marijuana is often produced in small quantities in hydroponic tanks in peoples' basements.);

(ii) costs of concealing transportation and retailing activity; and

(iii) the costs of bribes and fines.

Thus, the black market supply curve may intersect the demand curve at a higher price and lower quantity than would have been the case in a legal market. See Figure 7.

MODULE 13. APPLICATION - ALLOCATIVE MECHANISMS IN THE FACE OF PRICE "FLOORS"

When prices are not allowed to fall below a certain level, that level is referred to as a *price floor*. Examples of situations in which floors are set on the prices of products include:

- agricultural products
- minimum wages
- cartel prices

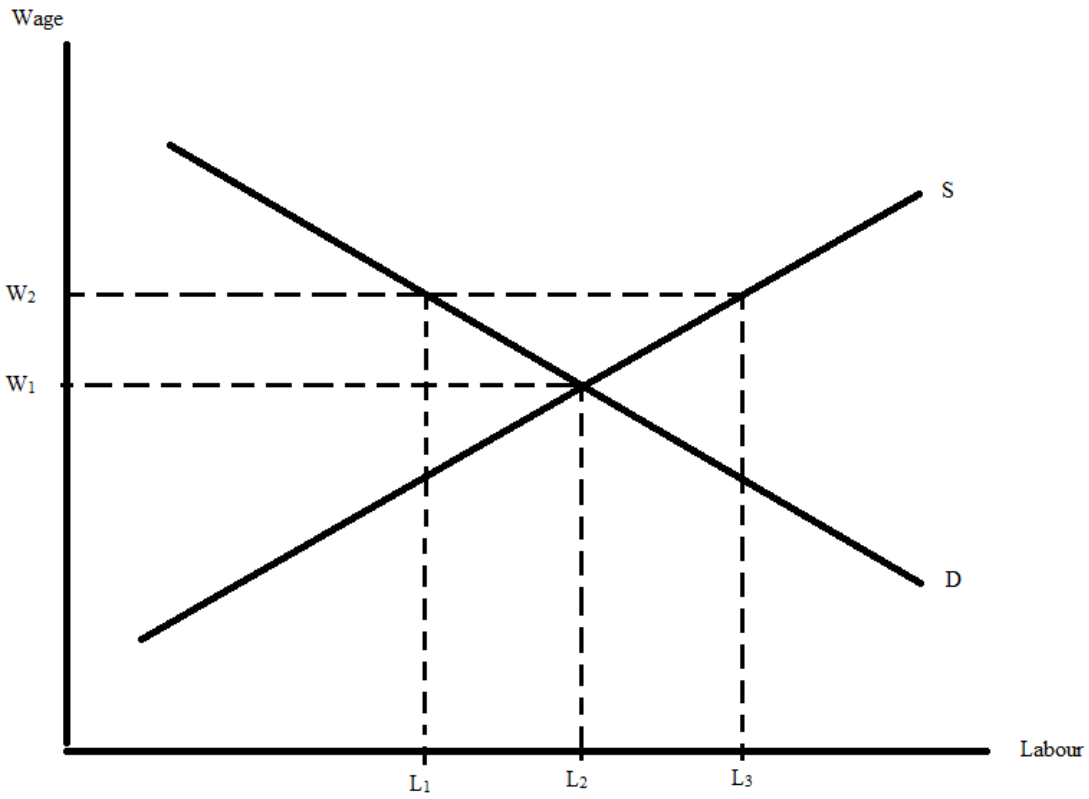
If the price floor is above the market-clearing price, excess supply is allocated among the scarce consumers using:

- quotas
- black markets

Minimum wage

Probably the best known, and certainly the most extensively investigated, price floor is the minimum wage. Assume, in Figure 7, that the average wage is W_1 per hour. Assume also that there has been no minimum wage. What effect would the imposition of a minimum wage equal to W_2 have on the allocation of resources (workers)?

Figure 7. Minimum Wage



• ***the level of employment*** - as the "price" of labour rises, from W_1 to W_2, the quantity of labour demanded will decrease. That is, the level of employment will decrease (as long as the demand curve for labour is downward sloping). In the diagram, L_2L_1 workers will lose their jobs.

• ***the rate of unemployment*** - as the "price" of labour rises, the number of workers employed will fall, hence the number of unemployed will rise. Also, as the minimum wage is higher than the equilibrium market wage, some individuals who would not previously have been willing to work will now become willing to supply themselves to the market. This creates the increase in supply from L_1 to L_3. When we add together the decrease in employment, L_2L_1, and the increase in the number of workers in the labour market, L_1L_3, it is seen that unemployment is now L_2L_3, (whereas it previously had been zero.)

• ***average annual incomes***- notice, in Figure 7, that when the minimum wage is introduced, L_2 workers manage to keep their jobs. As these workers now earn a higher wage than before (W_2 instead of W_1) they are better off, even if the workers who were laid off are now worse off. So, whether introduction of the minimum wage can be considered to be a "net gain" depends on whether we think that the gain to those who kept their jobs exceeds the loss to those who lose theirs.

An interesting alternative interpretation is this: Assume that the demand for labour is wage inelastic; that is, assume that the percentage increase in wages exceeds the (absolute value of the) percentage decrease in employment. For example, assume that when wages increase 25 percent, from $8.00 per hour to $10.00 per hour, employment decreases 10 percent, from 100 to 90 workers. In that case, total income per hour (added across all employed workers) will increase from $800 ($8.00 x 100) to $900 ($10.00 x 90). Now, assume that, because there is high turnover of workers in this industry, the 10 percent decrease in employment will be shared equally among all workers; that is, the average worker will be unemployed 10 percent of the time.

In this case, all workers will be employed 90 percent of the time, during which they will earn $10.00 per hour, and will be unemployed 10 percent of the time, during which they will earn nothing. But because percentage increase in the hourly wage was greater than the percentage decrease in the number of hours, workers' average annual income will increase. For example, workers who used to earn $16,000 per year working 2,000 hours at $8.00 will now earn $18,000 per year working 1,600 hours at $10.00 (and 400 hours at $0). In this case, it is possible that they will consider themselves to be "better off" with the minimum wage (although employers and consumers of the product being made by these workers will be worse off).

• ***discrimination*** - after the imposition of the minimum wage, there is an excess supply of workers equal to L_2L_3 - each employer has more job applicants than jobs. Thus, the employer can resort to discrimination among workers without any cost to him- or herself.

• ***training*** - if workers are relatively unproductive while they are being trained, employers may only be willing to provide training if workers will accept relatively low wages. Imposition of a minimum may prevent employers from lowering wages sufficiently to induce them to provide training.

[Minimum wages will be revisited in more detail in Module 22.]

MODULE 14. APPLICATION - THE LONG-TERM DECLINE OF AGRICULTURE

It seems that no matter how hard they work, farmers are doomed to face constant downward pressure on their profits. The result is that the number of farms has declined continuously over the last century or more. This section presents the argument that this phenomenon arises from a combination of three factors: that agricultural markets are competitive, that demand for agricultural products is income inelastic, and that output per worker (productivity) tends to rise at the same rate in agriculture as it does in other industries.

Assume that productivity increases at ten percent per year in both agriculture and manufacturing - and that the latter leads to an increase in income of each *manufacturing* worker of ten percent. It is usually observed that demand for agricultural products is *income inelastic* – that is, if income increases by x percent, consumers' demand for agricultural products will increase by less than x percent. This is because the *quantity* of food you can consume is limited, meaning that increases in expenditure have to come primarily from purchase increased *quality*.

Assume, for example, that each ten percent increase in the income of manufacturing workers results in a four percent increase in demand for agricultural products. If output per worker in agriculture increases by the same percent as it does in manufacturing (ten percent) but quantity demanded increases by only four percent, agricultural supply will exceed demand and the price of agricultural goods will decrease. If that decrease is sufficient, farm revenue will decrease.

For example, assume that the price per unit of agricultural products had previously been $5.00 and that the average farmer had produced 100,000 units of output, implying total revenue per farm of $500,000. But with an increase in output per farm (with the same number of inputs) of ten percent, to 110,000 units, price falls to $4.00. With an increase in farm productivity, total farm revenue falls by $60,000, to $440,000.

Furthermore, note that it was assumed that the increase in farm output arose from an increase in output per worker – there was no need for increased inputs (from, for example, land or fertilizer). In that case, the cost of production remained unchanged. As a result, the $60,000 reduction in farm revenue translated into a $60,000 reduction in farm profits. But, as agricultural firms are competitive, they were earning very low profits before the increase in output per worker, so a decrease in profits may mean they are now earning negative "profits."

Q.E.D. After an economy-wide increase in productivity, farm profits initially become negative. Some farmers will leave the industry until the quantity supplied will fall sufficiently that the equilibrium price of agricultural products rises to the level at which economic profits are just enough to stop farmers from leaving the industry.

How can farmers avoid this outcome? One way might be to increase their prices. But, if the market truly is competitive, when some farmers increase their prices while others do not, consumers will stop buying from the high price producers and switch to the low price farmers. This is why one of

the most common responses to declining farm profits is for farmers to ask the government to mandate a price floor on their products. For example, if, in the example above, the government was to set a floor of $4.75, farm revenue would rise to $522,500 (= $4.75 x 110,000).

D. MARKET ECONOMIES - COMPETITION

MODULE 15. COMPETITIVE INDUSTRIES - DEFINED

The real world is such a complicated place that, to solve even basic models of the economy requires the power of huge supercomputers. Nevertheless, we often find that we can obtain significant insights by creating simple, abstract models and then adding complications one at a time.

When physicists wanted to predict how fast an object would fall, for example, they began by imagining that the object was travelling through a vacuum. For many purposes, they found that the predictions from this simple model were "good enough" – for example, for predicting how fast a cannon ball would fall at sea level. In cases in which the simple model yielded predictions that were insufficiently accurate for their purposes, scientists then added complications to the model - such as air resistance, wind, and shape (of the object).

Similarly, when economists wanted to predict which goods would be produced in the economy, and in what quantities, they began with the simplified supply and demand model that was discussed in Part C of these notes. That model could then be adjusted to take account of additional factors. For example, the effect of changes in average income could be modeled by shifting the demand curve; and the effect of technological change by shifting the supply curve.

One of the factors that affects prices and quantities, but which was not discussed in Part C, is the level of competition among firms in the supplying industry. Economists generally distinguish among three types of industries, each of which can be expected to behave quite differently. These are *competitive industries*[4], in which there are hundreds of firms all producing roughly the same products; *monopolies*, in which there is only one firm supplying a product; and *imperfectly competitive industries*[5], in which there is only a small number of firms. It is competitive industries, which require the simplest models, that will be discussed in Part D. Monopolies and other imperfectly competitive industries will be discussed in Part E.

Characteristics of a Competitive Industry

An industry is said to be competitive if it meets the following criteria:

a. It is composed of *numerous, relatively small firms*. The classic example of a competitive industry is farming. Although, say, an individual potato farm might cost millions of dollars and extend over hundreds of acres, the amount produced by an individual farm is very small relative to the amount of potatoes produced in the world.

[4] In the following notes, I will use the term "competitive industry." Most economics textbooks refer to such industries as "perfectly competitive." I find this term to be confusing as "perfect" is not being used here in its normal sense, of "best possible" or "without fault"; rather it refers to an idealized model that abstracts from the complexities of the real world.

[5] Imperfectly competitive industries are often further divided, into *monopolistic competition* and o*ligopoly*, depending on the number of firms in each.

b. The industry is *easy to enter*. By "easy" is not meant that just anyone could set up a firm: for example, a grain farm in Canada can cost $5 million dollars. What is meant is that if the firms in an industry are profitable, there will be individuals who will be able to raise the amount of money necessary to start up a new firm. Indeed, in the last twenty years or so, the advent of large investment firms, hedge funds, and "angel" investors has meant that if an industry is sufficiently profitable, there will be people who will be able to raise billions of dollars to enter into competition with the existing firms. (See, for example, the airline industry.)
c. All firms *produce approximately the same product*. What this implies is that if a firm increases its prices relative to its competitors, it will lose its customers, as its competitors are producing the "same" product. This assumption clearly applies in the case of a standardized good, like wheat or graded potatoes. It is less clear for goods that are "approximately the same," like pizza restaurants, or corner grocery stores, or real estate lawyers; but for products that are very similar to one another, the competitive industry model may describe firms' behaviour sufficiently closely to make the model useful (like the falling cannon ball).
d. All economic actors are *equally well-informed*. By "actors" is meant the people making economic decisions. With respect to sellers, what the assumption implies is that all owners of businesses are equally well-informed about the techniques for making or growing their product, about the availability and prices of inputs, about access to consumers and the prices consumers are willing to pay, etc. With respect to buyers, the assumption mostly implies that consumers are aware of the prices that all sellers are charging. This assumption reinforces the three preceding assumptions: there can only be (a) "many competing firms" if it is (b) "easy for new firms" to enter the industry. And the latter requires that potential new entrants into the industry are "able to obtain information" about how to produce and market the product. Third, firms will only be able to (c) "produce the same product" if all firms have equal access to information about production methods.
e. There are only limited possibilities for obtaining "*economies of scale"*. By "scale" economists just mean the size of the firm – the amount of output it produces. A firm is said to experience economies of scale if, as its output level increases, the average cost per unit of output falls. (Economies of scale are discussed in more detail in Module 16.) If the firms in an industry did experience economies of scale, then any firm would be able to reduce its average costs, and therefore reduce its prices, simply by getting larger. The first firms to do so would be able to undercut the prices of the smaller firms and, therefore, drive the small firms out of business. Soon, there would only be a small number of large firms; but that would not be a competitive industry.
f. Every firm acts as if its goal was to *maximise its "economic profits"*. Two comments here: First, by "economic profits" is meant the firm's revenues (the amount of money its customers pay to it) minus its costs, where "costs" include the owner's opportunity costs (see Modules 3 and 9). For example, assume that a store has sales of $500,000 per year and *accounting* costs (for rent, wages, costs of goods sold, heat, lighting, etc.) of $470,000 per year. If the owner does not pay herself a salary, the firm's accounting profits will be reported to be $30,000. But if the owner could have earned $80,000 per year in the next best use of her time, say working as a teacher, the firm's *economic* profits would have been reported to be *minus* $50,000.

Second, notice that assumption (f) does not say that firms would maximise their profits; it says that they would act "as if" they were trying to maximise profits. This assumption implies that even if firms did not try to maximise profits, the forces of supply and demand would act in such a way that it is only those firms that adopted behaviours that lead to maximum profit that would survive. In this way, the "evolution" of a competitive market is analogous to the evolution of a biological species: even though no individual member of a species actively chooses the characteristics that will maximise its reproductive success, the "survival of the fittest" will ensure that it is those characteristics that will survive. Similarly, it will be argued that in a competitive market it is only firms that act in such a way as to maximise economic profits that will survive.

MODULE 16. ECONOMIES AND DISECONOMIES OF SCALE

If the firm's average costs per unit of output decrease as output increases, we say that the firm is experiencing *economies of scale* ("scale" just means the level of output); if average costs increase as output increases, the firm is experiencing *diseconomies of scale*; and if average costs remain constant the firm is experiencing *constant returns to scale*. As these factors will be very important in the construction of any industry's supply curve, in this module, the sources of these cost relationships will be investigated both at the level of the firm and at the level of an entire industry.

1. At the Level of the Firm

1.1 Economies of scale

Three factors are usually identified as potential sources of economies of scale (average costs decrease as output increases):

a. *Specialisation and division of labour*. Instead of one worker doing all of the tasks, different workers specialise in doing specific tasks. In automobile factories, it is not common for one worker to follow a single vehicle down the assembly-line, installing each part of the engine and body. Rather, as the vehicle moves down the line, it passes various workers, each of whom has his or her tools ready at hand and each of whom has invested in learning a specific job at a high level of productivity.
b. Use of more *efficient techniques*. There is no point buying a combine to harvest the wheat from a 10-acre plot. But when you expand to 1000 acres, a combine becomes worthwhile. Similarly, an assembly line may not make sense in a firm that is producing 100 cars per year, but may become efficient when the firm starts to produce 10,000 cars per year. These factors are sometimes referred to as *indivisibilities*, as the efficient production method – here use of a combine harvester or an assembly line – cannot be divided up when you want to produce fewer units of output. The concept of indivisibilities also applies to advertising – a TV ad for a single restaurant may cost the same as an ad for a chain of restaurants – because TV time cannot meaningfully be divided up.
c. Lower costs due to *volume buying*. A restaurant buying 100 pounds of beef per month may have to pay more *per pound* for that beef than a restaurant buying 1,000 pounds.

1.2 Diseconomies of Scale

It is often argued that firms should be able to avoid diseconomies of scale simply by duplicating themselves. For example, if a restaurant's cost-minimising output level is 200 meals per day, the owners might be able to maintain the low costs associated with that size by building additional 200-customer restaurants. Then, no matter how many customers it served per day, its average costs would never rise. The question, then, is: How might duplicating buildings or factories result in *higher* average costs (diseconomies of scale)?

a. Difficulty *managing increasing numbers of employees.* As the number of buildings/factories increases, so does the number of employees. With that increase, it may

become more and more difficult for managers to monitor employees to make sure that they are working as hard and effectively as possible.

b. Declining *quality of information*. The larger is the "pyramid" of employees, the greater is the number of levels through which information will have to pass before it reaches the managers at the top. As information passes through more hands, it may become less reliable.
c. *Ability of managers* to absorb the amount of information generated. As the size of the firm increases, so does the total amount of information that must be absorbed by the decision-makers at the top. This may lead them to make less efficient decisions, the larger is the company.
d. Declining *availability of resources* – if a firm (or industry) becomes large enough, it may begin to use a significant percentage of the available resources. If those resources begin to run out, their prices may increase. For example, a firm that has been relying on a local, abundant source of electricity may expand to the point where it requires more electricity than can be produced locally. At that point, it may have to bring electricity from a distant, more expensive source. Or a company like McDonalds may sell so many hamburgers that it begins to put upward pressure on the domestic price of beef.

1.3 No economies of scale

A firm is said to be experiencing no economies of scale if its average costs per unit of output do not change when it increases the scale (i.e. its output) of its operation. Little needs to be said about this situation other than it occurs when none of the conditions for increasing or decreasing returns to scale are in effect.

2. At the Level of an Industry

Even if individual firms do not experience economies or diseconomies of scale, it is possible that an *industry* might do so. Imagine, for example, an industry that is composed of firms which each experiences constant returns to scale. When any one of these firms increases its output, its average costs will not increase. But if every firm in that industry was to increase output, average costs for the industry as a whole might either increase or decrease.

2.1 Economies of scale

When an industry expands, its firms will need to purchase additional goods and services from their suppliers. As the suppliers increase their scale to meet this increased demand, *they* may experience economies of scale and, therefore, decreased average costs. To the extent that these savings are passed on to the purchasers, the average costs of the latter will also decrease. Hence, the industry may experience economies of scale even if the individual firms do not.

For example, as was discussed in Module 9, when the pizza restaurant industry first developed in North America, many of the ingredients were imported from Europe and were, therefore, quite expensive. As the number of pizza parlors increased, however, local suppliers

became established, creating sufficient savings in transportation costs that the average cost of pizzas fell. Although no individual restaurant had been big enough to generate these savings, the industry as a whole did so. Initially, at least, the industry experienced declining average costs as the scale of the industry increased.

2.2 Diseconomies of Scale

Conversely, an individual firm might represent such a small percentage of the industry that if it expanded it would have very little impact on the overall demand for inputs such as labour, land, and buildings and therefore on the cost of those factors. However, if there was to be a significant increase in the total *number* of firms, there might soon be a shortage of inputs and their price would be driven up. Hence, whereas the individual firm did not experience diseconomies, the industry might.

This effect is most easily seen with respect to farm land. When demand for a crop, such as quinoa or kale, first arises, farmers who raise that crop may be able to "bid" the required land from relatively low-valued uses. But as demand for the new crop increases, land will have to be taken from increasingly valuable alternatives and the price of land will rise, thereby increasing the average cost of production. Again, although the amount of land used by any individual farm did not increase sufficiently to affect the price of land, the amount demanded by the industry increased enough to drive up the costs faced by all of its members. The industry experienced diseconomies of scale even when individual farms did not.

MODULE 17. COMPETITION - SUPPLY AND DEMAND REVISITED

In this module, it will be seen that the model of supply and demand that was developed in Part C was actually a simplified description of a competitive market. Section 1 will comment briefly on the form of the demand curve for a single *firm* in a competitive industry. In Section 2, the supply curve for a competitive *industry* will be derived from the average cost curves that were described in Module 16. Section 3 will analyse the determination of price and quantity in such an industry.

1. The Demand Curve

The demand curve for the product of a *single competitive firm* is a horizontal line at the industry price for that product. That is, the firm will be able to sell any amount that it wishes at the industry price.

To see why, consider an industry with 10,000 farms, each producing 100,000 pounds of an agricultural product (say, potatoes), implying a total output of one billion pounds. Holding everything else constant, if one farm doubles its output to 200,000 pounds, total output for the industry will increase by 100,000 pounds, that is, from 1.0000 billion pounds to 1.0001 billion. It is not likely that such an infinitesimally small increase will have a measurable effect on the market price of this product. As any other plausible increase in a single farm's output – say, even tripling or quadrupling its output – will also have no noticeable impact on market price, economists assume that the demand curve for the *firm's* product can be treated as if it is horizontal, even if the demand curve for the *industry* is downward-sloping.

2. The Supply Curve

The determination of a *competitive industry's* supply curve proceeds in two steps. First, the relationship between average costs and output is derived for the individual firm. Second, from the firms' demand curves and average cost curves, it is possible to calculate the long-run relationship between price and quantity supplied for the industry as a whole.

2.1 The competitive firm's average cost curve

The industry's supply decision is determined in part by how individual firms' average costs (i.e. costs per unit of output) vary with the amount that they produce. First, at low levels of output, most competitive firms experience economies of scale - as output increases, firms are able to reduce their costs by using more efficient production techniques or by employing specialization and division of labour. For example, as grain farms increase in size, they can introduce larger and more technologically advanced machinery for tilling, fertilizing, and harvesting; and as coffee shops increase in size, they can invest in more sophisticated espresso machines. Thus, mid-sized firms will have lower average costs than will small firms.

Second, at some output level, there will be no further opportunities for economies of scale – average costs will reach their minimum point. For example, when harvesting machines reach a

certain size, it may not be possible to build larger ones that operate at lower costs; and high-output espresso makers may not operate at lower average costs than mid-sized ones.

Once the firm has reached the size where it is producing at the lowest possible average cost, it may be able to maintain that cost level as it expands output simply by reproducing its production facilities. For example, if the average cost per bushel of wheat can be minimised with a 1,000 acre farm, a wheat producer might be able to maintain that same cost by operating multiple farms, each of 1,000 acres. And a coffee shop that minimises costs by operating an outlet that serves 100 customers per hour, may be able to maintain that same cost level by establishing more 100-customer outlets. Over a "middle" range of output, therefore, average costs will be approximately constant.

[Note that if this is the case, it may be possible to have an industry in which relatively small firms – for example, independent, 100-customer coffee shops – can compete on a level basis with much larger firms – for example, Starbucks.]

Finally, even if efficiently-sized facilities are reproduced exactly, firms may encounter diseconomies of scale and average costs may rise as the firm expands output beyond some point– for example, because it becomes more difficult to manage the increasing numbers of employees. This appears to be the case with respect to agriculture, where costs are lower at small scale than at large scale, where employee-operators have to be hired. Over this "higher" range of output, therefore, average costs may rise with output.

To summarise, the competitive firm's average cost curve is usually assumed to be "U-shaped" with respect to output, as depicted in Figure 8. There, curve AC represents average cost at each output level, $5.00 is the minimum average cost that can be obtained, and 100,000 units is the maximum output that can be produced before average costs rise above $5.00.

2.2 The competitive industry's supply curve

When price determination was first discussed, in Modules 9 and 10, it was assumed that a supply curve answered the question: "what is the *maximum amount* that firms would be willing to supply at any given price?" Alternatively, however, those same curves could be thought of as answering the question: "what is the *minimum price* firms would be willing to accept in order to induce them to supply a given amount of output?" It will be convenient to adopt the latter interpretation when deriving the competitive *industry's* supply curve. In what follows, it will be argued first that the supply curve will consist solely of combinations at which price equals minimum average cost; and, second, that minimum average cost is a function of total output.

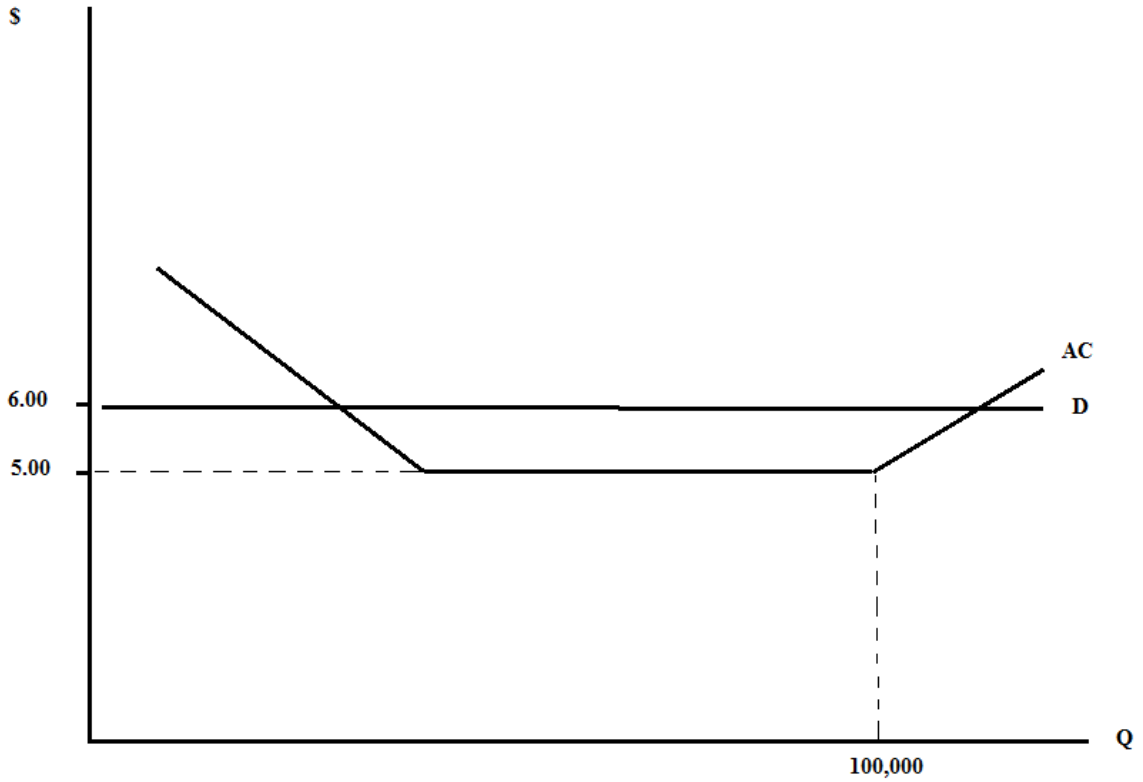
In the long run, price equals minimum average cost

The first step in deriving the *industry* supply curve is to add a demand curve to the *firm's* average cost curve in Figure 8. (Remember that demand for an individual firm is horizontal at the industry price.) For example, assume that the market price is, initially, $6.00 per unit, that there are 8,000

firms in this industry, and that each of them produces 100,000 units of output - for a total of 800 million units. As average cost is $5.00 at this level of output, each firm will earn economic profit of $1.00 per unit ($6.00 minus $5.00) and earn total economic profit of $80,000 (80,000 units at $1.00 per unit).

It might be tempting to assume that if 800 million units are offered at $6.00 per unit, that is one point on the industry's supply curve. But that is not correct. Remember that one of the components of average cost is the firm's opportunity cost, most of which is the profit that it could make in its next best use. So, if the firms described in Figure 8 receive more than $5.00 per unit, they will earn more than they could in other industries. And, conversely, they will have earned enough to attract investors from other industries (where firms are earning lower profits).

Figure 8 Competitive Firm's Average Cost Curve



This means that $6.00 is not a "stable" or "equilibrium" industry price for 800 million units. At $6.00, additional firms will be attracted into this industry, increasing supply and placing downward pressure on prices. The market will only stabilize when the price has fallen to equal average cost. And, as the demand curve is horizontal, this can only occur when price equals minimum average cost.

[When the demand curve is above any portion of the average cost curve, firms that operate at an average cost below the market price will earn economic profits, so new firms will be attracted, and

those firms will choose to operate at a scale that allows them to incur average costs below the market price. This will continue to occur until price equals average cost at the minimum average cost, here of $5.00.]

To summarise:

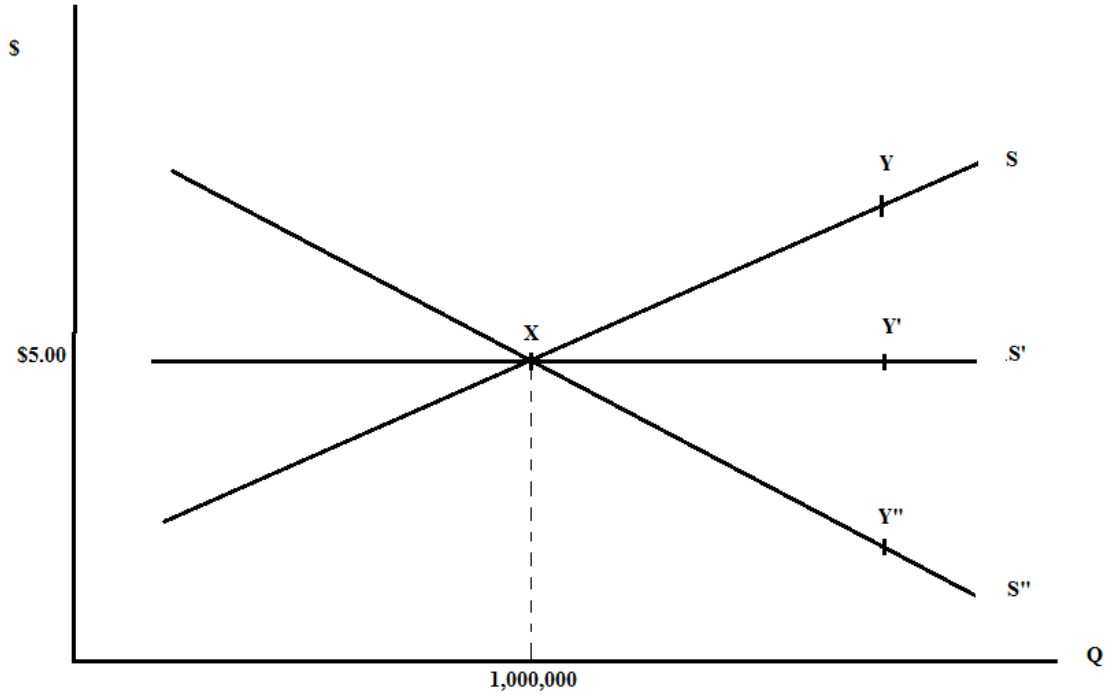
> ***A price/quantity combination at which price exceeds average cost cannot be a point on the industry's long run supply curve as, at such a combination, output will increase and price decrease. Hence, the supply curve is composed solely of combinations at which price equals minimum average cost.***

Minimum average cost is a function of total industry output

If price equals minimum average cost at each point on the industry supply curve, what does that curve look like? To answer this, consider Figure 9, in which it has been assumed that, after the entry of new firms described above, there are 10,000 firms producing 100,000 units each, at an average cost of $5.00. As price equals average cost, economic profits are zero and there will be no incentive for new firms to enter the industry. Hence, the combination of: price equals $5.00, and output equals one billion units, which has been plotted as X in Figure 9, is a point on the industry supply curve.

Now, assume that demand for this product increases. Initially, supply will be insufficient to meet demand and prices will rise above average cost: firms will earn economic profits. New firms will be attracted into the industry and output will increase until prices equal average costs again. Whether the new equilibrium will occur at prices higher or lower than, or the same as, $5.00 depends on whether the *industry* experiences economies, diseconomies, or no economies of scale.

Figure 9. Industry Supply Curve



Industry experiences diseconomies of scale
As the industry expands, it will need to purchase additional inputs: workers, machines, buildings, energy, land, etc. If there is a shortage of any of these factors, increased demand will drive up their prices, causing each firm's average costs to rise. Hence, in this case, as the industry expands to meet the increased demand, each firm's U-shaped average cost curve will shift upward and the point at which price will equal minimum average cost will be higher than it was initially. The result is that, as output increases, the minimum price at which firms will offer their product will also increase: the supply curve will be upward sloping. In Figure 9, the new price/quantity combination has been plotted as point Y – on a supply curve that looks like S.

For example, as the amount of land used by grain farmers is such a very large percentage of the total amount of land available, an increase in the size of the grain-growing industry can be expected to cause an increase in land prices. Hence, the minimum average cost of producing grain will increase and the supply curve (for the industry as a whole) will be *upward* sloping. In such an industry, an increase in demand will result in an increase in both quantity supplied and market price – the industry supply curve will be upward sloping.

Industry experiences no economies of scale
It is also possible that as industry output increases, the costs of inputs will remain relatively unaffected, in which case the minimum price at which firms will offer their product will be

constant: the supply curve will be *horizontal*. The new price/quantity combination has been plotted as Y' in Figure 9 – and the supply curve as S'.

For example, the amount of gardening tools, fertiliser, and workers used by landscaping firms may be such a small portion of the total amounts used by the economy as a whole that an increase in the supply of landscaping services will have little effect on the costs of those inputs. In such an industry, an increase in demand will result in an increase in quantity supplied but no increase in price (once additional firms had entered the industry).

Industry experiences economies of scale
Finally, it is also possible that, as demand for the inputs into an industry increases, the firms supplying those inputs will experience economies of scale, allowing them to offer their products at lower costs. The average costs of the industry purchasing those inputs would decrease, allowing firms to supply output at a reduced cost: the supply curve will be *downward* sloping. The new price/quantity combination has been plotted as Y" in Figure 9 – and the supply curve as S".

Such a supply curve might well result when a new product is brought to market. As consumers come to accept that product, and the size of the market increases, the producers of inputs into that product may experience economies of scale that allow them to reduce their costs, thereby reducing the average costs of the purchasing firms. In this case, an increase in demand will result in an increase in quantity, but a *decrease* in price.

4. Summary

In Module 9 it was *asserted* that supply curves might be upward-sloping, horizontal, or downward-sloping, and that equilibrium in the market would occur where supply equaled demand. In this module, it has been seen that the *theory* underlying the three possible supply curves discussed in Module 9 was derived assuming that the market was competitive. It will be seen later in these notes that when markets are not competitive, the supply/demand model does not apply.

MODULE 18. APPLICATION - INSURANCE AS A COMPETITIVE INDUSTRY

The insurance industry exhibits most of the characteristics of a competitive industry, (as defined in Module 15). There is a large number of relatively small firms, they all produce a fairly standardized product, the information necessary to operate an insurance firm is widely available, the opportunities for obtaining economies of scale are limited, and it is easy to enter the industry.

With respect to the latter, firms in one sector of the insurance industry, such as property insurance, can often move quickly into other sectors, such as automobile insurance, if profits increase in the latter. And firms in related industries, such as credit cards, banks, and automobile manufacturing, often have sufficient funds and expertise to establish new insurance companies.

Hence, we would expect that all firms will be induced to offer their products at the lowest average cost and that the industry as a whole will operate at a very low, or zero, economic profit.

With respect to these characteristics, insurance is like many other competitive industries. What makes it different from other such industries is the nature of its costs, the valuation of its benefits, and the difficulties it faces managing its customers' actions.

1. Measuring the Costs of Insurance

The "price" of an insurance policy is called its *premium*; and the cost of "producing" that policy is the value of payments that the insurer expects to make to the policyholder. (There are other costs, such as the cost of employing salespeople and accountants, but as these are a small percentage of total costs they will, for simplicity, be ignored initially.) To the extent that insurance is a competitive industry, premiums will be driven down to equal the average cost of the benefit payments. What will be considered here is how that value is calculated.

A simple numerical example will help to answer this question. Assume: that it is automobiles that are being insured, that the probability that the average driver will have an accident in any year is 5 percent (i.e. one accident every twenty years), and that the average cost of an accident is $10,000. If insurance companies promise to compensate their policyholders for their accident costs, what is the cost to the insurance company of each policy?

To answer this, assume that a typical company insures 20,000 drivers. That number is large enough that, if the average driver has one accident every twenty years, then in one year 1,000 policyholders will make a claim against their insurance policy (and 19,000 will make no claim). Thus, the total annual cost of claims will be:

$$(1{,}000 \times \$10{,}000) + (19{,}000 \times \$0) = \$1{,}000{,}000$$

And the average annual cost per driver will be:

$$\frac{(1{,}000 \times \$10{,}000) + (19{,}000 \times \$0)}{20{,}000} = \frac{\$1{,}000{,}000}{20{,}000} = \$500$$

This can be rewritten:

$$\frac{1{,}000}{20{,}000} \times \$10{,}000 + \frac{19{,}000}{20{,}000} \times \$0 = \frac{\$1{,}000{,}000}{20{,}000} = \$500$$

or

$$(0.05 \times \$10{,}000) + (0.95 \times \$0) = \$500$$

That is, the average annual cost of insuring one driver can be found by multiplying the probability of each possible outcome by the value of that outcome, and then adding together all of the resulting figures. (The "possible outcomes" in the example above were $10,000 and $0, and their respective probabilities were 0.05 and 0.95.)

As a more realistic example, assume that there are four possible outcomes: the driver has no accident ($0 claim); a minor accident, causing $500 damage; a serious accident, causing $5,000 damage; or a major accident, causing $50,000 in damages. Assume, further, that the probabilities of these four outcomes are 76 percent, 20 percent, 3 percent, and 1 percent, respectively. [Note: the probabilities have to add to 100 percent, which is the probability that "something" will happen.] In this case, the average annual cost per insured driver will be

$$(0.76 \times \$0) + (0.20 \times \$500) + (0.03 \times \$5{,}000) + (0.01 \times \$50{,}000) = \$750$$

In both of these cases, competition among insurance companies will drive the price of insurance down to the relevant average cost of claims - $500 in the first case and $750 in the second – plus administrative costs. If the latter were, say, $75 per policy, the equilibrium price for a policy would be $575 in the first case and $825 in the second. Whether automobile drivers would be willing to pay these amounts depends on the value that they place on insurance.

2. Measuring the Value of Insurance

For simplicity, assume again that there is a 5 percent probability per year that a typical driver will have an accident costing $10,000, and a 95 percent probability of no accident. What will the value to that driver be of an insurance policy that pays all of the driver's costs whenever an accident occurs? That is, what is the maximum premium the driver would pay for such a policy?

The answer to this question is complicated by the fact that the driver does not know with certainty what the payout will be from the policy. Nineteen times out of twenty it will be $0, and in one time out of twenty it will be $10,000. Pessimistic drivers might look at these odds and say "with my luck, I'll be the one in twenty drivers who has an accident this year costing $10,000." He or she might be willing to pay up to $10,000 for insurance against that possibility. Whereas extremely optimistic drivers might say something like "bad things never happen to me, so I'll be one of the

nineteen in twenty who doesn't have an accident." Those drivers may not be willing to pay much more than $0 for an insurance policy.

As you might imagine, economists do not find this way of thinking about drivers' attitudes towards insurance to be very satisfactory. It is too imprecise. Accordingly, economists, and statisticians, have developed a concept that provides a slightly more formalized approach, called *expected value*, to think about the value of uncertain events.

The simplest way to think about the expected value of an automobile insurance policy is to recognise that purchasing insurance is analogous to betting in a casino. Buying a policy that pays $10,000 five percent of the time (and nothing 95 percent of the time) is like betting on a single spin of a roulette wheel that has twenty slots, one of which pays $10,000, and nineteen that pay nothing. How much would you pay to spin the wheel once?

To answer this, economists (and statisticians), first ask how much you would earn on average if you could spin the wheel a very large number of times – say 20,000. The answer, as we saw above, is that the average payout would be

$$\frac{(1{,}000 \times \$10{,}000) + (19{,}000 \times \$0)}{20{,}000} = \frac{\$1{,}000{,}000}{20{,}000} = \$500$$

which can be rewritten

$$(0.05 \times \$10{,}000) + (0.95 \times \$0) = \$500$$

That is, the average prize across a very large number of bets can be found by multiplying the probability of each possible outcome by the value of that outcome, and then adding together all of the resulting figures.

Economists call this value, averaged across a large number of bets, the expected value of a *single* bet. Further, they define individuals as being *risk loving* if they would be willing to pay more to participate in this gamble than the expected value of a single bet – for example, they might be willing to pay $600 to spin the roulette wheel described above once, even though that wheel only pays $500 on average.

Individuals are defined as being *risk averse* if they would not pay as much as the expected value of a "bet"; and as being *risk neutral* if they would be willing to pay an amount equal to, but not more than, the expected value.

As drivers appear, generally, to be risk averse, they will be willing to pay more than $500 for the first insurance policy that was described above. Indeed, it is quite possible that they would be willing to pay $600 or more. Thus, if the cost to insurance companies of "producing" an insurance policy was $500 plus, say, $75 in administrative costs, risk averse drivers may be willing to pay more than that cost. In that case, a mutually beneficial exchange could occur. Drivers would pay, say $585 for something that was worth $600 (or more) to them.

Note: If insurance companies were also risk averse, they might not be willing to offer insurance policies at a price equal to the expected cost of producing those policies - $575 in this case. But, because insurance companies issue thousands of policies, there is very little risk that their actual costs will differ significantly from the expected costs. As a result, they will be willing to offer policies at prices close to their expected costs. Insurance is possible because, by aggregating the risks of individual risk averse consumers (drivers in this case), insurers can create a pool of "bets" that is virtually risk free (making the insurers risk neutral).

[Note also that this analysis explains why many large companies and government agencies do not buy automobile insurance for their employees. If their automobile fleets are large enough, the average number and value of claims will approximately equal the expected value. There will be no further advantage from pooling those claims with an insurance company.]

"Bonus" question: why are people risk averse?

To see why people might be risk averse in some situations and risk loving in others, consider the following simple example. Imagine that you have $7.00 in your wallet that you plan to use to buy a sandwich and drink for lunch. I now offer you the following bet: I will flip a coin. If it comes up heads, I will pay you $5.00; if it comes up tails, you will pay me $5.00. Although the expected value of this bet is $0 (= (0.5 x $5) - (0.5 x $5)), it might not be reasonable for you to accept it.

The reason for this is that if you win, you will have $12.00 for lunch, whereas if you lose, you will have only $2.00. The difference between the satisfaction you will get from what you can buy for $7.00 (if you don't bet) and $2.00 (if you bet and lose) will probably be greater than the difference between the satisfaction from $7.00 and $12.00.

On the one hand, betting and *losing* might mean you get, say, just a cup of coffee (for $2), whereas *not* betting gets you a cup of coffee *and* a sandwich – the difference is quite a significant loss, call this L. On the other hand, betting and *winning* might mean you are able to buy, say, a "superior" sandwich instead of only an "ordinary" one – perhaps only a minor gain, call this G. The expected value of the bet is now (0.5G – 0.5L), which will be less than 0 if L is greater than G.

So, even when the expected *money* value of a bet is zero (or even positive), the expected *pleasure* from that bet may be negative, if the pleasure gained from each dollar added is less than the pleasure lost from each dollar subtracted. In this case, the consumer may, reasonably, choose not to engage in such a bet – he or she will appear to be risk averse.

This may be an explanation for why drivers buy automobile insurance. If you pay, say, $575 per year for the automobile insurance policy described above, you will have $575 per year less to spend on goods and services than if you had not purchased that policy. But you will also receive $10,000 once every twenty years from the insurance company.

You will buy this policy if the pleasure you gain from that $10,000 exceeds the pleasure you lose by spending $575 on insurance every year for twenty years: $11,500. It seems quite possible that this might be the case.

$575 per year is about $11 a week. That might mean, for example, that once a week you would take your lunch to work instead of buying it; or you might ride your bike to work instead of taking public transit.

Compare that loss of pleasure, accumulated over twenty years, with the loss you would suffer if you had not purchased automobile insurance and had become involved in an accident that cost you $10,000. This might force you to replace your comfortable apartment near the city centre with a smaller, less comfortable one much further out; or to give up five years of your annual holidays in the sun at $2,000 each. It is possible that such a catastrophic loss could result in a loss of pleasure that was much greater than would have been experienced had you made a large number of smaller adjustments to your spending patterns over a twenty-year period. You would appear to be risk averse.

MODULE 19. APPLICATION - ADVERSE SELECTION AND MORAL HAZARD IN INSURANCE

Insurance companies face two problems that arise because they are not fully informed about their customers' behaviour – *adverse selection* and *moral hazard*.

1. Adverse selection

In Module 18, it was assumed that drivers were homogeneous with respect to the probability of having an accident. When this assumption is relaxed, to consider that some drivers will be more likely than others to have an accident, it is seen that insurers may not be able to offer a premium structure that is attractive to safe drivers.

Consider the following simple extension to the first example discussed in Module 18. Assume, again, that the average probability an insurance company's clients will have an accident is five percent per year; but assume now that that figure arises because the probability of having an accident is two percent for half of its clients – its "safe" drivers – and eight percent for the other half – its "risky" drivers.

If the insurer knew which category each driver fell into, it would offer two different policies: the safe drivers would be charged the expected value of their claims, \$200 (=0.02 x \$10,000 + 0.98 x \$0), plus administrative costs – say, \$75 – for a total of \$275; and the risky drivers would be charged the expected value of their claims, \$800 (= 0.08 x \$10,000 + 0.92 x \$0), plus \$75, or \$875

But it is often the case that, although the drivers themselves know whether they are cautious or risk-taking, the insurance company does not – it only knows that drivers who "look like" its clients average one claim every twenty years. So, if it has, say, 20,000 clients, they will make 1,000 claims per year and its average cost per driver will be \$500 plus administrative costs of \$75. In a competitive market, it will charge a premium of \$575.

This creates a conundrum for safe drivers. They know that they are cautious, implying that the expected cost of an accident for each of them is \$200. They would have to be *very* risk averse to pay \$575 for a policy that they expected would pay them only \$200 on average. Many of them may choose not to insure themselves (or, in a country where automobile insurance was mandatory, to buy only the minimum required level of insurance). This is called *adverse selection* – safe drivers have "selected themselves" out of the insurance market.

Risky drivers, on the other hand, are given a bargain when they are lumped in with safe drivers. Even though their expected costs are \$800, their premiums are only \$575. We would expect all of them to buy insurance. The result is that only half of the potential clients buy insurance, even though most of them would like to do so.

[Furthermore, assume that the "risky" group is also divided into two: perhaps half have a five percent probability of having an accident and half an eleven percent probability. When premiums are $875, many in the less risky group will also not buy insurance and rates will climb even higher.]

The way that the insurance industry responds to adverse selection is to take great pains to categorize potential clients by their potential risks. This is why, for example, automobile insurance companies charge higher premiums to young men than to young women, and higher premiums to people who have had accidents in the (recent) past than those who have been accident-free. As these categorizations are unlikely to divide drivers perfectly, some drivers who are confident that they are better drivers than others in their category may choose to purchase less insurance than they would have had their risks been identified accurately.

2. Moral hazard

One of the unintended effects of insurance is that it reduces the incentive for those who purchase insurance to take actions that would reduce their accident costs - *moral hazard* This can happen in two ways.

First, insurance can reduce the individual's incentive to *minimize the costs* of any claim. For example, if health insurance pays for all medical bills, regardless of which doctor the patient uses, the incentive to find the most cost-effective doctor will also be reduced.

To minimize these effects, insurance companies often require that the client pays for the first $X of any claim, or the companies only pay for some percentage of a claim. With respect to unemployment insurance, for example, individuals are not usually paid for the first two weeks of lost earnings and are compensated for only a portion of their lost earnings – typically about sixty percent – after two weeks.

Second, insurance can reduce the *incentive to avoid risks*. If, for example, the cost of health insurance is the same for people who live a sedentary life as it is for those who engage in risky activity, such as rock climbing or hang gliding, the incentive to avoid those risks will be reduced.

To minimise these effects, insurance companies reward clients for taking actions that reduce risk and penalize clients for engaging in risky behaviour. For example, with respect to homeowner's insurance, premiums might be reduced for clients who install secure door locks (to reduce the probability of theft) or smoke detectors (to reduce the risk of fire). And with respect to health insurance, insurers might charge higher premiums to those who smoke or drink alcohol.

It may appear that automobile insurance companies penalize clients for risky behaviour by charging higher premiums to those who have had accidents than to those who have not. But this is not strictly the case. To see why this is so, consider the extreme case in which drivers who had an accident in the past were no more likely to have an accident in the future than were drivers who had no accidents.

In this case, the expected cost of future accidents to these two groups of drivers would be the same. And because the insurance industry is competitive, premiums will always be forced down to equal expected costs. Even if an insurance company told its clients that it would increase their future premiums if they had an accident – in order to encourage them to drive safely – they would have difficulty imposing such an increase. If they tried to do so, drivers would simply to switch to other insurers, who would be willing to charge them premiums based on their future, expected costs.

Of course, in the real world, there are various categories of drivers, from very safe to very unsafe. And we see that insurance companies charge higher premiums to those have had accidents than to those who did not. But this does not occur because insurers are "penalizing" drivers for their past behaviour. Rather, it occurs because drivers who have had accidents "reveal" that they may be in a higher risk category than drivers who have not had an accident (and, therefore, have higher expected costs).

Refer to the example in the section on adverse selection, where there are two categories of drivers, with accident probabilities of two percent and eight percent, respectively. If, before any of them have had an accident, the insurance company is unable to determine which category each of them belongs to, the expected probability of an accident for each driver will be five percent and all drivers will be charged the same premium.

At the end of one year, two percent of the first group will have had an accident and eight percent of the second group. For example, if each of the two groups had 10,000 drivers, 200 from the first group will have had an accident and 800 from the second group – 1,000 in total. (Conversely, 9,800 from the first group and 9,200 from the second will not have had an accident – 19,000 in total.)

The probability that a driver who had an accident in the first year will have an accident in the second year is the weighted average of the probabilities of the two groups:

$$\frac{200}{1{,}000} \times 0.02 + \frac{800}{1{,}000} \times 0.08 = 0.68$$

And the comparable probability that a driver who has not had an accident will have an accident in the next year is:

$$\frac{9{,}800}{19{,}000} \times 0.02 + \frac{9{,}200}{19{,}000} \times 0.08 = 0.49$$

That is, assuming that the drivers' individual probabilities of having an accident have not changed, the effect of having had an accident is to miscategorise some of the safe drivers as risky and to correctly categorise some (but not most) of the risky drivers. This raises the expected cost of an accident for those who have been observed to have had an accident but has very little effect on those who had not had an accident.

MODULE 20. APPLICATION - WHY ARE HOUSE PRICES RISING?

In many of the major cities of the world, rising house prices has become one of the hottest topics everywhere people meet. We all want to know what is causing the declining affordability of housing, and whether our governments can do anything about it. Two competing theories are that house prices are rising because of

(i) supply factors – prices are rising because the costs of building houses have been increasing;
(ii) demand factors – prices are rising because the number of people buying houses has been increasing (with "foreign buyers" often being identified as the main culprit).

This module will argue that, if the house building industry is competitive, supply factors are much more likely than demand to be the source of price increases. The module begins by arguing that the housing industry *is* competitive; and then analyses the effects of (i) an increase in average costs (supply) and (ii) an increase in demand. [Note: in each case, the analysis refers primarily to housing markets in North America, where there are far fewer government regulations than in Europe, especially with respect to the use of land.]

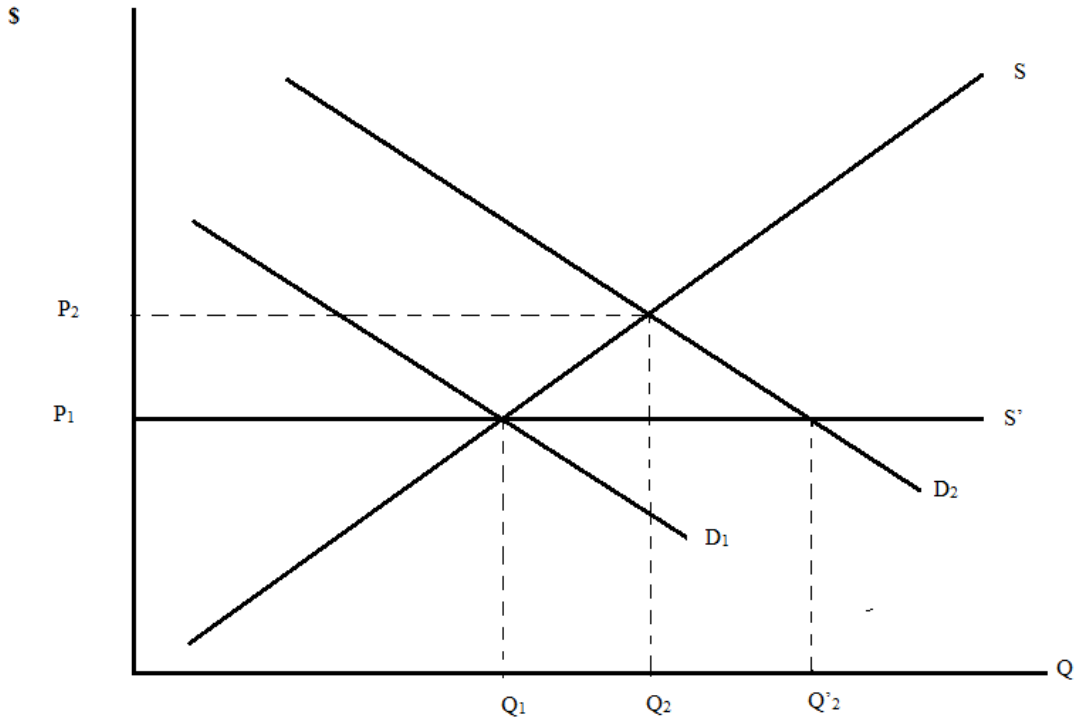
1. Is the House Building Industry Competitive?

The house building industry meets three of the most important criteria for competition:

(i) There are few economies of scale. Although it may be possible to reduce costs by buying inputs, such as wood and cement, in large quantities, these opportunities are limited, meaning that even firms building only three or four houses per year can often compete with firms that build hundreds.
(ii) All firms are equally well informed. Knowledge concerning how to build a house is very widely spread throughout the community. Large firms do not know more about how to install windows or build a wall than do small firms.
(iii) There is ease of entry. New firms can enter with little capital. Typically, a new firm starts by doing small renovations on existing houses, then expands to major renovations, and finally to construction of single family homes. Large firms are just firms that have been in business longer than have small firms.

Although housing is not a "perfectly" competitive industry, it meets the criteria sufficiently well that the model developed in Modules 15, 16, and 17 can be used to describe how the housing sector operates in most North American markets.

Figure 10. Effect of an Increase in Demand



2. Demand

Assume, in Figure 10, that the demand curve for housing is given by curve D_1 and that, initially, supply equals demand at price P_1 and quantity Q_1. Now, assume that demand increases, such that the new demand curve is D_2. If the industry supply curve is upward sloping – because the costs of inputs (perhaps land) increase as demand for them increases – the new equilibrium will occur at (P_2, Q_2): increased demand caused an increase in price (which is what most of us assume will happen).

However, if there was no shortage of workers, land, and materials in this city, the supply curve might well be horizontal, like S' in Figure 10. In that case, the price of housing would not increase as demand increased and the new equilibrium would occur at a point like (P_1, Q'_2): increased demand did not affect the market price. This is the outcome that might be expected in small rural towns or in depressed cities like Detroit, in both of which there is a lot of underutilized land and the unemployment rate is high.

Finally, it is possible that the supply curve could be downward sloping. However, as this seems unlikely in the housing market, this situation has not been depicted in Figure 10.

To conclude, Figure 10 indicates that it is not really *demand* that is determining price in the housing market; rather, it is the *slope* of the supply curve. Whereas even a small increase in foreign investment in the housing market might be associated with a significant increase in housing prices in a city with limited availability of land, such as in Vancouver, B.C., even a large influx of residents to a depressed city like Detroit might have very little effect on price.

3. Supply

On the other hand, regardless of the shape of the supply curve, an increase in average costs of production will lead to an increase in the price of housing and decrease in the quantity built. For example, assume that a season of bad forest fires in Scandinavia and western Canada has reduced the supply of lumber, leading to an increase in the cost of building materials. This will increase every firm's minimum average cost and, therefore, shift the supply curve of housing upwards, as in Figure 10a, where it is assumed that the industry faces diseconomies of scale. The price of housing will increase, from P_1 to P_2, and the number of houses will decrease from Q_1 to Q_2. And, as is seen in Figure 10b, this will occur even if there are no economies or diseconomies of scale (i.e. if the supply curve is horizontal).

Figure 10a. Supply Shift: Diseconomies

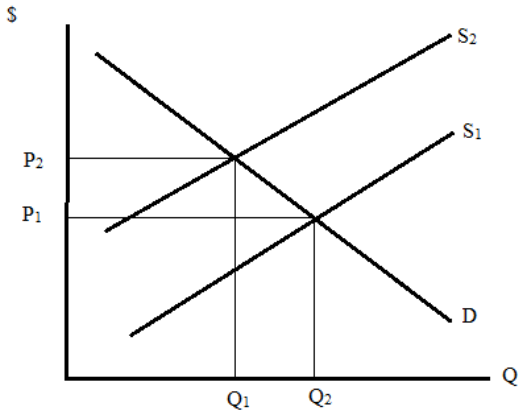
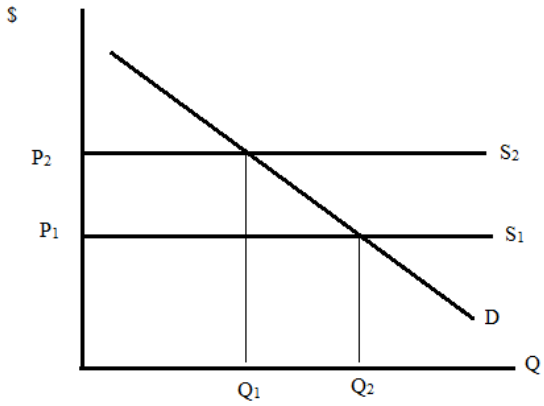


Figure 10b. Supply Shift: No Economies



4. Conclusion

This Application suggests that if the housing industry is competitive, an increase in demand will only lead to an increase in prices if the supply conditions are conducive (specifically, if the supply curve is upward sloping); whereas a decrease in supply will lead to an increase in price in all cases. In short, it is supply and not demand that is responsible for rising house prices.

Why does the model produce this asymmetry? Intuition would suggest that supply and demand would be equally important in determining prices. The answer is that whereas the supply curve

may sometimes be horizontal (or even downward sloping), the demand curve can be assumed to be downward sloping in virtually all cases. Why? When house prices are high, only the wealthiest individuals in society can afford their own houses. Hence, high prices are associated with low quantities demanded. As prices are reduced, however, more families find houses affordable and the number of houses demanded increases: the demand curve is downward sloping.

[With respect to some goods (say, number of restaurant meals or numbers of pairs of running shoes) quantity demanded increases as price decreases because each family buys more of those goods. But this is not (usually) what happens with respect to houses. Rather, the industry demand curve for houses slopes downward because some consumers have more money than others so, as prices decrease, more individuals become able to afford their own house.]

MODULE 21. APPLICATION - THE COMPETITIVE LABOUR MARKET

To this point, the discussion of competitive markets has concerned markets for consumer goods, which economists call *product markets*. This module extends that analysis to markets for *inputs* into production, which economists call *factor markets*. (The term "factor markets" arises because inputs into production are called "factors of production.") Specifically, this module and the next will focus on the factor market that is of greatest personal interest to most of us, the *labour* market, (that is, the market for workers).

First, in this module, it is argued that many, if not most, labour markets are competitive (in the sense set out in Module 15), implying that they can be described using a supply and demand model. In Module 22, the competitive model of the labour market is used to understand how the imposition of a minimum wage might affect employment and earnings.

By analogy with the model of competitive product markets, a competitive labour market is one in which there are many individuals with very similar work-related characteristics, competing for employment with many firms, none of whom employs a large percentage of the work force. Typically, when we talk about competitive labour markets we think of unskilled workers, such as labourers, fast food workers, cleaners, and clerical workers. However, even the markets for such professional workers as accountants, lawyers, and engineers meet most of the criteria for competition: there are lots of such workers, they are relatively interchangeable, and there is a very large number of firms competing for their services.

The analysis of these markets is divided into three components: demand, supply, and the interaction of supply and demand (to determine the "price" of labour – wages).

1. Demand for Labour

The demand for factors of production is said to be a *derived demand*, by which economists mean that such factors are not valued for themselves, but for the goods that they can produce. For example, whereas the customers of a hamburger restaurant value the product of that restaurant for the pleasure that hamburgers give them; the owner of the restaurant receives no value from his or her employees directly. Rather, workers' value derives from the price of the hamburgers that they can produce (minus the costs of other inputs). The more money the owner makes from the burgers that the employees make, the greater is the value that is placed on those employees.

The derived demand for an hour of an employee's work is found by multiplying the number of units of output that each employee produces in an hour by the net revenue received from each unit of product sold. If (i) a cook makes ten hamburgers per hour, (ii) each hamburger sells for $10, and (iii) the cost of each hamburger's ingredients (meat, bun, and prorated costs of the kitchen equipment and the building) is $8, then (iv) the value of the cook to the restaurant owner is $2 per hamburger times 10 hamburgers per hour, (($10 - $8) x 10), or $20 per hour.

If we assume that all workers in a competitive market are equally productive, then the number of units cooked per hour will not vary with the number of cooks hired. However, as we also assume that the demand curve for most consumer goods is downward sloping, when the number of workers employed by any industry increases, (and, therefore, the amount of output produced increases), the price of that industry's product will decrease (the industry moves "down" its demand curve). The result is that the wage that the firms in a labour market are willing to pay for the "next" worker will decrease as the number of workers increases; and the demand curve for labour, like the demand curve for the employers' products, will be downward sloping.

2. Supply of Labour

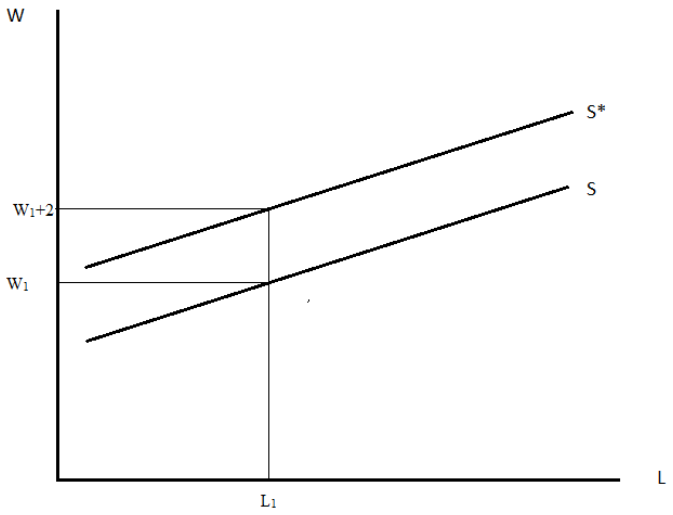
The minimum wage at which workers will supply themselves to employers in a labour market will depend primarily on the opportunity cost of those workers' time in their next best activities and on the cost to those workers of obtaining the skills necessary to work in that market.

Opportunity cost: The "next best" opportunity will be either the highest wage workers can obtain in other industries or the value of their time at home, if they choose not to work in the labour market. Assume that it is the former and, for simplicity, that there are only two industries, X and Y, that employ the required workers. When industry X begins to hire workers, it will hire them away from Y and will have to pay them at least the wage that they are currently receiving in Y.

As more workers are hired in X, the number available in Y will decrease. If demand for Y's product is such that it cannot afford to raise wages in response, or if Y is able to replace workers with machines or computers, X will be able to attract workers for a small increase in wages. If, however, Y's customers are willing to accept price increases (and workers cannot be replaced), Y may respond to X by increasing the wages of its workers; and X will have to pay higher and higher wages as it attracts more workers from Y – the supply curve will be "upward-sloping."

Similarly, if industry X attracts workers from outside the labour market – for example, from individuals who might otherwise have been going to school, looking after children, or enjoying retirement – it will have to pay them a wage sufficient to compensate them for the value of their foregone time in those activities. At first, X can be expected to attract those whose time has the least value – perhaps students who are not doing well at school or retirees who have found that their pensions did not go as far as they had hoped. As firms in X hire more and more people, however, they will have to pay higher wages as they hire people with higher and higher opportunity costs. Again, the supply curve will be upward sloping. An example of such a supply curve, S, has been drawn in Figure 11.

Figure 11 Supply Curve of Labour



Cost of obtaining skills: Assume that S in Figure 11 represents the supply curve of workers who have high school education, to an occupation such as carpentry; and that L_1 workers had been willing to supply themselves at wage W_1. Now assume that the government introduces a rule that requires workers to obtain three years of trade school (as well as high school) before becoming carpenters, at a total cost of E.

But remember that W_1 represents the wage that workers could obtain in their "next best" occupation. Thus, if carpenters' wages are not increased after the new rule is introduced, they will not spend E to obtain the required education – they can obtain W_1 elsewhere without spending E. If employers wish to continue attracting workers into carpentry, they will have to increase wages sufficiently to compensate them for spending E.

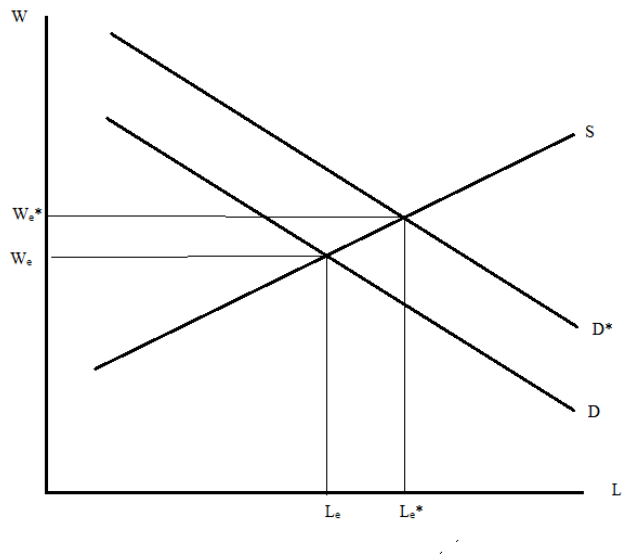
As individuals spend about 80,000 hours in the labour force over their lifetimes (approximately 40 years at 2,000 hours per year), the required increase will be E/80,000 per hour. For example, if E is $160,000 (composed of earnings foregone while at school, plus the cost of books and tuition), wages will have to be increased by $2.00 per hour, to W_1+2. As this increase will be required at every level of employment, the supply curve will shift vertically from S to S* in Figure 11. In short, all else being equal, the occupations that require the greatest investment in education will have the highest wages, simply in order to compensate individuals for the cost of that investment.

3. Supply and Demand – Wage Determination

Now, just as was done for prices and quantities of goods sold, it is possible to combine supply and demand to obtain the equilibrium wages and numbers of workers hired. This has been done in Figure 12, where a downward-sloping demand curve, D, has been superimposed on curve S from

Figure 11. There we see, using the same reasoning that was employed in Figure 4 of Module 10, that the equilibrium wage-employment combination will occur where the supply and demand curves intersect; that is, at (W_e, L_e).

Figure 12 Wage Determination



It is interesting to ask what will happen if demand should now increase from D to D* (in Figure 120. Assume for this purpose that Figure 12 represents wages of workers in automobile manufacturing. As demand for automobiles increases, the firms in this industry will need additional workers for their assembly lines. You can imagine that these workers will be drawn away from many other occupations: mechanics, labourers, agriculture workers, salespeople, bookkeepers, cooks, daycare workers, delivery drivers, etc.

If the automobile firms in question were in a country with a centrally planned economy, resolution of this question might prove overwhelming: which alternative occupations and industries should workers be drawn from, and how many? But in a market-based economy with competition, the response is simple: the firms post ads on-line and then sort through the responses that they receive. They do not have to worry about where those workers come from. All they know is that, if workers apply to them, they must prefer the wages the automobile firms are offering to those they are earning at their current employers (or to the value of their time outside the labour market). Ultimately, a new market equilibrium will be found in Figure 12, at (W_e*, L_e*).

From society's point of view, there will have been an efficient re-allocation of workers. The reason that the automobile industry was able to pay more for the workers it attracted than did their former employers is that the value consumers placed on automobiles had increased relative to the value they placed on the goods that workers were producing elsewhere. In short, when demand for automobiles increases relative to that for other goods, the market provides a method by which

resources (here, workers), can be transferred from the goods that are now in relatively low demand to those (automobiles) that are now in high demand.

MODULE 22. APPLICATION - MINIMUM WAGES IN A COMPETITIVE LABOUR MARKET

Recently, a considerable amount of public debate has arisen concerning the (apparent) increase in the income gap between low- and high-wage workers. Many commentators have responded by calling for an increase in the minimum wage. This module investigates the effects that such a change will have on the incomes of low-wage workers.

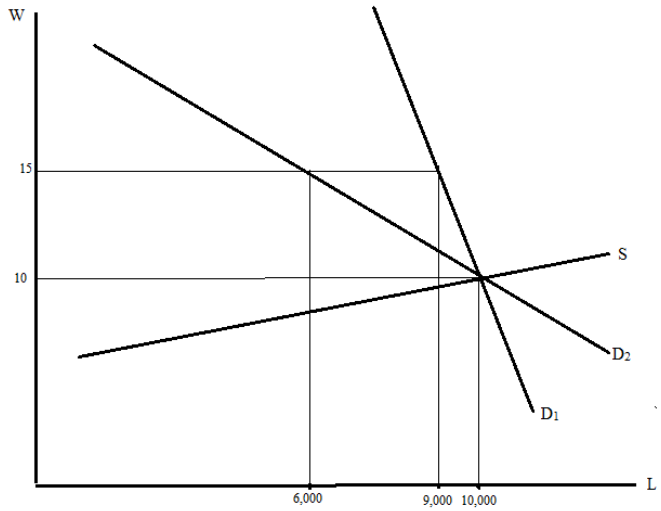
1. Impact on Employment

Imagine that Figure 13 describes the market for fast food workers, with 10,000 workers being employed at a wage of $12 per hour. If the government introduces a minimum wage of $15, the impact on employment will depend in large part on the *wage elasticity of demand* for workers (i.e. on the percentage change in the number of workers demanded divided by the percentage change in the wage rate – see Module 8). If the demand curve between $12 and $15 is wage *inelastic*, like D_1, when the wage increases by 25 percent (from $12 to $15), employment will decrease by less than 25 percent – here, it is 10 percent, from 10,000 to 9,000.[6] Ninety percent of the workers in this market experience a significant increase in wages, while 10 percent lose their jobs and incomes entirely. Importantly, total employment income (among those who keep their jobs) increases from $120,000 (= $12 x 10,000) to $135,000 (= $15 x 9,000) - so total income increases.

If, however, demand is wage *elastic*, the percentage decrease in employment will be greater than the percentage increase in wages. For example, when wages increase by 25 percent, from $12 to

[6] Technically, a straight line demand curve has a different elasticity at each point. D_1 only has an elasticity of 10%/25% in the region described.

Figure 13 Minimum Wage



$15 along D_2, employment decreases by 30 percent, from 10,000 to 7,000. A little more than half the workers will earn significantly more than initially, while a little less than half will lose their jobs entirely. In this case, total employment income decreases from $120,000 (= $12 x 10,000) to $105,000 (= $15 x 7,000).

Is one of these scenarios more likely than the other? The answer depends primarily on two factors: how "necessary" the product is and how easy it is to substitute machines, (including computers), for people.

Necessity

When the cost of labour increases, the firm's average costs of production will increase and so will the price that it charges for its product. If the good that the firm produces is a necessity, that increase in price may have only a small effect on quantity demanded. For example, if 50 percent of the firm's costs are composed of wages and if wages increase by 25 percent (while all other costs remain constant), the price of the firm's product will increase by 12.5 percent. If that product is a necessity, like eggs, water, or housing, quantity demanded, and therefore employment, may decrease by significantly less than 12.5 percent.

If, on the other hand, the firm's product is not a necessity, a 12.5 percent increase in the costs of production (and in the price of the product) might lead to a (percentage) decrease in quantity demanded that significantly exceeded the increase in price. For example, a 12.5 percent increase in the cost of airline travel might induce many consumers to stay at home for their vacations rather than to fly abroad, in which case demand for airline workers might fall by more than 12.5 percent.

Substitutability

If it is possible to substitute machines (including computers) for workers, a 25 percent increase in the minimum wage may induce firms to reduce employment significantly. For example, an increase in the wages of workers in warehouses and manufacturing facilities might induce employers to substitute robots for workers; an increase in the wages of fast food workers may encourage the development of computerised ordering systems; and an increase in the wages of truck drivers could increase the rate at which driverless vehicles are introduced. In each case, the development of new technologies may take enough time that the connection between a wage increase and an employment decrease will not be clear, but that does not mean that the connection does not exist.

Summary

As wage increases lead to increases in the prices of the goods being produced, and as demand curves are almost always downward sloping (i.e. higher prices mean fewer goods purchased), it is highly likely that an increase in the minimum wage will result in a decrease in the level of employment (in the industries to which the minimum wage applied).

Nevertheless, if demand for workers is wage inelastic, an increase in the minimum wage will increase the total amount of income being paid to those workers who maintained employment. In the example at the beginning of this section, an increase in the minimum wage from $12 to $15 resulted in a decrease in employment from 10,000 to 9,000 workers, implying that total employment income *increased* from $120,000 to $135,000. A large number of workers experience a 25 percent increase in their take-home pay, while a smaller number lost their incomes entirely. (Conversely, of course, if demand is wage *elastic*, total employment income will fall.)

2.2 Effect of labour turnover

The "conclusion" that many workers benefit from an increase in the minimum wage, at the expense of a small number of workers who are made worse off, assumes a static labour market. But many low-wage industries are plagued by high rates of turnover. That is, a high percentage of each firm's workers leave each month and are replaced by other workers. It is possible, therefore, that the higher earnings that workers receive after an increase in the minimum wage may be shared among all workers, as they move in and out of employment. In an extreme case, each of the 10,000 workers in Figure 13 might be employed for 90 percent of the year and unemployed for 10 percent. Thus, if they had each worked 2,000 hours per year at $12 before the minimum wage was introduced, they might work an average of 1,800 hours (and be unemployed for 200) after that introduction. The effect of the minimum wage might then be to increase their annual incomes from $24,000 (+ $12 x 2,000) to $27,000 (= $15 x 1,800). (Again, this gain occurs only if demand is wage inelastic. If demand is wage elastic, all workers will receive lower incomes when turnover is high.)

2.3 Composition of the low wage workforce

A large percent of low-wage workers are members of families that are not poor. Many, for example, are teenagers who live with non-poor parents. To the extent that this is the case, an

increase in the minimum wage may have little effect on the number of poor families, even if the average low-wage worker was to benefit from the minimum.

2.4 Variable productivity

In Module 21, a competitive market was defined to be one in which all workers had very similar work-related characteristics. In reality, what is considered to be "very similar" is often stretched a little. Walmart, for example, has a policy of hiring individuals with physical and mental disabilities; fast food restaurants often employ teenagers with very little work or life experience; and taxi companies accept immigrants with little proficiency in English. When wages are low, firms are often willing to intersperse these individuals with employees who meet more traditional requirements.

When a minimum wage is introduced, however, employers may find it more difficult to justify retaining their least productive workers. This will be particularly true if the increased wage has meant that firms hire significantly fewer employees, as is the case when demand is wage elastic. The result may mean that the lay offs that follow the imposition of a minimum wage will fall disproportionately on disabled workers, teenagers, and immigrants.

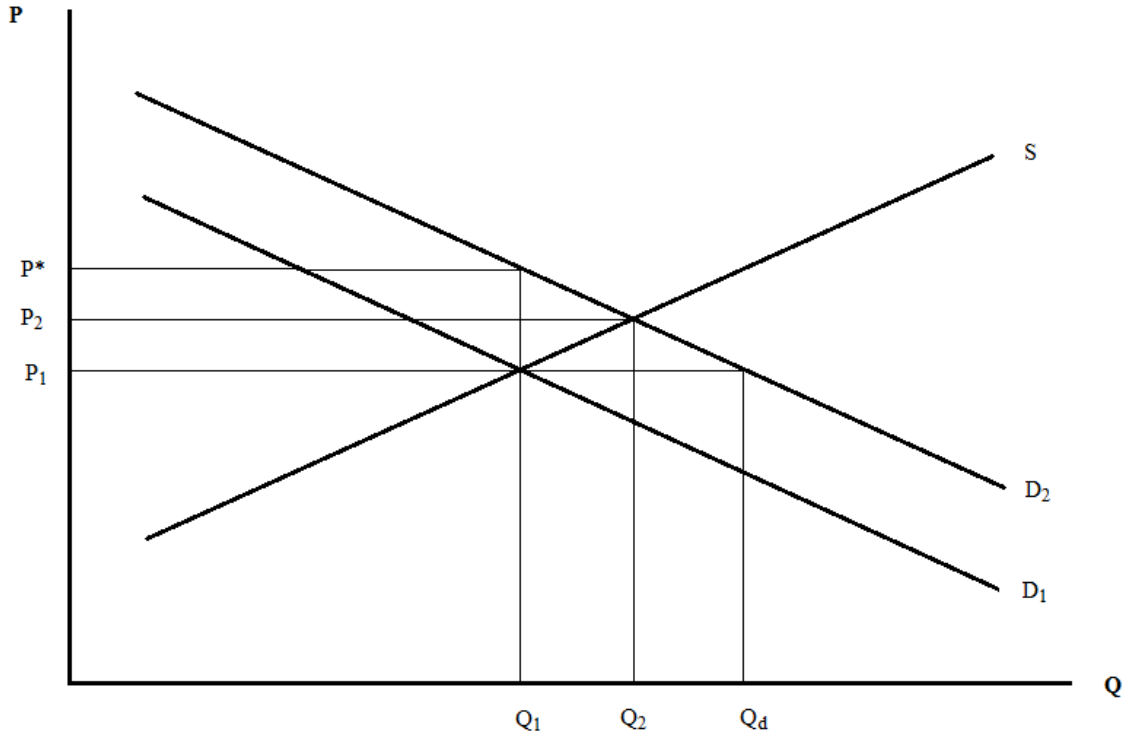
MODULE 23. ALLOCATION OF SCARCE RESOURCES IN A COMPETITIVE MARKET

Module 6 asked how effective centrally planned economies would be at allocating resources. This module now asks the same question with respect to economies composed of competitive industries. To simplify the analysis, focus on a representative competitive industry like hamburger restaurants.

Assume, in Figure 14, that demand for hamburgers is represented by curve D_1 and that supply is represented by curve S. The equilibrium price and quantity will be (P_1, Q_1).

Now, assume that demand shifts, so that the new demand curve is represented by D_2. Initially, there will be excess demand, of (Q_d - Q_1), at price P_1. With this excess demand, queues will develop for hamburgers and producers will realise that they are able to raise prices. As prices rise, quantity demanded will fall and quantity supplied will rise until supply and demand equal one another at (P_2, Q_2).

Figure 14 Allocation of Scarce Resources



Here we have a case of competing users chasing after scarce resources. The *scarce resources* are the goods and services – including workers, buildings, beef, and kitchen equipment - that will be

needed in order to produce the additional hamburgers. As these goods and services were previously inputs into other products, consumers of hamburgers will have to compete with the consumers of those other products for these inputs.

The ability of consumers of hamburgers to attract resources away from other uses depends upon the value which consumers place on burgers *compared to* that which they place on other products that could be produced using the same inputs. That the supply curve has not shifted in Figure 14, whereas the demand curve has, indicates that there has been a shift of preferences in favour of hamburgers, relative to the demand for other products. This will allow hamburger purchasers to attract resources (the resources that will be necessary to produce extra hamburgers) away from other uses by offering higher prices for those resources.:

The mechanism by which this happens is that, when demand for hamburgers increases, demand for the *inputs* into hamburgers increases also. As a result, the prices of those inputs increase, causing an upward shift in the supply curves of all other goods that use those inputs (because their costs have increased). That shift causes the prices of the other goods to increase, thereby reducing the quantity demanded for those goods and freeing up resources for the production of hamburgers.

Two effects are important: First, when there is an increase in consumer preferences for hamburgers versus other products, consumers' resulting willingness to pay higher prices for burgers allows manufacturers to pay higher prices for inputs and, therefore, to attract the necessary inputs from other users. Economists would say that this results in an "efficient transfer of resources," *to* firms producing hamburgers *from* firms producing goods that are less valued than running shoes.

Second, if the inputs required to produce extra hamburgers were formerly distributed among many different industries, the operation of supply and demand will ensure that the inputs reallocated to hamburgers will come from the industries whose products are least valued.

For example, one of the inputs that the hamburger restaurants will need is extra workers. Assume that the relevant workers were initially employed in two other industries, A and B. When hamburger restaurants raise wages to attract new workers, firms A and B will have to raise their wages in response, in order to retain their workers. This will raise the costs of production to both A and B, their respective supply curves will shift upwards, and the equilibrium prices of their products will increase.

If consumers of A are more "price sensitive" than are consumers of B (that is, if quantity demanded changes by a relatively large amount at A when prices change), the equilibrium quantity of A will decrease by a greater percentage than will the quantity of B. (For example, A might be a luxury good that consumers are willing to give up when prices increase, whereas B is a necessity that consumers will be reluctant to give up.) That will free up more workers from A than from B. So, hamburger restaurants will get more of their additional workers from the industry, A, that is relatively unimportant to consumers than from the industry, B, which is relatively important. And this is the efficient outcome for society.

The analysis of this module suggests that *if* an economy was composed of competitive industries, the market would answer the question of "what should be produced" efficiently. But economies

are *not* composed of competitive industries. Economists designed the competitive model to study how markets would operate in an ideal world, in much the same way that physicists study how objects fall in a vacuum. Before we can ask how markets allocate resources in the real world, it will be necessary to modify the competitive model to allow for the many market imperfections that we find in the real world.

Some of those imperfections are identified in Module 24; and Parts E and F of these notes study how those imperfections affect the operation of a market-based economy. It is only when those imperfections are taken into account that an appropriate comparison can be made between market and planned economies.

2. Comparison of a Market Economy with a Centrally Planned Economy

Notice that when demand for hamburgers increases, no government agency, or regulatory board, or trade union has to tell workers to move from their previous employers to the hamburger industry, and no human agency has to decide that the new workers for the hamburger industry will come from firms for whom demand is price sensitive. Instead, it is, in Adam Smith's words the *invisible hand* of the market that will make these decisions, organically.

In a centrally planned economy, however, the planner will have to make these decisions. To see how difficult this will be, consider a fairly straightforward product like housing. Imagine, initially, that the planner has allocated a two-bedroom apartment to each family in the economy. Now, assume that, with increasing income, consumers begin to demand "more" housing – perhaps more bedrooms, or a larger kitchen, or an attached garage. The planner will face (at least) three problems, each of which will be extremely difficult to answer efficiently:

Has demand increased? The first question that will bedevil the planner is determining whether demand has increased. As it is the planning agency that has built and allocated the existing housing, how will consumers "reveal" their increased demand? They won't be able to bid up the "price" of three-bedroom houses because (i) houses don't have prices; (ii) the planning agency hasn't build any three-bedroom houses (or at least not enough); and (iii) even if there was an existing stock of three-bedroom houses, the people currently living in them wouldn't have an incentive to give them up to other consumers (until the planner had built alternative housing for them).

To determine whether consumers would like larger, better appointed houses, the planner will have to ask them. But this raises the second question:

How strong is any increased demand? Just asking people whether they want bigger better houses isn't enough. The planner also has to determine how strong that demand is. If demand isn't very strong, the case for reallocating resources from other products to housing will also not be very strong. But how does the planner determine that strength? In the market economy, it is measured by the amount of money people are willing to give up to get more houses. In the centrally planned economy, it is going to require assessment of the nature of consumers' answers to questions about their "wants" and "needs."

But even if the planner is able to get a rough idea about the strength of demand for new housing, how will it determine whether that demand is enough to justify taking resources from the production of other products?

Where will resources come from? Where will the workers come from to build the new houses? From farming? Banks? Shoe stores? Steel plants? And where will the wood come from? Will workers have to be allocated to cutting down more trees, and building bigger sawmills? Where will the steel come from for more nails and more dishwashers and microwaves? Etc. Etc. How will the planner even begin to answer these questions?

In a market economy, the myriad of such questions is answered without anyone even thinking about them. When people reveal their preferences for more houses by spending more on housing, builders use the extra money they receive to bid inputs from other industries, and not just "any other" industries, but from those industries that make the products that are least valuable to consumers.

This seems to imply that market economies are inherently superior to centrally planned economies. But, not so fast. Before we can conclude that, we first have to deal with the "grass is greener fallacy."

3. The Grass is Greener Fallacy

Imagine that two neighbours, A and B, are arguing about which of them has the greener grass. They decide to resolve the dispute by asking C to adjudicate for them. When C investigates A's grass, he notes a number of brown spots. B says, immediately, "See, what did I tell you? My grass is greener than yours. Even C agrees."

Why is B's argument fallacious? It is because the argument between A and B is not about whether A's grass is perfect; it is about whether A's grass is greener than B's, and this cannot be determined without looking at B's grass also.

The *grass is greener fallacy* is particularly rampant in political debate. During election campaigns it is common for candidates to spend most of their time criticizing their opponents' platforms or performance. Much less time is spent proposing policies of their own. The implicit assumption is that if I can show that my opponent's polices are flawed that must mean that my own policies are better than hers (even if I have not actually proposed any meaningful policies of my own.)

The grass is greener fallacy also applies to debates between proponents of planned economies – such as China's and Russia's – and those of market economies – such as North America's and Europe's. Perhaps typical of this debate are the arguments raised in Section 2, above, showing that centrally planned economies have difficulty allocating resources efficiently.

The fallacy is that Section 2 implicitly compared those economies with an economy composed of competitive markets. But real world markets deviate from the competitive ideal in many important ways. Before we can draw meaningful conclusions about the efficiency of market economies, we

first need to investigate how they really operate, warts and all. That will be the function of the modules in Parts E and F.

MODULE 24. SHORTCOMINGS OF MARKET ECONOMIES

The idealized model of competitive markets that has been discussed in Part D provides many useful insights into the operation of market economies. Nevertheless, it deviates from the real world in many important ways. The remaining parts of these lectures, will discuss the following important modifications to the competitive model that will allow for a more nuanced insight into how markets work.

1. Imperfect Competition

The competitive model assumes that industries are composed of large numbers of relatively small firms, all producing roughly the same product. Although this does apply to many industries, a significant number of industries, called *monopolies* (from the Greek "monos: one" and "polein: seller"), have only one firm; and others, called either *oligopolies* (from "oligoi: few" and "polein: seller") or *monopolistic competition*, are composed of small numbers of firms which often produce goods that differ from their rivals' in important ways.

1.1 Monopoly

Monopolies arise primarily because firms have been given patents or copyrights over their products, which allow them to prevent other firms from competing with them; or because economies of scale are such that average costs of production are lowest if all output is produced by one firm. The clearest example of the latter is delivery of water or natural gas to residential areas. As it is much less expensive to have one water line or gas pipeline in each street than to have two or more, once one firm has installed such a line, any additional firms are at such a disadvantage that they will not try to compete.

The most important difference between a monopolistic industry and a competitive one is that whereas each competitive firm is so small relative to the industry that it has no control over the price of its product, the monopolist is the only firm in the industry, so it is free to choose whichever price it likes. This usually means that unregulated monopolists will set much higher prices, and produce less output, than would comparable competitive industries.

1.2 Oligopoly

Most manufacturing industries - such as those producing automobiles, computers, and steel – financial firms – such as banks and investment companies – and retail chains – such as department stores and clothing stores – are composed of small numbers of dominant firms (plus, often, a large number of smaller firms that tend to be "followers" of the larger firms).

The most important characteristic that distinguishes oligopolistic industries from competitive ones is that each firm is aware that the decisions it makes about price and output are closely observed by its rivals. When an oligopolist is thinking about lowering its prices, for example, it has to consider how its rivals will respond. If they lower their prices also, it will attract fewer additional customers than it would have if they had left their prices unchanged. This interrelationship among firms, which is often referred to as "conditional interdependence," will be analysed in Section E using *game theory*.

1.3 Monopolistic competition

Industries that are monopolistically competitive fall between oligopolistic and competitive industries. They are composed of large numbers of firms, but not such a large number that each firm can ignore how its rivals behave.

2. Externalities

Implicit in the construction of the competitive model is the assumption that the prices charged for inputs reflect the values of those inputs in their next best use. A pig farmer, for example, is only able to obtain feed, land, and buildings by "bidding" those factors away from alternative uses. It is the combination of this assumption with the assumption that the price consumers will pay for pigs reflects the value that consumers place on them that allows us to conclude that competitive industries allocate resources in such a way that they are used to make the most valuable goods and services.

External costs
In some cases, however, the prices that firms pay for their inputs do not reflect the opportunity cost of producing them. A simple example occurs when the inputs in question are wild animals, such as fish. As no one "owns" the fish in the ocean, there is no one who can charge individuals for the opportunity cost of catching them – which is that an increase in this year's catch will lead to a reduction in next year's. The result is that too many fish will be caught.

This is an example of an *external cost*. The individual fisherman's actions place costs on society (fewer fish in the future) that are not borne by him – they are "external" to him – and, therefore, he does not take them into account when deciding what actions to take.

Another common example of external costs occurs with respect to air and water. If my neighbour takes his garbage to the dump, he will have to pay the transportation costs and the dump fees himself. Whereas if he just burns his garbage in his back yard, he won't have to pay for dumping it into the air. As no one "owns" the air, the *private cost* (to my neighbour) of using air is zero, even though the cost to his neighbours, the external cost, is that they have to suffer the smell and nuisance of his actions.

If individuals like my neighbour do not have to consider the external costs they impose, they may well choose to burn their garbage rather than take it to the dump, even when the latter is the

efficient action. Similarly, the large scale pollution of both the air and water by manufacturers occurs not only because they are "bad corporate citizens," as is often the charge against them, but also because they have taken advantage of society's implicit offer to let them use air and water for "free." The numerous methods that economists have suggested for dealing with external costs will be discussed in the notes in the next section.

External benefits
It is also possible to have external *benefits*. This occurs when the price that a firm receives for its product is less than the value that society places on it. A simple example is a fireworks display. Unless you hold your display in a remote country area, it will be difficult for you to charge the people who benefit. The result is that if such displays are left to private firms, they may be underproduced.

A more important example of the potential effects of external benefits is armies. If I raise an army to protect my property, it may be difficult for me to prevent my neighbours (within maybe a hundred kilometres or so) from benefitting from my actions. If so, they may try to *free ride* on my beneficence, and not contribute anything to my costs. The benefits they receive will be external to me and I will not take those benefits into account when deciding how much to spend on armed forces. (Indeed, I might reasonably be expected to pay little or nothing.) Armies will be underproduced.

3. Principal-agent Problem

When developing the model of a competitive labour market, it was implicitly assumed that the employer could easily measure the output of each worker. For example, in one case it was assumed that a cook was able to produce $20 worth of hamburgers per hour. In many situations, however, it is difficult for the employer to determine how productive individual workers have been – either because it is difficult to measure the value of the worker's efforts or because it is difficult to disentangle the individual worker's contributions from the contributions of the other workers in a group.

A common example of this situation occurs with respect to real estate agents. Assume that you could have sold your house for $500,000 with little or no effort. If you hire a real estate agent, what you would like her to do is to obtain a price that is as much above that as possible. That suggests that your contract with the agent should offer a fee that is a function of the *difference* between the sale price and $500,000. But most real estate contracts call for the client to pay the realtor approximately three percent of the *total* sale price of the house. So, if the realtor sells the house for $550,000, she will receive$16,500, of which only $1,500 is compensation for the effort it took to raise the price by the "last" $50,000. That does not seem like a very good return on the realtor's efforts. She would probably be better off spending her time getting another listing than trying to get much more than $500,000 for you.

There are many examples of problems like these. Some involve employer-employee relationships; but others involve the hiring of experts, such as doctors, lawyers, and accountants, to perform

professional services for you. In all such cases, the person doing the hiring is known as the *principal* and the person being hired is known as the *agent*. The notes that follow will discuss the many ways that principals write contracts with their agents in such a way as to create the greatest harmony between the principal's goals and the agent's behaviour.

4. Discrimination

It is difficult to explain why there might be racial or sexual discrimination if labour markets are competitive. Assume that discrimination means that employers hire from the dominant group even when their wages were higher than equally productive workers in a minority group. For example, assume that both male and female cooks are worth $20 per hour to their employers, but restaurants pay $20 to male cooks and $16 to female cooks. The firms with the greatest percentage of female cooks will have the lowest costs and will be able to undercut the prices of the firms with the highest percentage of male cooks. Either the only firms that would survive would be those who hired only female cooks, or all firms would be forced to pay the same wage to males and females.

Modules 43 and 44 will show that there are numerous barriers, not considered in the model of the competitive market, that might explain how discrimination could prevail in the long run.

5. The Government as Producer

In Part B of these notes, it was argued that planned economies would have difficulty solving the fundamental economic problems of: what, how and for whom? And Parts C and D explained why economists believe that competitive market economies may have success in answering these questions. Economists tend to forget, however, that even in the most market-oriented economies, a significant percentage of economic decisions – perhaps a quarter to a third – are made by planners. We don't call the agencies responsible for this function "central planners;" rather, we call them government departments – education, armed forces, transportation, etc. These agencies face the same problems identifying what the optimal policies are that were identified with respect to the central planners of such communist countries as Russia and China. Section G of these note investigates how government agencies might solve these problems.

E. MARKET ECONOMIES - IMPERFECT COMPETITION

MODULE 25. MONOPOLY

The competitive model that was discussed in Part D assumes that industries are composed of large numbers of relatively small firms, all producing roughly the same product. Although this does apply to many industries, a significant number of industries, called *monopolies* (from the Greek "monos: one" and "polein: seller"), have only one firm.

Monopolies arise primarily because firms have been given *patents* or *copyrights* over their products, which allow them to prevent other firms from competing with them; or because economies of scale are such that average costs of production are lowest if all output is produced by one firm.

The clearest example of the latter is delivery of water or natural gas to residential areas, by *utilities*. As it is much less expensive to have one water line or gas pipeline along each street than to have two or more, once one firm has installed such a line, any additional firms are at such a disadvantage that they will not try to compete.

A third type of monopoly, which will be discussed in Section 2, below, arises when the government orders all firms in an industry to act collectively, such as when dairy farmers are required to sell their product through a *marketing board*.

The most important difference between a monopolistic industry and a competitive one is that whereas each competitive firm is so small relative to the industry that it has no control over the price of its product, the monopolist is the only firm in the industry, so it is free to choose the price it will charge. It will be shown in the notes that follow that this means that monopolists will set much higher prices, and produce less output, than would comparable competitive industries.

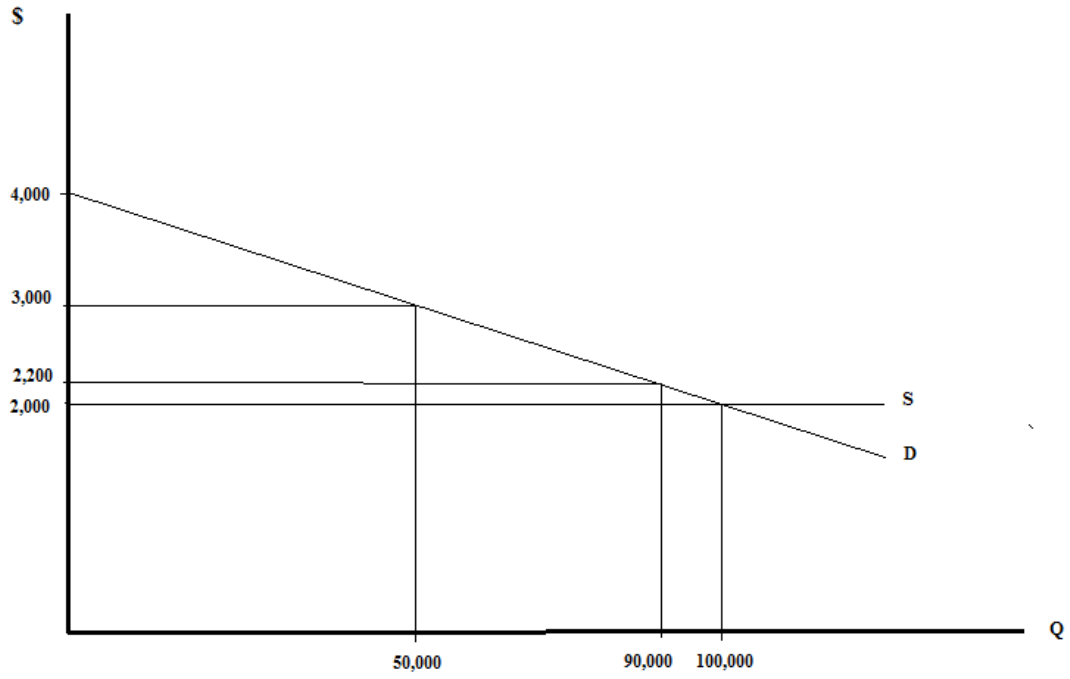
The simplest way to describe how monopoly affects prices and output is to start with a model of a competitive industry. Then, ask what would happen if the firms in that industry were all to be bought by one owner (or if they were all to work together as if they were a single firm), with no other changes. This way, we can see what the effect of monopolization is, independent of any other factors such as differences in cost structures. Accordingly, Section 1 will describe the characteristics of a simple competitive industry. Section 2 will then ask what would happen to the prices and quantities in that industry if it was to be monopolized.

1. Example of a Competitive Industry

Imagine an agricultural industry composed of identical farms, each of which produces 100 tons of product per year at an average cost of $2,000 per ton (i.e. $1.00 per pound). If it is assumed that there are no economies or diseconomies of scale in this industry, the cost per ton will always be $2,000 and firms will be willing to supply any amount of output at that cost. Hence, the industry supply curve will be a horizontal line at $2,000. This curve is labeled S in Figure 15.

Assume also that demand is given by curve D in Figure 15. In this case, the price will settle at $2,000 in the long run: if the price was higher than $2,000, firms would earn economic profits (profits that exceed the farmers' opportunity costs) and new firms would be attracted, driving the price down; and if the price was to fall below $2,000, profits would fall below farmers' opportunity costs and firms would leave the industry. Hence, in Figure 15 the equilibrium quantity will be 100,000 tons (at the intersection of supply and demand). If each firm produces 100 tons, the equilibrium number of firms will be 1,000.

Figure 15 Marketing Board Acting as a Monopolist



2. The Cost of Monopolising a Competitive Industry

What would happen if the industry described in Section 1 was to be monopolised? In order to focus strictly on the effects of monopolisation itself, and to ignore any gains that might arise from economies of scale (for example from bulk purchases of inputs) or losses arising from diseconomies of scale (for example from difficulties of managing a large enterprise), imagine that the government requires that the farms described above sell their product through a marketing board (which is directed to act as a monopolist).

In this case, the average cost curve for the product - a horizontal line at $2,000 – will not be affected. That is, the supply curve from Figure 15, S, can be used to represent the marketing board's cost curve. At the same time, it seems reasonable to assume that the demand curve for the product will not be affected – consumers will not care whether they are buying from individual

farmers or from a marketing board - and the demand curve for the latter will be D, also from Figure 15.

Because the marketing board is a monopolist, it can set any price that it wishes. If its goal is to maximise economic profits, it will choose a price above $2,000, as at every such price, price will exceed cost. Assume, for example, that it chooses a price per pound of $2,200. From Figure 15, it is seen that, at that price, consumers will purchase 90,000 tons. If the number of farms does not change, the marketing board will allow each of the 1,000 farms to produce 90 tons, implying a profit per farm of $18,000 (= 90 x ($2,200 -$2,000)). (The 90 ton limit here is usually known as the farm's *quota*.)

Note that each firm's profit increased because, although its revenue decreased by $2,000 (from $2,000 x 100 tons to $2,200 x 90 tons), that decrease was more than made up by a $20,000 decrease in its total cost (from $2,000 x 100 tons to $2,000 x 90 tons).

The profit maximising rule for the marketing board is to increase its price as long as the incremental gain, from reduced expenses, exceeds the incremental loss of revenue. As it happens, in Figure 15, this will occur at a price of $3,000 and quantity of 50,000 tons. (You can check this by asking what would happen to profits if the price was to increase to $3,200, at which quantity demanded would be 40,000 tons.)

At $3,000, the marketing board will require that each of its member farms decrease output, from the 100 tons they were each making when the industry was competitive to 50 tons. At that price-output combination each farm will make an economic profit of $1,000 per ton (which equals the price per unit of $3,000, minus the cost per unit of $2,000), or $50,000 in total.

It might seem that $50,000 is not a particularly large profit. But remember that that is the farmer's *economic* profit – its profit in excess of the farmer's opportunity cost (the income the farmer would have earned in his/her "next best" occupation). Imagine, for example, that that income was $75,000. That figure was one of the components of the $200,000 it had cost to produce 100 tons. At that output, the farmer was earning zero economic profits, so his/her total income was the $75,000 already incorporated into total cost. At 50 tons, the farmer's opportunity cost is still $75,000, but now economic profits are $50,000, so the farmer's total income has increased to $125,000. In this light, an economic profit of $50,000 per year becomes quite substantial.

In terms of economists' traditional focus on the "allocation of scarce resources among competing ends", the social cost of this monopoly is that, at 50,000 tons, consumers revealed that the value to them of additional units of this product, $3,000, exceeded the opportunity cost of producing those units, $2,000 – they would rather have more of *this* product than more of the products that could be produced with the inputs required to make this product. But, in order to maximise its own profits, the monopolist/marketing board has chosen to restrict output.

MODULE 26 APPLICATION - FURTHER COSTS OF MONOPOLY

Although the misallocation of resources discussed in Module 25 represents an important social cost of monopoly, economists have also been investigated a number of other costs. This module briefly discusses three of these.

1. Objectives other than Profit-maximisation

At the equilibrium price in a competitive market, the highest profit that firms can earn equals the opportunity costs of their owners' investments of time and money. (If profits were higher, additional firms would be induced to enter the industry, driving prices down.) Hence, each firm will have an incentive to take all steps necessary to ensure that its profits are maximised, as at any lower profit level it will not cover its opportunity costs and the owner will be better off employing his/her resources elsewhere.

In a monopolistic industry, on the other hand, the firm can expect to earn profits that exceed its opportunity costs. Thus, even if monopolistic firms earn less than the maximum possible, their owners may still earn profits that far exceed what they could earn in their next best activity. In such cases, owners might be willing to accept lower-than-maximum profits in exchange for the opportunity to use their resources in ways that are not socially optimal.

1.1 Satisficing

One possibility is that owners might choose to do just enough work to produce a "satisfactory" level of profit – an approach that is sometimes called *satisficing*. They may find it easier and less stressful to maintain the status quo than to work hard trying to find ways to improve their products. For example, monopolists might not work as hard as owners of competitive firms at ensuring they find the lowest possible costs of producing their goods; and they might devote fewer resources to developing products that their consumers find desirable.

1.2 Indifferent customer service

Another manifestation of satisficing is that monopolists may not work very hard at providing customer service. The competitive firm that provides slow service or fails to respond to customer complaints about poor quality will soon find that other firms will have stolen its clients. The monopolist, however, does not need to concern itself with the possibility of losing customers to other firms as there are no other firms.

1.3 Sexual and racial discrimination

If employers show a preference for hiring one group of workers, say white males, when a second group, say black females, has equal skills, the wages for the first group will rise relative to those of the second. Hence, in a competitive industry, those employers who hire primarily from the first

group will have higher costs than will employers who hire primarily from the second. All else being equal, the latter employers will be able to undercut the prices of the former and drive them out of business. In this case, economists argue that it will be difficult for employers to discriminate among workers of equal productivity.

If monopolistic employers have a "taste for discrimination," however, they may be willing to accept reduced profits in exchange for the ability to discriminate against minority groups – women, blacks, homosexuals, immigrants, etc. – and those reduced profits may still be significantly higher than the employers' opportunity costs. Thus, even if minority workers have lower wages than do workers from the majority group, there will be no competitive pressure from other firms to hire workers from the minority and discrimination will continue.

(The causes of wage discrimination will be discussed in more detail in Modules 43 and 44.)

2. Lobbying

Because monopoly provides the opportunity to earn substantial profits, there is a strong incentive for entrepreneurs to obtain monopoly power and, once such power has been gained, to maintain it. As such power is often granted, or taken away, by government regulation, substantial amounts of money are often dissipated in lobbying activities. As these activities generally provide little or no advantage to society, any such expenditures may be considered to be a cost of monopoly.

3. Bribery

An extreme form of lobbying is bribery – instead of just arguing for preferred treatment, monopolists might pay bribes to government officials to purchase the outcomes they desire. Whereas in autocratic countries such bribes are often overt, in democratic countries monopolists (or those seeking monopoly powers) may resort to indirect approaches, such as promising to hire government employees at exorbitant salaries after they have left government, in return for favourable rulings while they are in government.

MODULE 27. APPLICATION -

SOME PRACTICAL IMPLICATIONS OF MONOPOLY

Module 25 used a simple model to argue that monopolies misallocate resources, in the sense that they produce less output than consumers are willing to pay for. Nevertheless, society is often willing to countenance monopolies because they provide benefits that are not easily measured in the simple model. This Application identifies some of these benefits and investigates the policies that have arisen in response to the tension between the benefits and the costs of monopoly.

This analysis will be divided among four sources of monopoly: the three that were mentioned in Module 25 – patents, utilities, and marketing boards – and one that has become of particular importance in recent years – that arising from *network externalities*.

1. Patents

Even though "patent monopolies" charge prices that exceed the opportunity cost of production, and will produce fewer units than are socially desirable, it may still be to society's advantage to offer patents to firms that invent new products. The reason for this is that there are often very high costs to invention, with a high risk that those investments will not be successful. Patents are a way of rewarding this investment.

Once a product has been patented, a significant portion of the difference between the product's price and its cost of production is a return on the inventor's investment in innovation. Although a drug manufacturer, for example, might charge $10 for a pill that costs $0.10 to make, much of the remaining $9.90 may represent compensation for the billions of dollars that had been spent on research and development.

This raises a public relations problem for the manufacturer: it becomes very tempting for critics to claim that a $10 price for a product that costs $0.10 to manufacture is "excessive," and to demand that prices be rolled back – a claim that is particularly appealing to voters when the product in question is a drug or medical procedure that has substantial health benefits.

Governments have responded in four ways to the demand for price controls on patented goods:

Maintain the status quo: The government might simply argue that the long-term costs of reducing monopoly prices – that entrepreneurs will devote fewer resources to the development of new products – outweigh the short-term benefits of reduced prices.

Subsidise private research: The government could subsidise private firms' research into the development of new products, in return for an agreement that firms will accept lower prices for those products once they have been brought to the market. This approach encounters two problems. First, when new products are in the development stage, it is not obvious which ones should be subsidised. As government agencies are not in the business of inventing new products, they may

waste a lot of money on unpromising schemes. Second, once a product has been developed, it may not be clear what the price level is that will appropriately compensate the manufacturer for its investment.

Government-run investment: The government could operate its own research facilities. The primary problem with this solution is that it is often difficult to identify a reward scheme that will induce its employees to invest the desired time, effort, and resources into the development of the most beneficial products. [The difficulties of constructing incentive schemes for government employees will be discussed in Section G.]

"Prize" for innovation: An interesting idea that has been suggested with respect to drugs is to replace patents with a "prize" for the development of new products. In this scheme, a world-wide pool of prize money would be created and each year the agencies that had developed the most valuable new products would be awarded substantial monetary rewards. As these rewards would replace patents, all drugs would be sold at prices that reflect their costs of production, rather than their current monopoly prices.

2 Utilities

In the past, the reason governments were willing to accept the monopolisation of utilities, like the delivery of electricity or natural gas to residential areas, was that it was wasteful to have more than one electrical line or gas pipeline on any street. Recently, however, governments have created a certain amount of competition in these industries by separating production from delivery.

For example, one company might be given a monopoly over the construction and operation of power lines that deliver electricity; but that company would be required to transmit electricity provided by any producer, at a regulated price. Each producer would then be free to sell its electricity to any consumer. (As electricity is a homogeneous product, each producer would simply be required to enter as much electricity into the grid as it had contracted to sell – there would be no need to ensure that the electricity consumed by one of its customers was the "same" electricity the producer had added to the grid.)

In this approach, producers are free to generate electricity using any method they prefer – coal fired plants, gas generators, hydroelectric plants, wind farms, etc – and charge any price that consumers are willing to pay. As there are often numerous producers on a single electricity grid, competition may be fierce among producers, creating an outcome that approximates a competitive market.

A similar approach could be taken with respect to railways. Instead of each railway firm building its own rail system, ownership of the railways could be separated from ownership of the tracks. Numerous railways could then contract use of the tracks from the track's owner. This way, even if geography limited the number of tracks that could be built in an area, there would be competition among railway companies.

The primary difficulty with this type of utility market is that the company that transmits the product is still a monopoly - whether it is delivering electricity, natural gas, water, telephone services, railway tracks, or cable TV. To avoid the misallocation of resources associated with monopoly, governments usually respond either by regulating the transmission firms very closely or by running those firms as government agencies.

3 Marketing boards

When governments allow the creation of marketing boards, for example for chickens, milk, or cheese, they usually do so because they have decided that the misallocation of resources that arises from increased prices is offset by the reallocation of income to low income farmers. Hence, the public policy issue that arises is not: "how should the government reduce the misallocation effect of monopoly?" as that misallocation has already been accepted. Rather, the policy question is: "does the marketing board raise the incomes of farmers?"

The answer to that question is, perhaps surprisingly, that farmers' incomes are only increased in the short run. To see this, return to the example that was introduced in Part 2 of Module 25. There, at the profit-maximising price of $3,000 per ton, each firm produced 50 tons at a cost of $2,000 per pound. Thus, each firm earned economic profits of $50,000 per year, which would amount to $2,000,000 over a 40-year period (perhaps the "life" of the farm).

Any new firm wishing to enter this industry will now have to pay existing firms something in the neighbourhood of two million dollars just for the right to belong to the marketing board (*plus* they would have to pay for the cost of land, buildings, and machinery - perhaps another two million dollars)! These new firms will earn zero economic profits, because the high price they pay to enter the industry must now be included in their opportunity costs. The government's goal of raising the incomes of farmers will only be reached for the first generation of farmers. All subsequent generations will earn the same net incomes that had been available before the marketing board was established, yet consumers will pay the higher prices created by the marketing board.

4 Network externalities

A positive network externality occurs when the benefits to one user of a product increases as the number of other users increases. The classic example is the telephone: when there is only a small number of users, the telephone may be of little use; but as the number of users increases so too does the value of the telephone system to each user. Similar effects have been noted for fax machines, and email. In each case, if the first firm to enter the industry manages to develop a large customer base before other begin to compete with it, its products will be more valuable to potential customers than will those of the new entrants (who do not yet have a "network" of users). This may make it so difficult for new firms to compete that the first firm will obtain significant monopoly power.

This appears to be the situation for Facebook. Government regulators have argued that Facebook's first entrant advantage has been so great that it has virtually become a monopoly. It is not clear yet, how regulators will deal with this power.

MODULE 28. IMPERFECTLY COMPETITIVE INDUSTRIES

Competition and monopoly inhabit opposite ends of a spectrum that ranges from industries characterized by vast numbers of small firms, each producing the same product (competition), to those that have only one firm, producing a unique product (monopoly). In between, however, lies a large set of industries in which firms each produce a product that is differentiated in a meaningful way from those of its rivals[7]. These industries, referred to as *imperfect competition*, are generally divided between two types: those in which there are many small firms - *monopolistic competition* – and those in which there is only a small number of relatively large firms. - *oligopoly*.

The most important characteristic that distinguishes imperfectly competitive industries (whether monopolistic competition or oligopoly) from competitive ones is that they sell *differentiated products*. This differentiation arises from variations in *product content* and *location*.

1. Product Content

Although General Motors and Toyota are rivals in the automobile industry, they each do their best to convince consumers that their trucks and cars are different from the others'. This attempt at differentiation can extend down to the smallest of firms, such as pizza parlors, plumbing contractors, and beauty parlours.

2. Location

Products can also be differentiated based on the location of the firm's retail outlets. For example, consumers may prefer one gasoline station or coffee shop over another based, not on the characteristics of its products, but on its proximity to the routes that consumers follow from home to work. Drivers who normally take 16th Street to work often show a distinct preference for gasoline stations on that street over those on, say, 18th Street; and commuters who walk along 5th Avenue on the way from their subway stop to their office will rarely deviate to 6th Avenue to purchase coffee when there is a satisfactory shop on 5th.

3. Impact

Through differentiation, the firm hopes to convince its customers to continue buying from it even if it raises its prices above those of its rivals. That is, it hopes to create a downward-sloping demand curve for its products, unlike a competitive firm, which would lose most, if not all, of its customers if the price of its product was to rise even slightly above the market price.

[7] Imperfectly competitive firms are often referred to as "rivals," rather than "competitors," to distinguish them from firms in competitive industries.]

MODULE 29. MONOPOLISTICALLY COMPETITIVE INDUSTRIES

1. Characteristics

As the name implies, monopolistically competitive industries have characteristics in common with both monopoly and competition.

Similarities to monopoly: Like a monopoly, each monopolistically competitive firm attempts to distinguish itself in some way from its rivals, in order to capture a portion of the market. It might, for example:

- Create a product that was preferred by its customers to those of its rivals. Starbucks, for example, was able to charge more for its espresso-based coffees than were firms that relied on the traditional drip coffee.
- Offer its product at a lower price than is charged by its rivals. This, for example, was the source of McDonalds' early success.
- Provide superior customer service. At one time, once you checked your car into a service station for repairs, you were left on your own to get to work or home. Now, service stations provide you with a ride to your office or home, and pick you up when your car is ready.
- Move their offices to locations that are more convenient for their customers than are the offices of their rivals; or open at hours that are more convenient.

Each of these actions may provide the monopolistic competitor with a slight amount of protection from its rivals' predation; and, therefore, may allow it to charge a price that is above its costs.

Unlike monopolies, however, monopolistically competitive firms are unable to rely on artificial barriers to entry, like patents and marketing boards.

Similarities to competition: Like competitive firms, monopolistically competitive firms are assumed to be unable to obtain significant economies of scale– even small, independent coffee shops and accounting firms are able to keep their costs low enough that they can compete with the largest, international firms in their industries. Furthermore, as small firms can compete with their larger rivals, entrepreneurs will find it relatively easy to raise enough capital to allow them to enter a monopolistically competitive industry. Thus, just as was assumed with respect to competitive industries, there will be few barriers to entry.

So many industries meet these criteria that it is likely that the most common type of firm that you will encounter in your daily activities is a monopolistic competitor. These include firms in:

- The personal service sector – hairdressers, plumbers, lawyers, physiotherapists;
- The food industry – manufacturers of potato chips, mustards, wines;
- The clothing sector – both manufacturers and retailers of jeans, shoes, coats; and
- The restaurant industry – from hot dog stands to haute cuisine restaurants, Asian to African foods

2. Behaviour

Because there are limited barriers to entry and few economies of scale, even small firms can enter monopolistically competitive markets. Thus, economic profits are driven towards zero, just like in competitive industries. The only way that firms can obtain above-normal profits is either to identify a niche market in which they can become one of a relatively small number of competitors or to find a new way of reducing costs.

For example, if there is an untapped market for Asian food, entrepreneurs may open Chinese food restaurants. If they have guessed correctly, the first entrants into the market will have the market to themselves and will be able to earn economic profits. Soon, however, new firms will enter, and profits will be driven towards zero. But then an entrepreneur may notice that most of the firms in the market are offering food from the same area in China and, therefore, will attempt to distinguish its product by offering food from another region. Again, if this generates economic profits, additional firms will enter. Then new firms might expand beyond Chinese foods into Vietnamese, Thai, Malaysian, and Indonesian cuisines; and then regional variations of each of those cuisines.

Alternatively, those entrepreneurs that are unable to identify new products, may distinguish themselves by finding lower cost methods of production, as McDonalds did.

In each case, early entrants are rewarded, with economic profits, for having provided a product that is valued by consumers at the lowest cost possible. This reward is like the profits that drug companies obtain in return for the costs they have incurred for research and development. The innovative restaurant receives above-normal profit in return for having taken the risk of identifying, and investing in, a valuable product that had not previously been available to consumers.

In the long-run what we see in each monopolistically competitive industry is that, although the first firms to identify a new niche may earn profits above the owners' opportunity costs, the firms that subsequently fill the remaining market will earn zero or close to zero economic profit.

3. Implications of Monopolistic Competition

Advantages: Monopolistic competition offers a solution to the question that bedevils central planning agencies: how to decide which goods and services should be produced, given that planners know little about either the strength of consumers' demands for alternative goods or the opportunity costs of producing those goods. Under monopolistic competition, consumers reveal their preferences through their willingness to accept the prices offered in the market; the opportunity costs of goods are revealed through the prices charged by the producers of inputs; and entrepreneurs are induced to supply any good for which a profit can be made. No central plan needs to be made. Rather, each participant in the market merely needs only to follow its own goals.

Disadvantages: There is an incentive for monopolistically competitive firms to collude with their rivals to manipulate prices. Fortunately, however, the ease of entry to monopolistically competitive

industries makes this practice difficult to implement, as any success at raising prices to the point that firms were earning economic profits would attract new firms into the industry – firms that were not parties to the initial collusive action.

The primary disadvantage to monopolistic competition is that some firms will be able to devise sufficiently distinct products that they will be able to charge prices that exceed opportunity costs – that is, they will, in a limited way, be able to act like monopolies. In most cases, this ability will be constrained by the potential for rivals to develop similar, competing products, thereby limiting the profit-making capacity of the innovator. Nevertheless, the first firms into a sector may be able to expand quickly enough that they can capture limited economies of scale, from production and advertising, thus allowing them to continue earning economic profits even after new, smaller firms have entered the market.

Summary: Monopolistic competition offers the distinct advantage that it encourages the development of new products at the lowest possible cost, an advantage that may outweigh the cost arising from the possibility that some firms will be able to restrict output below the optimal level.

MODULE 30. APPLICATION - FRANCHISES

Many monopolistically competitive firms, particularly in the fast food and retail clothing sectors, find that the most efficient way to expand is to increase their numbers of outlets. 7-11, for example, has over 60,000 locations worldwide, McDonalds and Subway each has over 35,000, and Gap and H&M over 3,500. This allows these firms to obtain economies of scale, primarily from the bulk purchase of inputs and from the aggregation of their advertising campaigns.

But they also face diseconomies of scale because there is a disconnect between the goals of the owners and those of the managers of the individual outlets. If the managers are paid monthly or annual salaries, for example, they may have little incentive to ensure that the employees in their outlet work hard, that customers are treated with patience and respect, or that costs of locally sourced inputs are kept as low as possible. And, as the number of the firm's outlets increases, the firm's ability to influence the individual managers' behaviour decreases.

This difficulty is exacerbated by what economists call *asymmetric information*. To deal with this problem, firms are seen to develop *self-enforcing employment contracts*, of which *franchises* are one type. Each of these concepts is discussed here.

1. Asymmetric Information

When A hires B to perform a service, A is called the *principal* and B the *agent*. For example, in a real estate contract, the person selling a property is the principal and the person who has been hired to find a buyer is the (real estate) agent; and when a firm hires a worker, the employer is the principal and the employee is the agent.

In principal-agent relationships, the principal wishes to ensure that the agent will work as diligently as possible to obtain the principal's goals. So, we might imagine that the contract between them would specify that the agent would follow those goals. But notice that the agents will usually be much better informed about how diligently they have worked than will the principals. In the example at the beginning of this Application, for example, the outlet manager will know much more than the firm's head office about how much effort she put into ensuring that her employees worked hard, that her customers were treated well, and that she sourced the least expensive and highest quality local products. This difference in knowledge is referred to as asymmetric information.

A number of types of contracts have been developed to ensure that agents will act in their principals' interests, even when the principals are less informed than their agents. A simple form of such a contract might tie managers' salaries to measures of managerial performance, such as employee turnover (a measure of the manager's human resource skills), customer satisfaction surveys, and complaints from suppliers.

But the problem is that these are only imperfect indicators of managers' efforts to maximise the firm's profits, in part because they omit many aspects of managerial performance that are of importance to the firm, in part because many of them can be manipulated by the manager, and in part because they do not measure managerial performance directly. Accordingly, a number of more sophisticated contractual relationships, called "self-enforcing contracts," have been developed.

2. Self-enforcing Contracts

A self-enforcing employment contract is one whose terms are such that it is simultaneously in the interest of employees to act in a way that maximises the firm's profits and in the interest of the firm to compensate its employees for any extra effort that is required of them.

For example, assume that if managers do as little as they can, the stores that they manage will earn $50,000 per year; but that if managers work diligently, their stores will earn $75,000. Assume also that the managers will consider themselves to be fully compensated for working diligently (instead doing as little as possible) if they are paid an extra $15,000.

It can be seen that it will be to both parties' self-interest to agree to a contract in which the managers agree to work diligently in return for a promise by the firm to pay them anything between 60 and 100 percent of the profits that exceed $50,000. For example, the parties might agree that managers will receive 80 percent of any incremental profits (and that the firm will receive 20 percent). In this case, managers will receive $20,000 if they work diligently (out of the assumed incremental profit of $25,000). By our assumptions, that is sufficient to compensate them for their extra effort. At the same time, the firm will experience a $5,000 increase in its profits.

Would such a contract be self-enforcing? That is, would the parties conform to the requirements of their agreement even if it was not enforced by a third party (say, the courts)? The answer is "yes." The managers will live up to their promise, of working diligently, because by doing so they will earn more than enough to compensate them for their extra effort. And the firm will (probably) live up to its promise (to pay the managers 80 percent of the incremental profit) because failure to do so would discourage future managers from agreeing to similar contracts, and therefore would prevent the firm from earning extra profits in the future.

3. Franchises

A popular form of self-enforcing contract among monopolistically competitive firms is the *franchise*. In this case, the firm, called the *franchisor*, sells the right to operate one or more of its outlets to a third-party manager, called the *franchisee*. In return for payment of a purchase price and annual *royalties* (usually a percentage of the outlet's profits), the franchisee benefits from the firm's regional and national advertising and receives the right to use the seller's name, its methods of doing business, and its products.

To the franchisee, the primary advantage of this relationship is that she is provided training and advice that would otherwise have to be obtained through lengthy trial and error; and that she will benefit from economies of scale the franchisor can obtain by buying materials and supplies in bulk, and by advertising nationally. To the franchisor, the franchise arrangement helps to avoid the diseconomies of scale that would be incurred if it used employee-managers to run its outlets.

From a societal point of view, the disadvantage of the franchise arrangement is that franchisee-agents do not have the same incentive to respond to consumer preferences as would the principal if it operated its outlets itself. The reason for this is that, whereas the principal would receive 100 percent of any profits it obtained by responding to customer demand, the franchisee will receive only some fraction of that amount (as specified by the royalty agreement). For example, if the franchisee receives only 50 percent of any incremental profit that its outlet generates, the incentive for the franchisee to respond to customers will be only 50 percent as great as would the firm's incentive had it maintained 100 percent control over revenues, costs, and profits.

MODULE 31. OLIGOPOLY - INTRODUCING GAME THEORY

Industries that are composed of a relatively small number of firms are referred to as *oligopolies*, from the Greek "oligos: few" and "polein: seller." Most such industries are characterised by economies of scale, differentiation of products, and interdependence of decision making. This module first investigates the importance of these three characteristics; and then asks how they affect the prices that will be chosen by oligopolistic firms.

1. Characteristics

1.1 Economies of scale

The primary characteristic of an oligopolistic industry is that all firms obtain economies of scale over very large ranges of output. As a result, only a small number of firms will be big enough to operate at the lowest possible average cost. (Once those firms have each grown large enough to minimise their costs, they are producing enough output to meet industry demand, leaving no room for additional firms.) Typically, oligopolistic industries - like automobile manufacturing, airlines, and supermarket chains – are composed of five to eight large firms, with perhaps a fringe set of smaller firms that fill niche markets.

1.2 Product differentiation

Typically, the firms in an oligopolistic industry produce goods that are not perfectly interchangeable. Chevrolets have different characteristics from Toyotas, Samsung smartphones differ from Apple iPhones, and Levis' jeans are not perfect substitutes for Gap's. The most important implication of this characteristic is that, unlike competitive firms, when oligopolistic firms raise their prices above those of their rivals, they may be able to retain a significant portion of their customers.

1.3. Conjectural interdependence

When an industry is composed of a small number of firms, economists say that those firms are *conjecturally interdependent* upon one another.

By interdependent we simply mean that each firm is aware that the actions of its rivals will have important effects on its own ability to maximise profits. For example, if firm A's rivals increase their advertising budgets or reduce their prices, to attract additional customers, it is likely that many of those customers will be drawn from A.

By conjectural we mean that when an oligopolistic firm is deciding whether a change in its behaviour will be to its advantage, it has to speculate (or conjecture) about how its rivals will respond. For example, if it thinks that it would be able to increase demand for its products by engaging in research that will improve those products, it will recognise that its rivals may respond to its decision by increasing their own research.

Thus, before it makes any substantive decision, each firm will try to predict how its rivals will respond to that decision. Economists use *game theory* to investigate how firms behave when they need to consider the reactions of their rivals.

2. Pricing Behaviour

2.1 Introduction to game theory

Consider a very simple industry composed of two dominant firms, A and B, (a "duopoly"). For example, A might be Airbus and B Boeing. Both firms are currently charging $50 million per unit (airplane) and are making $5 billion in profits per year. Assume also that if they were both to raise their prices to $60 million per unit, their profits would increase to $6.5 billion. If they were able to collaborate with one another, it seems likely that they would be able to agree to such a price increase. But, normally, governments discourage collaboration of this sort. What we will ask here is: will duopolists raise their prices even if they cannot collaborate?

You might think that the answer was obvious: if these firms can increase their profits by raising their prices, they will do so. But the concept of conjectural interdependence alerts us to the possibility that each firm may be reluctant to make such a change, given that the level of profit it will make depends crucially on what its rival does.

In order to analyse the two firms' decision-making processes in this situation, economists use an approach known as *game theory*. A simple example of a game is described in Table 4. There, the vertical column on the left represents the firms' profits when A charges $50 million and the column on the right when it charges $60 million. Similarly, the two rows represent outcomes when B charges $50 million and $60 million, respectively.

Table 4: Pricing Game 1

		FIRM A	
		$50 MILLION	**$60 MILLION**
FIRM B	**$50 MILLION**	**$5.0 billion** **$5.0 billion**	**$4.5 billion** **$6.0 billion**
	$60 MILLION	**$6.5 billion** **$6.5 billion**	

In Table 4, the number in the top right corner of each "box" represents A's profits and the number in the bottom left corner represents B's profits. For example, as the top left box represents the situation in which both firms charge $50 million, the figure in the top right of that box, $5.0 billion, represents A's profits and the figure in the bottom left, also $5.0 billion, represents B's profits.

Assume that, initially, both firms are charging $50 million and that A is thinking about raising its price to $60 million. If it assumes that B will *not* change its price, it will expect to lose many of its customers to B. Although the resulting reduction in sales will reduce its profits, the increase in profit *per unit* (because A is now charging $60 million per unit instead of $50 million), will mitigate its loss of profit somewhat. Let us assume that it will now earn $4.5 billion per year. At the same time, B will experience an increase in sales and, therefore, an increase in profits, say, to $6.0 billion. The top right-hand box can now be filled in: if A raises its price and B does not, A will earn $4.5 billion (the top right number) and B will earn $6.0 billion (the bottom left number).

Finally, the bottom right hand box in Table 4 reports the profits the two firms will earn if both of them raise their prices to $60 million – which was assumed above to be $6.5 billion each.

Firm A might now reason as follows:

> "If I raise my price and B leaves its price constant, my profits will fall to $4.5 billion and B's will rise to $6.0 billion; but B will realise that if it raises its

price to match mine, it will earn even more: $6.5 billion (see the lower right-hand box in Table 4) and so it will follow me by raising its price. “

That is, A anticipates that if it raises its price to $60 million, B will follow. Hence, A will raise its price, B will follow, and both will have maximised their profits without having collaborated with one another.

2.2 Will oligopolists always raise prices in concert with one another?

In the example presented in Table 4, it was in Firm B’s interest to respond to A’s price increase by increasing its own price. But it is easy to construct a game in which B cannot be assumed to follow A’s lead. For example, in Table 5, B’s profit in the top right-hand box has been increased from the $6.0 billion used in Table 4 to $7.0 billion. In this case, B’s best option will be to hold its price at $50 million when A increases its price to $60 million – because if B now raises its price to $60 million, its profits will fall to $6.5 billion. And if A anticipates that B will not follow its price lead, A’s best option will be to hold its price at the initial level of $50 million. In that case, both firms will be “stuck” earning profits of $5.0 billion when cooperation would have allowed them to earn $6.5 billion each.

Table 5: Pricing Game 2

		FIRM A	
	$50 MILLION	**$60 MILLION**	
FIRM B			
$50 MILLION	**$5.0 billion** **$5.0 billion**	**$4.5 billion** **$7.0 billion**	
$60 MILLION		**$6.5 billion** **$6.5 billion**	

This is why oligopolists often have a strong incentive to form *cartels*, that is organisations which coordinate actions among rivals, to allow them to maximise their profits. Cartel behaviour will be discussed in Module 33.

2.3 Will oligopolists lower their prices when that behaviour will reduce their profits?

We saw, in Part D, that if competitive firms are earning economic profits, new firms will enter the industry, increasing output and driving down both prices and profits. This is generally a desirable outcome from society's point of view, as consumers will gain from both greater output and lower prices. What will be asked here is whether oligopolists can also be expected to lower prices and increase output when they are earning economic profits. The answer (in the simple model developed here) is that, if a reduction in price leads to a reduction in profits, oligopolists will (usually) *not* lower their prices.

Imagine, in Table 4, that the initial position is that both A and B are charging $60 million and earning $6.5 billion (i.e. that their initial position is represented by the lower right box of that table). If they lower their prices to $50 million, their profits will decrease, to $5.0 billion (but will still be well above each firm's opportunity costs). Firm B realises (i) that if it lowers its price to $50 million and A does not change its price, B's profit will fall to $6.0 billion and (ii) that Firm A would retaliate to B's reduced price by lowering its own price (to $50 million), causing B's profit to fall even further, to $5.0 billion. Thus, B has a double reason to hold its price at $60 million. The firms cannot be assumed to reduce their prices, even though that would be in society's interests.

Alternatively, assume that the firms' positions can be described by the data in Table 5. If both firms initially charge $60 million, B might reason as follows:

> "If I lower my price to $50 million, my profits will increase to $7.0 billion and A's will decrease to $4.5 billion. But, from that point, A could increase its profits to $5.0 billion by matching my price decrease (that is, by lowering its price to $50 million also). In that case, my profits would fall to $5.0 billion also. So, my optimal policy is to hold my price at the initial value, $60 million."

As long as B recognises that A will earn higher profits by lowering its price to $50 million when B charges $50 million, B will not lower its price. As A can be expected to follow similar reasoning, again neither firm will reduce its price.

It is possible to construct a scenario in which it will benefit one firm to lower its prices and for the second to hold its prices constant. For example, if A's profits were $5.5 billion when it charged $60 million and B charged $50 million, A would hold its price at $60 million when B reduced its price to $50 million from $60 million. But, given the assumptions at the beginning of this section, there is no scenario in which both firms will lower their prices.

3. Implications

In the simple model that was described in this module, oligopolists will sometimes raise their prices, without collaborating, when doing so would increase their profits; and will rarely lower their prices when doing so would reduce their profits. They will have earned their reputation for charging "too much" and for producing "too little."

But these predictions are based on (at least) two implicit assumptions that may not be true in real world situations: First, they assume that firms can easily observe what their rivals are doing, and react accordingly; and second, that oligopolies consisting of, say, eight or ten firms, will behave similarly to those that consist of only two. The implications of altering these assumptions will be discussed in the next two modules.

MODULE 32. OLIGOPOLY WITH INCOMPLETE INFORMATION

In Module 31, it was assumed that the two firms were well-informed about their rivals' actions. Firm B, for example, could easily observe whether Firm A had raised or lower its prices. Using this information, it could act quickly to counteract any negative effects of A's policies.

But in many cases, firms have only *incomplete information* about their rivals. Two examples are particularly important. First, if producers sell to large intermediaries – for example, if oil companies sell to pipelines or refineries – it may be difficult for rivals to obtain information about the terms of the contracts between these firms. Second, firms may be able to keep information about their investments in research and development secret from their rivals.

1. Incomplete Information about Rivals' Prices

Many agricultural products, such as wheat, coffee, and cocoa, are sold by national marketing agencies, each of which acts like an oligopolist. Assume that the price of such a product is $6.00 per kilogram and that, at that price, countries A and B each earn profits of $100 million. If A reduces its price to $5.50 and B keeps its price at $6.00, A's profits will rise to $120 million (because it can attract customers away from B), and B's profits will fall to $75 million. Assume also that if both firms lower their prices to $5.50, they will each earn profits of $85 million. This information is summarised in Table 6.

Following from the analysis in Module 31, it can be seen that if B observes that A has reduced its price (that is, if B has *complete* information about A's pricing), B's best response will be to reduce its price also – because when A charges $5.50, B earns higher profits if it charges $5.50 ($85 million) than if it holds its price at $6.00 ($75 million).

But if B has *incomplete* information about A's actions – for example, if A can do an under-the-counter deal with one its major customers – A will be able to earn $120 million until B discovers what has happened. The opportunity to earn that much in the short run may be sufficiently attractive to A that it will consider lowering its prices even though it knows that B will eventually realise what A has done and will lower its prices too.

Table 6: Pricing with Incomplete Information

		FIRM A	
	$6.00	**$5.50**	
FIRM B			
$6.00	**$100 million** **$100 million**	**$120 million** **$75 million**	
$5.50	**$75 million** **$120 million**	**$85 million** **$85 million**	

Assume, however, that B reasons in a similar way: it thinks that if *it* secretly lowers its price, it will be able to increase its profits significantly until A discovers what has happened. Will A and B both lower their prices – even though both realise that the result will be that both will earn lower profits than they had initially? Surprisingly, perhaps, the answer is yes.

The mathematician, John Nash, developed *game theory* to answer this question. He argued that if neither firm knew what its opponent was doing (because, in this case, neither knew whether or not the other had decreased its prices), each firm would first ask: "for each possible action by my rival, what would my own best action be?" In Table 6, it can be seen that, regardless of whether A knows if B has lowered its price, A's best action is to lower its price. If B has *not* lowered its price, A could increase its profits, from $100 million to $120, million by reducing its own price; and if B *has* lowered its price, A could again increase its profits, from $75 million to $85 million, by reducing its price. Similarly, it can be seen in Table 6, that B's best action, regardless of what it thinks A has done, is to reduce its own price.

Thus, contrary to the result that was obtained when it was assumed both parties had complete information about their rivals' actions, when parties have incomplete information, they may be induced to reduce their prices, even when doing so results in lower profits.

2. Incomplete Information about Rivals' Investments in Research

Although firms may find it difficult to keep their pricing policies secret from their rivals, it may be possible to disguise how much they are investing in research and development. Hence, each firm will be concerned that its rival will suddenly announce that it has made a dramatic breakthrough that will give it a distinct advantage. For example, Airbus might secretly develop a new technology that make its planes more fuel efficient than Boeing's; or Samsung might develop a new smart phone that finally gave it an edge over Apple. When firms are concerned about this possibility, game theory suggests that they may each invest more in research than would be consistent with profit maximization.

Table 7: Investment with Incomplete Information

		FIRM A	
	Do Not Invest	**Invest**	
FIRM B			
Do Not Invest	**$2.0 billion** **$2.0 billion**	**$2.5 billion** **$1.0 billion**	
Invest	**$1.0 billion** **$2.5 billion**	**$1.5 billion** **$1.5 billion**	

For example, in Table 7, Firms A and B are both assumed to earn $2.0 billion profits if they do *not* invest in new technology and $1.5 billion if they *both* invest. If one of them invests and the other does not, the one that invests will earn $2.5 billion and the one that does not will earn $1.0 billion.

In this case, if neither of them knows whether the other is conducting research, each of them might reason as follows: If my rival does *not* invest in research, my best policy is to invest, as my profits will rise from $2.0 billion to $2.5 billion; and if my rival *does* invest in research, again my best policy is to invest, as my profits will rise from $1.0 billion to $1.5 billion. Both firms may be driven to invest even when the result is lower profits for both of them.

3. Implication of Incomplete Information

Because the firms analysed above had incomplete information about their rivals, they were induced to choose actions that reduced their profits. One implication of this result is that firms can be expected to attempt to collude with one another to avoid these actions. This collusion can result from communication as informal as lunch between CEOs at their golf club; or as formal as international treaties among countries or contracts between firms. The latter examples, known as *cartels* will be discussed in Module 33.

MODULE 33. APPLICATION -

COLLUSION AMONG OLIGOPOLISTIC FIRMS

In Module 31, it was argued that if oligopolists do not cooperate with one another, they may be unable to maximise their profits. For example, in Section 2.2 of Module 31, it was seen that if duopolists are well informed about one another's pricing behaviour, they may fail to raise their prices even when that would increase both firms' profits. And in Module 32 it was seen that if firms are *not* well informed about their rivals' behaviour, they may be induced to choose both prices and investment levels that reduce their profits.

Furthermore, the results in Modules 31 and 32 were obtained assuming there were only two firms in the industry. It can be imagined that, as the number of firms increased, each firm would face increasing difficulty predicting what its rivals would do in response to its own actions, making it difficult for firms to coordinate.

In the face of these issues, oligopolists often attempt to collude with one another; that is, to agree among themselves on policies that will maximise their joint profits. When they act in this way, they are said to have formed a *cartel*. OPEC, the Organization of the Petroleum Exporting Countries, is a clear example of such an organisation.

This module will examine: how cartels operate, why it is difficult for them to maintain cohesion, and how oligopolistic firms respond to these difficulties.

1. The Role of the Cartel

There are at least three different ways that cartels can help their members to maximise profits: collaborate, disseminate, and coordinate.

Collaborate: In Modules 31 and 32 it was seen that if oligopolists act independently, they may choose prices or investment levels that fail to maximise their profits. To avoid this, they might use a *collaborative* process to choose these factors. The duopolists described in Table 5, for example, might agree to raise their prices to $60 million; those in Table 6 to refrain from lowering their prices; and those in Table 7 to curtail their expenditures on research and development.

Disseminate: Even if firms that act independently are able to select a policy that *increases* their profits, it is not certain that they will be able to identify the policy that *maximises* their profits. The game described in Table 4, for example, assumed that the firms could choose between only two prices, $50 million and $60 million. In that case, game theory predicted that the firms would independently choose the higher of those prices.

But, in reality, these firms would be free to choose from a wide range of prices, say from $40 million to $75 million, and each of those prices would be associated with a different profit level. In this case, it is likely that two (or more) firms acting independently would encounter difficulty finding the price level that maximised their profit levels, especially if no firm was knowledgeable about the profits that its rivals would earn at each price level. In this case, the members of a cartel might agree to pool and *disseminate* the information that was needed to find their best price (or investment) level.

Coordinate: When an oligopoly consists of more than two or three firms, the number of options available may become so great that it will be difficult for individual firms to respond quickly to the changes that are taking place in their industry. One role of the cartel may be to allow firms to coordinate their actions in such a way that they do not inadvertently work against one another's interests.

2. Maintaining Cohesion

A cartel is only of value to its members if it can identify and enforce policies that increase its firms' profits; so, if firms can maximise profits without a cartel, firms will have little incentive to collaborate with one another. For example, in Module 31, it was argued that if each firm could easily observe any price changes its rivals made, firms would choose the profit-increasing increase in price even when they acted independently.

In Module 32, however, it was argued that if firms were not able to observe one another's actions, they might choose to reduce their prices even when that led to a reduction in their profits. For example, in Table 6, reproduced below, if Firm A does not know whether B has reduced its price from $6.00 to $5.50, John Nash argued that A would first calculate what its best action would be under two scenarios: (i) if Firm B held its price at $6.00; and (ii) if B lowered its price to $5.50. As can be seen from Table 6, A's best action in both cases would be to reduce its price to $5.50. If B reasons in the same way, both firms will lower their prices, even though that results in a reduction in their profits. In this case, it might be in their interests to form a cartel, and jointly agree to hold their prices at $6.00.

Table 6: A Pricing Game with Incomplete Information

		FIRM A	
	$6.00	**$5.50**	
FIRM B			
$6.00	**$100 million** **$100 million**	**$120 million** **$75 million**	
$5.50	**$75 million** **$120 million**	**$85 million** **$85 million**	

But this outcome comes with inherent difficulties. The most important of these is that not only will each firm's total profit be higher at $6.00 than at $5.50, but so will its profit *margin* – the profit earned on each additional sale. For example, assume that, in the absence of collusion, the firms had initially set their prices at $5.50, as Nash predicted, but that the cartel has induced them to raise their prices to $6.00. Assume also that at $5.50 each firm's profit margin had been $1.00, whereas at $6.00 it is now $1.50.

At this much higher margin, each firm will have a strong incentive to break the cartel agreement. Firm A might, for example, try to attract customers from B by offering an under-the-table price reduction, say to $5.50, with a profit of $120 million. If it could hide this price reduction from B for long enough, the extra profit it earned might be sufficient to compensate for the reduction in its profits that will result once B reduces its price in response.

In practice, this incentive proves to be so strong that, in the absence of an enforcement mechanism by which the cartel can ensure compliance with the cartel's policies, many cartels will fail. In theory, one enforcement mechanism might be to draw up a contract among members of the cartel. However, as such contracts are generally not in the public interest – they are designed primarily to raise prices and reduce output – the courts will refuse to enforce them. Hence, cartels are forced to rely on extra-legal means. Three such methods are considered in the following section.

3. Enforcing Compliance with Cartel Policies

Government regulation: Cartels are often observed to engage in extensive lobbying of governments, arguing for the introduction of regulations that they contend are in the public interest but whose primary effect is to reduce competition among firms. They might, for example, lobby for regulation of safety or pollution emissions, knowing that these regulations raise costs more for small firms, who lack economies of scale, than for large firms.

Self-enforcement by dominant firms: If there is one firm that dominates an industry, that firm might act either to "punish" its rivals for charging a price below that agreed by the cartel, or to support the cartel price by reducing its own supply. For example, it has been argued that Saudi Arabia has acted to police OPEC's agreements by threatening that, if the cartel's members breached that agreement, it will flood the market with oil, thereby driving the world price well below the level preferred by OPEC's members. And, in some cases, Saudi Arabia has acted as a facilitator, reducing its own supply of oil in order to ensure that the cartel price is supported, despite breaches by other members.

Mergers: Perhaps the simplest way for two firms to enforce a "cartel" price is for them to merge into one firm. This may be one reason why so many mergers have been observed among automobile firms in recent years; and it is why governments look unfavourably at such coalitions. On the other hand, the firms involved usually argue that a merger allows them to reduce costs, and therefore prices, through economies of scale.

F. MARKET ECONOMIES - SHORTCOMINGS

MODULE 34. EXTERNAL COSTS

Implicit in the discussion of market-based economies, in Parts C and D of these notes, were the assumptions that (i) the prices that suppliers charged for their goods reflected the full opportunity cost of producing those goods; and (ii) the prices that consumers paid for their purchases reflected the full value that they placed on those purchases. The first of these assumptions implied that when a firm paid $5.00 for the inputs into its product, the value of the alternative goods that could have been produced with those inputs did not exceed $5.00. The second assumption implied that when consumers paid $5.00 for a good, they must have placed a value of at least $5.00 on that good. These two assumptions together allowed us to conclude that when consumers voluntarily purchase a good, they are sending a signal that they prefer that good to any other that could be produced with the same inputs.

But these two assumptions do not always hold.

In some cases, firms or individuals are able to avoid paying the full costs that their actions impose on other members of society – those actions are said to have imposed *external costs*. Air and water pollution, global warming, and overfishing of the oceans are all examples of inefficiencies that arise when costs are external in this way.

In other cases, individuals and firms are not fully compensated for the benefits that they provide, in which case we say that their actions have provided *external benefits*. Underutilization of public transit and the recent re-occurrence of measles outbreaks are examples of issues that have arisen when benefits were external.

These types of costs and benefits are collectively known as *externalities*. This module will examine how external *costs* affect the allocation of society's resources. Module 35 will extend this analysis to external *benefits*. Modules 36, 37, and 38 will investigate some of the effects of externalities on the economy.

1. A Simple Example

Begin with a simple example. Assume that you have some waste that the city garbage collectors will not accept – perhaps you have torn down an old shed in your yard and you have a pile of wood to dispose of. You think of three things you might do with this wood. You could pay someone $200 to haul it away. You could throw it over the fence into your neighbour's yard. Or you could build it into a big pile and set it on fire.

Start with the second of these options first: throw it over the fence. The immediate problem with this approach is that one of the privileges that attaches to ownership of property is that the owner has the legal right to exclude other people from using that property. So, when you throw your waste over the fence, your neighbour will have the full force of the law behind him when he either (i) insists that you take your waste back and compensate him for the harm that you have done; or

(ii) offers to sell you the right to use his land as a disposal site. In either case, you will be required to pay the full costs of your actions. If those costs are less than $200, you will prefer to compensate your neighbour rather than hire a firm to haul away your waste.

But what about the third option? If you build a bonfire in your back yard, you will be able to dispose of your waste at very little cost to you. Why? Although the smoke from the fire will carry particulates into the air over your neighbour's property, and the properties of many other people downwind from the fire, those individuals do not "own" the air in the same way they own their land. Accordingly, they may not have a well-defined right to exclude others from using the air.

Implicitly, the rule is often: because no one owns the air, anyone can use it as they like. And where that is true, you do not have to compensate your neighbours for any harm caused by the smoke from your fire. This, of course, makes the third option (the bonfire) tempting – particularly, it might be imagined, when the individuals who have been harmed are spread over a large area and, therefore, are not known to you.

The situation in which no one owns the rights to the resource being used – here, the air – is one of the main sources of external costs. Individuals who create external costs to society do not have to compensate their victims because the victims have no rights over the resource that has been compromised. Two inefficiencies result. First, resources which are not owned are used in inefficient ways. For example, in the case described above, you appropriated the air for your own purposes, regardless of whether that was the most valuable use for it.

Second, because you do not bear the external costs of your actions, the market does not deter you from engaging in activities that cause such costs. For example, assume that you had been deciding between an action that would produce very little waste – say, repairing your shed – and one that would produce a considerable amount of waste – say, tearing down your shed and replacing it. If you know that you will not be required to pay for the external costs that would be caused if you were to burn any waste, you may be induced to replace the shed even if repairing it would have been the socially desirable action.

2. Two Sources of External Costs

It is sometimes useful to distinguish between two broad circumstances in which firms or individuals may not be required to pay for the full costs of their actions; that is, in which costs may be external to the party generating them. The first of these, discussed above, occurs when one or more of the resources being used has no owner. The second arises when resources are "owned by everyone."

2.1 Resources have no owner

The smoke from your bonfire will impose an *external* cost on your neighbours if no one owns the air into which that smoke drifts. Lacking ownership, your neighbours would have no legal method

of preventing your actions and, therefore, they could not demand compensation from you for any harms you cause them.

Similar effects arise with respect to water and wildlife. If chemicals that leach into a lake from nearby farms kill fish or pollute drinking water, the users of the lake will have no recourse against the farmers if they do not have ownership rights to the water. The *market-based* economy described in Parts C and D will provide no incentive for farmers to reduce their purchases of chemical fertilisers and pesticides. As a result, governments and other organisations may be asked to step in to correct this misallocation of resources. [It is these alternative methods of allocating resources that will be discussed in Modules 36 and 37.]

Other situations in which external costs arise if there is no clear owner of a resource include:

- *Noise pollution*: If you do not own the "airwave" rights to the air over your property, you may not be able to prevent your neighbour from playing loud music in her garden. And even if you did own those airwave rights, how high would those rights extend? If it was, say 200 feet, that would not allow you to obtain an injunction preventing airplanes from flying so low over your house that the noise of their engines would create a significant nuisance.
- *Greenhouse gases*: As no one owns the right to the troposphere (the part of the atmosphere where most greenhouse gases are trapped), the market economy does not provide a method for imposing costs on those who cause global warming – the costs of the latter's actions are external.
- *Habitat destruction*: Assume that users of land, such as farmers, house builders, and ski hill operators, expand into areas that provide habitat for endangered species. Many people would consider the harm that is done to those species to be a cost. But if no one owns wild species, there will be no market mechanism for requiring land users to take that harm into account. The result is that "too much" valuable wildlife habitat may be used for other purposes.
- *Overfishing*: As no one owns the rights to the fish in the open ocean, fisherman do not have to pay for the fish they catch. As a result, they may not take into account the effect that their catch today will have on future fish stocks, and stocks will decline.

2.2 Resources are owned in common – the tragedy of the commons

In many cases, it may be useful to think of resources as being owned equally by everyone. The classic example is a piece of land called the *common* that all inhabitants of a village can use for grazing their animals. (If you go to England today, you can still see villages built around a common.) Because each villager has an equal right to use the common, no villager can stop any of the others from using it as they wish. In turn, this means that when farmer A is deciding how many sheep or cows to graze on the common, he does not have to take into account what the effect would be on the other villagers. Having common ownership turns out to have the same effect on resource allocation as having no ownership.

This leads to the now-famous ***tragedy of the commons***. When farmer A grazes his sheep on the common, the *private* cost *to that individual* is that there will be less grass available for his sheep in the future. But grazing also imposes an *external* cost on all of the other farmers who share in

that land: there will be less grass available tomorrow for their sheep also. Two effects arise. First, each farmer overgrazes the common because he only takes into account the private costs imposed by his own sheep. Second, when each farmer sees that the others are overgrazing, he may become concerned that there will be insufficient grass left for his sheep, and respond by overgrazing his sheep also.

[Note: The tragedy of the commons is an example of the *fallacy of composition* – which is that what is true for one person must be true for everyone. Here, although one farmer may gain by grazing more of the commons, that does not imply that every farmer will gain by doing so.]

There are many examples of this kind of market inefficiency:

- *Overfishing*: as many types of fish are concentrated in areas that are geographically limited, the community of fishermen who catch those species often act as if they owned the fish stocks in common. As predicted above, this may lead each of them to ignore the impact that their actions have on the group as a whole. That impact is that an increase in this year's catch will reduce the number of breeding fish available to produce next year's stock. When one fisherman increases his catch, the private effect may be so small that the individual can ignore it. For example, if one fisherman among a hundred doubles his catch, next year's breeding stock will be reduced by about one percent – perhaps a small price to pay for such a large increase in today's profits. But if each fisherman thinks the same way, soon there will be very few fish left for next year. Because the individual can treat the reduction in the stock that he causes as an external cost, each fisherman may overfish the ocean.
- *Overhunting*: if the animals in the forest or birds in the sky are treated as if they were owned by everyone equally, each individual may be induced to act as though the cost of killing a single elk or eagle or grizzly bear was negligible, even though the cumulative effect of a large number of hunters behaving this way may be to drive the species towards extinction.

3. Policy

Because a market-based economy fails to take external costs fully into account, alternative methods of dealing with those costs may be appropriate. Numerous techniques have been proposed for this purpose. They will be discussed in Modules 36 and 37.

MODULE 35. EXTERNAL BENEFITS

In the same way that imperfect definition of ownership rights can lead to external costs, it can also lead to external benefits - that is, to situations in which the actions of one individual (or group of individuals) create benefits for third parties. A classic example of such an activity is a fireworks display. If you cannot *exclude* others from seeing your display, you will have created an external benefit for them, and you will not be fully compensated for your action. In that case, you will tend to underproduce it.

For example, assume that the cost to you of putting on a fireworks display is $10,000 and that the benefit to those who wish to see the display would be $15,000. Putting on the display would provide a $5,000 net benefit to society. But if you cannot exclude people who have not paid for it from seeing the display, you may be able to collect only $5,000 from observers (say, from people who have paid to enter a fairground) and you will not put on the display.

Some examples of situations in which individuals' actions provide external benefits to others include:

- *Epidemics*: At the beginning of each winter, health authorities encourage us to get flu shots. The *private* benefit is that we will each be less likely to suffer the costs associated with a bout of the flu. But each vaccination also provides an *external* benefit – when the probability that I will get the flu is reduced, the probability that I will pass it on to others is also reduced. Although those others might have been willing to pay me to get my flu shot, there is no obvious mechanism by which this can be done. Could I threaten my friends and co-workers that I would not get vaccinated unless they paid me? Perhaps. But how would I collect from the strangers that I might infect while I was shopping, taking the bus to work, or just walking down the street? As a result, I will be less likely to get a flu shot (most people don't) than if I had been able to benefit from the full effect of my actions.

- *Security patrols* Assume that there has been an increase in burglaries in your neighbourhood. One of the policies you might take to protect your property would be to hire a security company to patrol the area, watching for suspicious activity. One of the benefits would be that you would be burgled less often. But those benefits would be enjoyed equally by your neighbours – you would provide an external benefit to them. Unless you could find a way to encourage your neighbours to share the cost, you might choose not to invest in security even though the total, social benefits exceeded the cost. [Conversely, not only will you invest in locks and alarms for your own residence, but you may spend more on these than you would have had you been able to hire a security company. You pay for locks and alarms because you obtain all of the benefits (and the costs) of those precautions. And you invest more on them than otherwise because the crime rate will be higher than if your neighbourhood had been patrolled by security guards.]

• *Lighthouses*: If a shipping company builds a lighthouse to warn its ships of dangerous rocks, the benefits will be shared by every ship that passes by that area. Unless the company can find a way to extract a fee from all of the ships that benefit from the lighthouse, the private benefit to the builder will be less than the social benefit and the lighthouse may not be built. [The shipping owner might instruct his lighthouse keeper to turn on the light only when one of his own ships was nearby. But if the owner knew that his ship was near dangerous rocks, he wouldn't need a light to warn him.]

• *Public transit*: Assume that the benefit to you of taking the subway to work is that you arrive ten minutes earlier than if you had driven your car. Assume also that if that ten minutes is worth $1.00 to you, you will not take the subway (you will use your car) if the subway fare is $3.00. But when you take the subway, you reduce the amount of congestion on public roads. If that reduction is worth, say, $4.00 to other commuters, the total benefit you create by taking the subway is $5.00 ($1.00 to you plus $4.00 to everyone else). If the other commuters could compensate you for the benefit you have provided them, taking the subway would create a net social benefit of $2.00 ($5.00 - $3.00). But as the market economy provides no obvious method for transferring money from drivers to subway users, the subway will be underutilised.

• *Education*: Individuals are observed to engage in less crime, the greater is their education. Hence, it is sometimes argued that education provides an external benefit – the greater is the education obtained by a given group of citizens, the less the rest of us have to spend protecting ourselves from crime. As those who obtain the education do not benefit from this reduction in spending, they may underinvest in education. [Note, however, that this argument assumes that it is the greater education that "causes" the reduced criminal activity. If what is actually happening is that it is people who are less likely to engage in criminal activity who choose to become educated, then the external benefit argument no longer holds.]

• *Armed forces*: Perhaps the most extreme example of something that creates an external benefit is a defence system. If I purchase a ring of intercontinental missiles to protect my property, I will provide protection, and therefore an external benefit, to everyone else who lives within thousands of miles of me (unless I can talk them into helping me pay for my missiles).

MODULE 36. APPLICATION:

POLICIES TO DEAL WITH POLLUTION

This module identifies a number of policies that have been proposed for dealing with those types of external cost that are usually referred to as "pollution" – and which were identified in Module 34 as arising when resources have no owner.

1. Leave the Parties to Work it Out Themselves

In Module 34, it was argued that when Mr. A was deciding how to dispose of his waste, he could have paid $200 to have it hauled away. Alternatively, he could have burned his waste in his back yard, in which case the smoke would have drifted into the yard of his neighbour, Mr. B. If A did not have to compensate B for the harm he had caused, A might choose to burn his waste, at no cost to himself, rather than pay $200 to have it removed. This is the classic case of a polluting activity, one which most people would agree called for some kind of government intervention.

But assume that the harm to B was greater than $200 – say, $350. Some economists have pointed out that if the parties were left to sort this out between themselves, B *could* offer to pay A to refrain from burning his waste. (B would be willing to pay up to $350 and A would be willing to accept anything more than $200.) There would be no need for a third party (for example, the government) to intervene.

Although this argument may make theoretical sense (to economists), it encounters a number of problems in practice:

- If there are more than three or four people involved, *transaction costs* - the costs of arranging meetings, hiring lawyers, working out the technical details, etc. - may be so great that it is not worth working out an agreement. For example, if the transactions cost of negotiating an agreement between A and B was $250, the net benefit from that agreement, would be *minus* $100 (= $350 - $200 - $250).
- In many cases, the actions of a single polluter may affect the air quality of large numbers of individuals downwind (or, in the case of water pollution, downstream). In these cases, it may only be possible to raise sufficient funds to induce the polluter to reduce its pollution if most of those who have been affected contribute. For example, each of a thousand individuals might be asked to contribute $100 in order to induce a factory to install equipment that removed pollutants from its smoke. But each such individual might say to him or herself: if I don't contribute, but everyone else does, surely enough money will be raised to compensate the polluter; and I will have benefitted without making a payment. This is the ***free rider problem*** – individuals who can benefit from an action without contributing to the cost may try to avoid contributing. If enough people "free ride," it will not be possible to raise enough money to proceed. [The free rider problem will be discussed in more detail in Modules 37 and 38.]

2. Set Limits on the Amount of "Pollution"

The government could set limits on the amount of pollution that each producer could emit; limit the number of cars on the highway at particular times; limit the number of fish which could be caught by each fisherman; etc. The primary drawback to this technique is that some polluters may find it much more costly to reduce their pollution than would others. It would be efficient to allow the firms that have the highest costs of abatement to produce the greatest amount of pollution and require the firms with the lowest costs of abatement to produce the least pollution.

Also, an efficient policy would only require firms or individuals to change their actions if the benefits of those changes exceeded the costs. It might not be desirable, for example, to require a factory to reduce its air pollution by 100 tons per year when the cost of doing so would be $1 million and the benefit only $100,000. But the government is likely to have only limited information about the costs of reducing pollution, and even less information about the money value of the benefits.

3. Create "Injunctive Rights" in the Air, Water, Fish, etc.

If people are overusing the air because they do not have to pay for it, the government could specify that a certain set of individuals "owned" the air and could give them the right to exclude others from using it. In the same way that you can now obtain an ***injunction*** (a type of court order) to prevent people from dumping garbage on your lawn, perhaps you could obtain an injunction to prevent people from dumping garbage (smoke) into "your" air (e.g. the air over your lawn). Similarly, if the exclusive rights to the fish or elk in a certain area could be given to a particular individual or set of individuals, they could prevent others from infringing on that resource unless the benefits exceeded the (opportunity) costs.

The main problem with this solution is that those who want to pollute may find it extremely difficult to identify who it is they should bargain with to purchase the pollution rights. For example, a factory that wished to pollute the air over a whole city would find it difficult, because of the transaction cost problem, to bribe every air owner in the city to allow it to do so, even if every such individual considered the cost of air pollution to be less than the amount the firm was willing to pay. Similarly, it would be virtually impossible for an airline to bargain with every property owner under its flight path, to allow it to produce noise pollution.

Injunctions are also less useful when the harm occurs to a *public good,* that is, to a good whose benefits are shared equally by everyone. Among those people who are concerned about the loss of endangered species, the loss of one member of such a species affects everyone equally. For example, if a farmer reduces the reproductive success of a pair of rare ducks by filling in a wetland on his property, the loss is felt by everyone who has an interest in the survival of that species. In that case, it is difficult to identify to whom injunctive rights should be given. Should every duck enthusiast be given an equal right to accept or reject the farmer's proposed actions? If so, how would the farmer identify each of those individuals, to allow him to bargain with them?

4. Create "Tort Rights" in the Air, Water, Fish, etc

Instead of giving those who are harmed by an external cost the right to obtain an injunction against that activity before it happens, the government could give them the right to sue for harms *after* they had happened. Harms for which you can sue another person are called ***torts***. The idea, from an economic perspective, is that a polluter knows that it will be required to pay for any harm that it causes. Hence, the polluter will only cause that harm if it believes that the benefit to it exceeds the cost to those who have been harmed.

One advantage of torts is that the polluter does not have to identify those who will be harmed in advance. It is only once the harm has been caused that payment has to occur and, by that time, it may be easy to identify who the victims were. Torts are most commonly used to deal with automobile accidents, where the victim is not known until the accident occurs, but is known after.

It might be difficult to use torts to deter those whose air or water pollution affects broad areas, however. First, it will often be difficult to identify how far the pollutant has spread and, therefore, who is entitled to compensation. Second, the cost of a pollutant will often differ widely among those who have been affected, making it very expensive to calculate the damages owed to each victim. (Nevertheless, harms of this type are often dealt with using a legal device known as a *class action suit*.)

5. Create Property Rights in the Pollution Activity

Instead of giving the property rights to those who have been harmed by pollution, the government could sell those rights to the highest bidder. A common recommendation, for example, is that the government should determine the maximum amount of permissible air pollution and then sell the *pollution rights* to that level of emissions to whoever was willing to pay the most for them. In that way, those who wanted to preserve fresh air could buy the rights and not use them, while those who wanted to use the air as their "garbage dump" could buy the rights and use them to emit pollutants. In some such systems, the government sets the price of these rights; in others, firms are allowed to buy and sell rights for whatever price they can get. [A variant of this system is used in some cities to reduce rush hour congestion: those who want to drive into the city at certain times have to "buy" that right. In some cities that have tried this, each car has a transmitter that signals to a counter every time that it passes. The city then sends a bill at the end of each month.]

The primary disadvantage of this system is that the free rider and transactions cost problems will often prevent those who want clean air from buying any rights - so the level set by the government will be the level which will result, whether that is efficient or not.

6. Tax Pollution

The government could impose a tax per unit of pollutant. If the government wishes to allocate resources efficiently, it will set the value of this tax equal to the cost which the pollution imposes on those who are harmed One problem with this solution is that the government may find it very difficult to determine what the value of that harm is, and therefore what the value of the tax should be.

7. Subsidise Pollution Abatement Equipment

Instead of penalizing firms or individuals for producing external costs, the government could pay them subsidies to reduce the costs of buying and using pollution abating equipment. There are numerous drawbacks to this scheme:

- As the government has only limited information about which technologies are available for reducing pollution in each sector, it may find it difficult to identify which equipment it should be subsidizing.
- The government will have insufficient information to determine which industries, firms, and individuals are the least-cost mitigators of pollution. Hence, it may direct its subsidies to industries in which the costs of abatement are relatively high, while providing insufficient subsidies to the lowest-cost abaters.
- If the government has no monetary measure of the cost of pollution, it will not know whether the benefits of abatement exceed the costs of the equipment that it is subsidizing.

8. Government Produces the Polluting Activity

Instead of regulating the amount of pollution that is generated by private firms and individuals, the government could nationalise all polluters and run them at the "optimal" levels of pollution. The drawbacks to this proposal are the same as those discussed with respect to option 7, "Subsidize Pollution Abatement." That is, the government will lack the necessary information either to determine which pollution abatement techniques are most cost efficient or to determine whether the benefits of any abatement exceed the costs.

MODULE 37. APPLICATION:

INTRODUCTION TO THE FREE RIDER PROBLEM

In Module 36, it was argued that the free rider problem could make it difficult to implement policies that have been designed to deal with external costs. For example, assume that a thousand fishermen have each been catching 100 tons of fish per year from which they each earn $100,000. It is now discovered that, unless the annual catch is reduced to 75 tons per boat, the stock of fish will begin to decline significantly, in which case the fishermen will each earn $25,000. (This is a case in which each fisherman's catch imposes an external cost on all of the others.) The fishermen meet and agree that each of them will reduce his catch to 75 tons, reducing their incomes to $75,000.

The classic statement of the free rider problem notes that each fisherman may think: "If everyone else reduces their catches to 75 tons, but I continue to catch 100, the impact of my actions on the viability of the total stock will be trivial, and I would continue to earn $100,000." And if every fisherman thinks this way – that is, if every fisherman attempts to *free ride* on the actions of the others - no fisherman will reduce his catch, and the stock will be endangered.

A similar argument can be made with respect to most other wild species. For example, if loggers are destroying the habitat of an endangered species of owl, they might agree that, in order to save the owl, they will each leave 100 acres untouched. But each logging company may reason that if everyone else does this, it will be able to ignore the agreement without affecting the survival of the owl. And if every company reasons the same way, no precautions will be taken.

Module 38 will use game theory to examine how the free rider problem may prevent society from reaching commonly accepted goals, in four quite different situations: subsurface rights, charitable giving, voting, and traffic congestion. First, however, this module shows that what was called the "classic" statement of the free rider problem, above, is not the only possible explanation for why people may withhold contributions to a common goal.

1. Two Explanations for Free Riding

Consider here two explanations for why people might free ride: (i) that they are *self-interested*; and (ii) that they are *cynical*. The simple game described in Table 8 can be used for both explanations.

In the construction of Table 8, begin with the data that were used above to describe a fishery. There:

- if none of the fishermen attempts to reduce his catch, the fish stock will decline to the point that each of them will earn only $25,000;

• if they all reduce their catches by 25 percent, they will preserve the stock, and each earn $75,000;

• if one of them does not reduce his catch, but all the others do, he will earn $100,000 and the others will each earn $75,000. (That is, if only one fisherman reneges on his promise, the effect on the others will be trivial.)

Table 8 Free Riding in a Fishery

		All Other Firms	
		100 tons	75 tons
Firm A	100 tons	$25,000 $25,000	$75,000 $100,000
	75 tons	$25,000 $20,000	$75,000 $75,000

• to these, add the assumption that if all but one of the fishermen attempt to catch 100 tons, the stock will fall to the point at which they will each earn only $25,000. At that point, if one fisherman does reduce his catch, he will earn less than the others, say, $20,000. (Note that if every fisherman "attempts" to catch 100 tons, they will each only earn $25,000 because the stock will be so depleted.)

In Table 8, the rows refer to the actions of one fisherman, Mr. A, and the columns refer to the actions of each of "all other" fishermen. Thus:

• in the top left-hand box, when all fishermen (including A) attempt to catch 100 tons, they each earn $25,000.

• in the top right-hand box, the others reduce their catches to 75 tons and earn $75,000, while A earns $100,000 because he has continued to catch 100 tons. [Actually, as the supply of fish will

have been reduced significantly, the price per ton may have increased, in which case A will earn more than $100,000; but this modification does not alter the argument.]

- in the bottom left box, the others earn $25,000 (because they have not reduced their efforts) and Mr. A, who *has* reduced his efforts, earns $20,000.
- finally, in the bottom right box, all parties catch 75 tons, and all earn $75,000. The latter is the outcome that the parties will reach if none of them reneges on their agreement to reduce their catch - that is, if none of them acted as a free rider.

1.1 Individuals are self-interested

In the classic version of the free rider, Mr. A is *self*-interested. He looks at Table 8 and says, "if everyone else lives up to our agreement and reduces their catch, my best policy will be to renege on our agreement and maintain my catch level (and earn $100,000 instead of $75,000)." [Game theory goes one further and suggests that even if he considers the possibility that everyone else will renege, his best policy will still be to renege (and earn $25,000 instead of $20,000). Either way the self-interested individual will free ride.]

1.2 Individuals are cynical

An alternative interpretation of the free rider is that individuals are cynical about their *neighbours*. In this interpretation, Mr. A's actions are not driven by a lack of consideration for the welfare of his neighbours; rather, it results from a lack of respect for his neighbours. In this interpretation, he assumes that all (or, at least, most) of them are self-centered and, therefore, that they will all renege on their agreement. In terms of Table 8, he discounts the right-hand column. His only choice is between the two options in the left-hand column, in which case he chooses to renege on the agreement, and earn $25,000, rather than live up to the agreement, and earn $20,000.

2. Implications

Regardless of whether individuals are driven by self-interest or by cynicism, it appears that it may be difficult to induce them to live up to agreements that maximise their joint welfare. What will be required is a means for enforcing those agreements. Some that have been suggested include:

- Social censure: if the community of fishermen is small enough and sufficiently closely knit, individuals may be discouraged from free riding by the threat of being censured by their neighbours – what the world of social media would call "shaming." That this method may work better in a small close-knit community than a large impersonal one might be one argument for restricting the size of communes, such as those maintained by religious groups such as Mennonite and Hutterite colonies and Jewish kibbutzim.
- Contract law: it may be possible to use formal contracts among the participants to enforce any agreement they had reached among themselves.
- Government intervention: fishermen might be open to government regulation, particularly if the form of that regulation had been selected by the participants. In the case described in Table 8, for example, the fishing industry might ask the government to set, and enforce, a 75 ton quota per boat.

MODULE 38. APPLICATION:

FOUR EXAMPLES OF FREE RIDING

Module 37 examined how the free rider problem might arise with respect to the harvest of wild animals. In this module four additional circumstances will be considered in which free riding might affect the allocation of resources: extraction of oil, traffic congestion, charitable giving, and voting.

3. Extraction of Oil

As oil pools often extend for thousands of square miles, the surface rights above those pools may be divided among hundreds, if not thousands, of owners. If each of these owners has the right to extract the oil from beneath his or her land, there would be many opportunities for conflict among them, and for the imposition of external costs. Two of these are particularly relevant to the discussion of free riding.

First, when it is anticipated that oil prices are going to increase in the future, it may be advisable to reduce current production in order to leave oil that can be sold at those higher prices. But if there is a large number of firms extracting oil from a given field, the benefit to any one of them of reducing current production will be very small. For example, assume that when one firm reduces its current production by, say, 100,000 barrels, that will increase the total amount available next year by that amount. But if the firm is one of twenty operating in this oil field, in the next year it will benefit from only one-twentieth of the 100,000 that were saved, or 5,000. The remaining 95,000 will constitute an external benefit to the other nineteen firms. Conversely, every time a firm increases its production by 100,000 barrels this year, its own production will decline by only 5,000 next year. The remaining loss of 95,000 will constitute an *external cost* to the other firms.

Second, when oil is extracted, water or natural gas often flows into the spaces that had been vacated, thus maintaining the pressure that is necessary to drive oil to the surface. If oil is removed too quickly, however, water and gas may not flow in sufficiently quickly to maintain pressure, and the percentage of oil that can be recovered will fall. In short, the faster the oil is extracted, the less will be recovered. Thus, a firm that increases its production this year, may reduce the total amount available for all firms in the future, creating an additional external cost.

Assume that the firms that are producing oil from a particular field decide that they would maximise their collective profits if they reduced total production each year by 25 percent. Assume further that, by doing so, each of their profits would also fall by 25 percent. Individual firms might now ask themselves what their optimal action should be if (i) all other firms lived up to the industry agreement; or (ii) no other firms lived up to the agreement. In either case, each firm will conclude that it would maximise its profits by breaching the agreement. In scenario (i) if all other firms live up to the agreement, any firm that continues to produce as much as it had done previously will suffer little or no reduction in profits. And, in (ii), if all other firms breach the agreement, there

will be no advantage to any one firm from reducing its production. In short, this is a situation in which we anticipate that many firms will free ride, causing resources to be extracted too quickly.

The effects of free riding can be overcome either if the firms that have access to the surface rights over an oil field reach a contractual (i.e. legally binding) agreement to restrict output; or if the government restricts the number of wells drilled or the amount of oil that can be extracted per oil company. In the latter case, it is often argued that the restrictions should be developed by the firms that will be affected, both because they are more knowledgeable than is the government and because it is felt they will be more likely to live up to any rule that has been reached with their collaboration.

[Note also that, although that case was not discussed explicitly here, the extraction of water from underground aquifers is subject to constraints that are similar to those faced by petroleum firms.]

4. Traffic Congestion

Assume that a city's commuters can be divided into two equal-sized groups. Members of the first group prefer to drive their own cars to work but would be willing to take public transit if they received a subsidy of $4.00 per day. Members of the second group have a stronger preference to use their cars than do those in the first group, perhaps because they do not live close to transit routes that would take them directly to work. They would need a subsidy of $10.00 per day to be induced to take transit.

Assume also that there is a considerable amount of congestion on the city's streets, meaning that the commute to work by automobile is subject to significant delays. If 50 percent of those who currently drive to work could be convinced to take transit, the average commute time for those who continue to drive would be reduced by 30 minutes per day, for which those individuals would be willing to pay as much as $6.00 per day.

A civic group proposes to reduce congestion by collecting money from individuals in the second group and using that money to subsidise the transit costs of those in the first. As long as the amount collected from the second group was less than $6.00 per person and the amount used to subsidise the first exceeded $4.00 per person, members of both would be made better off.

Accordingly, the civic group sends letters to all individuals in the second group requesting that they each contribute $5.00 per day, to be used to subsidise public transit. If we assume that there are 100,000 individuals in each group, and initially that all individuals drive to work, then a 50 percent decrease in total number of drivers could be achieved by inducing all individuals in the first group to take public transit. If the individuals in the second group each contribute $5.00, $500,000 will be raised, which exceeds the $400,000 that would be necessary to subsidise the first group.

What will the response to this letter be? The analysis of the free rider problem in Module 37 predicts that most individuals in the second group will think either (i) that enough other people will contribute to this fund that at least $400,000 will be raised and it will not be necessary for

them to contribute; or (ii) not enough people will contribute to this fund to raise the necessary $400,000, so it would be a waste of money for them to contribute to the fund. Either way, insufficient funds will be raised, even if everyone involved believes that the benefit of reduced congestion exceeds the cost.

Most cities resolve this problem by using taxes to subsidise transit. Implicitly, those who benefit from the reduced congestion agree to be "forced" to pay for that benefit, in return for giving up the ability to free ride.

5. Charitable Giving

Assume that you believe that insufficient funds are being spent on a charitable project that is of concern to you: for example, on the millions of people in refugee camps, or research into ovarian cancer, or the building of shelters for homeless people. You might think:

(i) so many other people are giving to this charity that there is not much need for me to contribute. Or, perhaps
(ii) the amount of money that is required to make much of a difference to the project is large enough that the amount I would be able to give is too small to have a measurable effect. Or,
(iii) so few people in my community will donate to this project that any money I provided would just be wasted.

In each case, you might reasonably convince yourself that you should not donate to the charity in question. The first example is clearly a case of the free rider problem and the second and third are variations on that problem. In each case, however, less is given to charities than individuals would consider desirable – the market economy has failed.

Numerous "solutions" seem possible. One is for concerned citizens to convince the government to tax them in order to fund charitable projects. The rationale here is that if most people would like to give to a charitable project but find themselves subject to the free rider problem, they could be made better off if each of them was "forced" to donate (in this case through taxes).

Alternatively, some charities have learned that if they divide their programs into small projects, they reduce the probability that potential donors will feel that their contribution will have little effect. For example, charities that deal with rural poverty in Africa and Asia often ask donors to purchase a goat or pig for an individual family. This meets all three of the objections that were raised above and, therefore, may decrease the impact of the free rider effect.

6. Voting

A variation on the free rider problem may help to explain why the turnout rate is low at general elections. Assume that you have decided that, among candidates A, B, and C, you would prefer

that B was elected. Should you take the time and effort to go to your local polling station and vote for B?

You might reason as follows: if the number of other people who will vote for B is sufficient that she will be elected without my vote, then the benefit to me from voting is zero - which is less than the cost (even if the cost is very small). Alternatively, if the number who will vote for B is insufficient for my vote to make a difference, then, again, the cost of voting exceeds the benefit. Indeed, you might reasonably conclude that the only situation in which your vote will give you a benefit that exceeds the cost is that in which it is possible that two or more candidates will receive the same number of votes, in which case your vote would be the tie breaker. Except in the latter case, you might (reasonably) have chosen to free ride on the efforts of those citizens who took the effort to cast their ballots. And those who do choose to vote, say because they feel it is their civic duty, will spend less time obtaining information about the issues than if the free rider problem had not been in place.

Interestingly, when looked at this way, it is difficult to see why anyone would choose to vote, as the chance that an election will be close enough that your vote would have an effect is extremely small. And if that is the case, it is equally difficult to see how the government could induce larger numbers of individuals to vote. Perhaps the number of voting districts could be increased, in order to reduce the number of voters in each district, and therefore increase the chance that any one citizen’s vote would decide the election.

MODULE 39. IMPERFECT INFORMATION

When the concept of a competitive market was introduced in Module 15, the focus was on the assumption that these markets are composed of numerous small firms, producing approximately the same product. Less emphasis was given at that time to a further characteristic of competitive markets: that all participants are equally well informed. This module, and the three that follow, will consider the implications of removing that assumption. In particular, they will investigate what happens to the allocation of resources when participants have *imperfect information* about one another.

Two types of imperfections will be considered: (i) buyers have only limited information about the quality of the goods and services that are for sale: the *quality measurement* problem, and (ii) employers are unable to obtain accurate measures of their employees' productivity: the *principal-agent* problem.

1. Quality Measurement

The classic example of a product whose quality may be difficult to measure is a used car. Assume, for example, that A has advertised her used car for sale on Kijiji and that B has responded to her ad. If A's car is reliable (it is a "peach"), B would be willing to pay as much as $5,000 for it; but if A's car has hidden defects (it is a "lemon"), B would be willing to pay only $3,000. In this case, B would like to find a way to determine whether A's car is a peach or a lemon. At the same time A would like to find a credible method of revealing to B that her car is a peach (if it is). The methods of resolving these two problems are referred to as *signaling* and *screening*, respectively.

1.1 Signaling

Assume, as above, that buyers of used cars are willing to pay $5,000 for peaches and $3,000 for lemons. Assume also that buyers are unable to distinguish between the two (perhaps because the difference is related to mechanical reliability, which is not easily observed by the average buyer). If there are equal numbers of peaches and lemons, the average value of a used car is $4,000. In a competitive market, that is the maximum price consumers could be expected to pay. [If they are risk averse, they would offer less.] So, if the owners of peaches value their cars at, say, $4,500, they might reasonably decide that they will keep their cars rather than accept this price.

Two things happen. First, even though the optimal allocation of used cars would require that peaches be transferred from the current owners to new buyers - because buyers value them at $5,000 and sellers value them at $4,500 - this will not happen. Second, as buyers are offering $4,000, the only used cars that will be offered on the market are lemons - what Nobel prizewinner George Akerlof called a "market for lemons." This type of inefficiency will arise any time that the goods being sold come in more than one quality level (which is true for just about all goods) and the sellers of peaches find it difficult to convince buyers that their goods are of high quality.

The sellers of peaches in this case have an incentive to find a way of making credible claims concerning the quality of their products. Economists call such claims **signals**. A useful signal has two characteristics. First, it must be associated only with high quality products. It does not, however, have to "cause" the high quality, nor does it have to be acquired intentionally by the seller. For example, imagine that short people are observed to make better bank tellers than tall people. It doesn't seem likely that being short makes you a better teller, and no one would "become" short in order to get a job as a teller. Yet, from the bank's point of view shortness might, in this case, act as a useful signal of employee quality when hiring new workers.

Second, it must be less expensive for the seller of a high-quality product to obtain the signal than it is for the seller of a low-quality product. For example, if buyers decided that the owners of peaches replaced their tires more often than did the owners of lemons, buyers might use the amount of wear on a car's tires as a signal of the quality of the car. But, as it costs the same for the owner of a lemon to replace her tires as for the owner of a peach, owners of lemons would soon learn to mimic peaches by replacing their tires. The amount of wear on tires would lose its value as a signal.

A number of different examples of signaling will be discussed in Module 40.

1.2 Screening

A **screen** is a method used by buyers to induce sellers to reveal the true quality of their products. Screens are commonly used by employers to separate high productivity job applicants from low productivity. Assume, for example, that firms are hiring workers to sew jeans. Low productivity workers sew 10 pairs per hour and are worth $20 per hour to their employers; whereas high productivity workers sew 15 pairs and are worth $30 per hour. If firms cannot distinguish between the two groups at the time of hiring, and if there are equal numbers of low and high productivity applicants, half of the firms' hires will come from each group and workers' average value will be $25 per hour. In a competitive market this is the wage that will be paid by all firms.

But each firm will realise that if it can deter low productivity workers from applying to it, it will be able to restrict its hires to high productivity workers. One method of doing this would be to offer a two-part compensation scheme, in which the firm paid an hourly wage below the market wage plus a piece rate for any number of jeans produced in excess of 10 pairs per hour. For example, it might offer to pay $22 per hour plus $1 for each number of jeans over 10. Low productivity workers would not apply to this firm because they would earn $22 instead of the $25 they can get elsewhere (because they only produce 10 pairs of jeans). High productivity workers, however, will prefer this firm because they would earn $27 ($22 plus a $1 bonus for each of the five extra pairs they produce). As a result, only high productivity workers would apply to the firm – the two-part scheme would have screened high productivity from low.

If all firms learn about the success of the two-part compensation scheme, more of them will begin to offer it, and all of the high productivity workers will be hired by those firms. This means that all of the remaining workers will be of low productivity and the firms that hire them will pay a flat rate of $20 per hour.

[Note: of course, if a firm that wanted high productivity employees hired workers without screening, it would soon learn which of its new hires were from the high productivity group and which were not. At that point it could fire the low productivity workers and hire a new set, from which half would, again, be of low productivity, and so on. By screening workers in the first place though, it could obtain the high productivity workers it was seeking without constantly hiring and firing workers .]

A number of examples of screening will be discussed in Module 40.

2. The Principal-Agent Problem

If person A hires person B to perform a task, A is called the *principal* and B the *agent*. Some common principal-agent relationships include employers and employees, house sellers and real estate agents, litigants and lawyers, and car owners and mechanics. Two things are common among all of these relationships:

1. The principal often cannot easily observe how much skill and effort the agent has brought to the task - there is an imperfect information problem. This is similar to the product quality problem identified above – the principal is not certain whether the agent is a "peach" or a "lemon." [In his book, *What the Dog Saw*, Malcolm Gladwell quotes a manager at a financial services firm who told him that it takes three years for his firm to determine whether a financial adviser will be successful at his or her job.]

2. The principal may wish to induce the agent to increase his or her efforts to achieve the principal's goals, but lack the information necessary to measure that effort. For example, when you hire an accountant to prepare your income tax returns for you, you would like her to do her best to ensure that you only pay the minimum required. But you don't know whether she is burning the midnight oil on your behalf, or just putting something together quickly between games of solitaire and posts to her Facebook page. You would like to find a method of inducing your agent to put in as much effort as you would have yourself.

When the principal is unable to observe the skill and effort that the agent has expended, the market may not operate efficiently. Buyers (principals) may receive less than they had anticipated from a transaction, which means both that they would have been better off spending their money elsewhere and that they may spend less in this market than they would have preferred had they been able to obtain better information. At the same time, sellers (agents) may earn less in this activity than they would have had principals been able to observe, and compensate them for, extra skills and effort.

Modules 41 and 42 will discuss a number of examples of policies that might be adopted to minimize the principal-agent problem

MODULE 40. APPLICATION:

EXAMPLES OF SIGNALS AND SCREENS

In Module 39 it was argued that if buyers and sellers were not well informed about one another's characteristics, it would be difficult to ensure that all mutually beneficial exchanges between them would be made. This module will provide examples of two techniques that are commonly used to increase the amount of information available in the market: signals and screens. Modules 41 and 42 will then introduce a number of policies that principals use to motivate agents when they have imperfect information about agents' productivity.

1. Signals

Although signals can be found everywhere, this section discusses three circumstances in which sellers have used signals to allow them to convince potential buyers that their product is of above-average quality.

1.1 Employees

Traditionally, economists have argued that workers' incomes increase with their education levels because education gives them skills that increase their value to employers. But this view has been challenged by those who argue that education's main function may be to act as a signal.

Remember that, to be a useful signal to employers, a characteristic need only to

- be "associated" with high quality - it does not have to cause an increase in quality; and
- be more costly for low-quality workers to obtain than it is for high-quality.

Education may meet both of these criteria. With respect to the first criterion, assume for example, that the kinds of people who will make productive employees are also the kinds of people who choose to go to university. In that case, even if university education did not cause workers to be productive, it could act as a signal of productivity.

If this was true, however, non-productive workers would be tempted to mimic productive workers by also investing in university educations, in which case education would lose its value as a signal.

They might be discouraged from doing so, however, if the second criterion above was met: if it was more expensive for unproductive workers to obtain university educations than it was for productive workers. That additional expense could come in the form of higher opportunity costs, because it took longer for unproductive workers to earn a degree; or in the form of greater effort required to pass university courses. In either case, it is at least theoretically possible that education could be a signal of quality, not a determinant.

Although most of the evidence supports the view that valuable skills are learned at university, some persuasive evidence has been presented to suggest that university education acts largely as a signal. Perhaps, for example, increased emphasis on earning a master's degree has come about because the cost of earning a bachelor's degree has fallen sufficiently that it no longer provides a strong signal of productivity to employers.

1.2 "Lemons"

In Module 39 it was argued that if buyers are unable to distinguish between high quality and low-quality goods, sellers may offer only low-quality goods, even when some buyers would be willing to pay for high quality. In this case, sellers of high-quality goods may try to provide a signal that distinguishes their products from those of low-quality sellers.

One such signal would be to offer a warranty, promising to repair defects that are found on their products. As high-quality goods will have fewer defects than low quality, the cost of a warranty to sellers of the latter will be higher than it is to the former. And if the cost differential is sufficient, only the sellers of high-quality goods will offer warranties and buyers will be able to use the presence, or absence, of a warranty as a reliable signal of quality. (Note: this does not necessarily mean that only high-quality goods will be offered. If some buyers are less concerned about product quality than others, they may still choose to purchase the lower quality goods, with no warranty.)

Also, sellers of high-quality goods may be able to use social media to create signals of their products' characteristics. For example, many restaurants report customer reviews on their websites. But, if restaurants curate their own reviews, there is always the suspicion that they may be leaving out poor reviews. Some high-quality restaurants have dealt with this issue by adding links on their websites to independent review sites, like Yelp. As the latter will provide more good reviews, and fewer bad reviews, to high-quality restaurants than to low-quality, just the fact that the link exists may be a strong signal to the discerning customer.

1.3 New products

If a firm has been in business for a long time, one of the best signals of the quality of its product will be its reputation, earned mostly through word of mouth and social media. But if a firm is introducing a new product that it believes is of higher quality than those of its rivals, it will look for signals that take less time to establish.

One of these is advertising; but not the kind of advertising that extols the virtues of the firm's product, although that won't hurt either. Rather, advertising becomes a signal when its sheer volume is the relevant factor, independently of its content. For example, imagine that two fast food restaurants, A and B, decide to introduce a new product to their menus: veggie burgers. Firm A has announced a massive advertising campaign for their new product.

Firm B's managers, when they meet in the privacy of their boardroom, admit that Firm A's version is better than theirs. Should B reply with its own advertising campaign? The answer can be found by applying the game theoretic approach used in Modules 31 and 32. If consumers don't yet have any information about the products, advertising campaigns will initially draw approximately equal

numbers of customers. But, once those customers have had an opportunity to compare the two burgers, their purchases of A's product will increase and of B's product will decline. As B's managers can predict that this is the likely outcome, they will realise that their best policy may be to avoid an expensive advertising campaign in the first place. The outcome? Only firms that are confident that they have a high-quality product will spend heavily on advertising; and that level of expenditure will become a credible signal of quality.

A similar example arises with respect to expenditures on company vehicles. Assume, for example, that two housepainters, J and K, are both planning to offer their services in a new neighbourhood. J thinks that, as the quality of his service is not related to the quality of his truck, he can save money by using a beat up old truck with a sign hand painted on the door. K, however, decides that the quality of his truck should reflect the quality of his service, and he buys a new truck with professional signs painted on the sides and back.

Consumers might conclude that K was better able than J to afford a new truck because he expects to be in business longer, and therefore is able to spread the cost of his truck over more projects. And, if the length of time a housepainter can expect to be in business is determined by the quality of his product, then the quality of his truck acts as a credible signal. [Only painters who have been in business long enough to have developed a reputation for quality service will be able to forego the expense of new vehicles.]

2. Screens

Whereas signals allow sellers to send information to buyers, screens allow buyers and sellers to extract information from one another. For example, in the first example discussed under signaling, above, it was the sellers (workers) who chose university completion as signal to buyers (employers) that they were of high quality. In screening, employers design systems to discourage low quality workers from applying for jobs. Two instances in which screening might be used are discussed here.

2.1 Employers

As it is often difficult for employers to distinguish between workers of low productivity and those of high productivity at the hiring stage, they will try to devise policies that screen between these two groups. One such policy, that was introduced in Module 39, is to offer a two-part compensation scheme in which employees are paid a relatively low hourly wage plus a bonus that is based on observed performance (such as number of pairs of jeans sewn per hour). As low productivity workers do not expect to perform very well, they may prefer to work at firms that offer a fixed hourly wage. The two-part scheme will have screened them out.

Similarly, *up-or-out* employment contracts may also act as screens. In these contracts, employees are offered relatively low wages when they are first hired, but receive significant wage increases after an evaluation period, say a year, if their performance has proven to be satisfactory. Otherwise, they are let go. Hence, "up" in this contract refers to the increase in wages and "out" refers to being

laid off. (University professors usually have this type of contract, with an evaluation period of five years.)

The original reason for offering this type of contract was that it was often easier to identify whether employees had been working effectively if the employer had a long period to observe them. Workers who are found to have been working productively are offered increases in wages, while those who have been found to have been unproductive are laid off.

But these contracts have a screening function also. As low productivity workers receive low wages during the evaluation period and expect to be laid off after that period ends, they will be discouraged from applying to firms offering up-or-out contracts. The employer will have screened them out.

2.2 Insurance

Insurance companies are sellers who would like to screen out high risk buyers. Automobile insurers, for example, would like to distinguish between high risk and low risk drivers. A common method for doing so is to offer a policy that promises to pay for all of the insured driver's accident costs minus a *deductible*, say $1,000. Individuals who know that they are at high risk of making an insurance claim will expect to incur this deductible more often than will individuals who are at low risk. Hence, low risk drivers may be more likely to choose a policy that has a large deductible than are high risk drivers and the deductible option will have acted as a screen.

MODULE 41. APPLICATION:

DEALING WITH THE PRINCIPAL-AGENT PROBLEM

An individual who hires another to perform a service is called the *principal* and the person who has been hired is called the *agent*. Two fundamental problems arise in such relationships.

First, the agent's motivation to perform the service is not the same as the principal's. The principal, for example, may be a business owner whose primary goal is to maximise the profits of her company; whereas the agent may be a bookkeeper whose goals are to maximise his wages while minimizing his effort, and who wishes to work hours that are as convenient to him as possible.

Second, in order to motivate the agent to align his goals with those of the principal, the latter may wish to offer incentives to the agent. But these incentives will only provide the desired motivation if the principal is able to determine how much effort the agent has expended on her behalf.

If you think of your own experiences as an agent (say, as an employee) or as principal (say, as the seller of your own house), you will recognise that a large variety of policies have been designed to bring the agent's actions in to line with those that are desired by the principal. Five of these policies are discussed in this module. Module 42 will then examine in more detail the types of policies used by retail firms to motivate their sales forces.

1. Piece Rates

If it is easy to determine how many units the agent has produced, it may be possible to base the agent's compensation on a *piece rate*, that is on a dollar amount per unit produced. A waiter, for example, might be paid $2 per customer served; or a truck driver might be paid $1 per kilometre driven. But for various reasons piece rates are not widely used.

The first is that the output of many, if not most, agents is not easily measured. How would you measure the "output" of an accountant, or a bank teller, or an advertising executive? Indeed, there probably are very few principal-agent relationships in which the principal can monitor accurately how hard the agent has worked or how much he has produced.

Even if output can be measured, it may be difficult to disentangle the output of an individual worker from that of her colleagues. For example, it may not be clear how much each worker on an automobile assembly-line contributed to the production of a single car.

A further problem is that piece-rate systems can create perverse incentives. Waiters who are paid according to numbers of customers served may be tempted to rush customers through their meals and shirk their responsibilities to keep the restaurant clean. And a driver paid by kilometres driven may pay little attention to the harm that would be done to the owner's truck if he was to drive too fast or was to ignore warning signs that the truck needed maintenance.

Also, it may difficult to choose an appropriate piece rate. If the employer is trying to establish the rate for a new product, it will first need to determine what the typical rate of production is per worker. It probably will not, for example, want to pay unskilled workers $8 per unit if those workers can easily produce ten units per hour. But if the firm bases its piece rate on workers' productivity during an observation period, workers will have an incentive to work as slowly as possible during that period. And even if an "appropriate" piece rate has been established initially, it will always be in workers' interests to complain that the rate is too low, creating ongoing friction between employers and employees.

2. Profit Sharing

Even if piece rates could be used to encourage agents to increase output, and therefore revenue, they offer little incentive for employees to reduce costs. One policy that is aimed at minimising this problem is to pay employees a percentage of the principal's profits – as profits are increased both by increasing revenue and by decreasing costs. But profit sharing quickly runs into an issue that has been seen before in these notes: the free rider problem.

Imagine that a company that has a thousand employees offers to pay its employees 25 percent of the company's profits. Assume also, that if the employees were to increase their effort significantly, the company's profits would rise by $10 million dollars per year. Each employee's profit share would increase by $2,500 ($2.5 million/1,000). If that was sufficient to compensate them for working harder, employees might well agree to this proposal.

But each employee might think: "If I do not increase my effort, and everyone else does, the total amount available for me will only be reduced by one-thousandth and I will earn only $2.50 less than if I had worked hard all year." Given those considerations, it would not be surprising if many employees chose to free ride. If enough employees do so, the remaining employees may decide that it is not fair for them to work hard and the profit-sharing scheme will fail.

For this reason, profit-sharing schemes are not common at large firms. They are, however, prevalent among small firms, especially those with fewer than ten employees. The reason for this is that if one employee among, say, four or five, was not pulling his or her weight, this would be apparent to the others. In that case, we can imagine either that the non-performing worker would be fired, or that social pressure would be placed on him or her to work harder.

[Note: this argument suggests that large companies might break themselves down into small "profit centres," each of which would have few enough employees that self-monitoring would work. This might be an additional reason for franchising - see Module 30.]

3. Up-or-Out Contracts

Module 40 introduced up-or-out contracts, in which employees were paid relatively low wages during an evaluation period, after which they would either receive long-term contracts with relatively high wages (up) or would be fired (out). The argument was made there that these contracts would act as a screen between low and high productivity applicants. Up-or-out contracts may also act as incentive schemes. If employees know that if they were found to be productive,

they would be offered lucrative contracts at the end of the evaluation period, that could act as an inducement to work hard (at least early in their careers).

These schemes are commonly used in occupations such as lawyer and university professor, in which it is difficult to determine in the short run whether an employee is working productively. The assumption is that unproductive workers will not be able to disguise their inadequacies over the long term. Thus, if the evaluation scheme is of sufficient length, the employer will be able to separate low-productivity workers from high.

4. Efficiency Wages

If it takes time – say, a year or two – for the firm to determine whether an employee has been working hard, it might offer a wage that exceeded the market wage – perhaps $23 per hour when the prevailing wage at other firms was $20. This could be beneficial to the firm for two reasons. First, the employee might feel a moral duty to work hard to justify the employer's generosity. Second, if the employer indicates that it will fire the worker if he or she is discovered to have been shirking, the worker may work hard in order to avoid losing the $3 wage differential. When a wage is set higher than the market level in order to induce workers to increase their efforts, economists refer to it as an *efficiency wage*.

If an efficiency wage that exceeds the market wage by $3 induces employees to produce more than $3 of additional output, that wage is profitable for the firm. And if a $3 wage differential is sufficient to compensate employees for working harder, they are better off also. Thus, even though the employer pays more than other firms in the market, and employees work harder than at other firms, both parties are better off.

[Of course, as an efficiency wage is, by definition, a wage that exceeds the wage paid by the average firm, it is not possible for all firms to offer efficiency wages. In the long run, the market will end up with one set of firms paying high wages to employees that work hard and another set paying low wages to shirkers.]

5. Rank Order Tournaments

Finally, imagine that it is easier for the firm to determine which of its employees is the most productive than it is to measure exactly how productive each employee is. The firm might operate a *"tournament"* in which it promised to promote the workers who were judged to be the most productive. Until the outcome of the tournament was announced, this system could encourage all employees to work hard enough to be place at the top of the ranking. On the other hand, once the outcome has been announced, the firm will have to deal with disgruntled employees who were not promoted, especially if they had worked hard and only failed to be promoted because someone else ranked above them.

MODULE 42. APPLICATION: COMPENSATION PACKAGES AT RETAIL FIRMS

The purpose of this module is to apply some of the lessons from the preceding modules to the establishment of compensation at retail firms (and other similar firms, like restaurants). Such schemes need to:

- *Ensure that employees' total earnings are competitive with other firms in the market.* This is pretty obvious, in the sense that wages plus fringe benefits, plus incentive pay have to be sufficient to attract workers from comparable firms. It is less obvious, however, in the sense that if a firm offers an incentive scheme that its employees really like, it may not have to pay an hourly wage that is as high as they can get elsewhere.

- *Screen out shirkers*. An incentive scheme can help the firm to discourage individuals from applying to work for it if they aren't planning to work hard; and it will encourage individuals to apply if they are planning to be productive. So, if firm A's competitor is offering a flat $15 per hour, shirkers might prefer to work there than with A, if A is offering, say, $14 per hour plus performance bonuses (because shirkers aren't planning to work hard enough to earn bonuses); whereas productive workers might prefer A to the flat-wage company – because they plan to make A's $14 plus $2 or $3 per hour in bonuses. In this sense, then, a well-designed bonus scheme can screen out shirkers.

- *Reduce turnover*. Retail firms' bonus schemes often reward employees for achieving annual targets. Thus, employees who are performing well after, say, six or eight months will have an incentive to stay to the end of the year, to get their bonuses. With respect to those people an annualized scheme will help the firm to reduce turnover, which is usually costly, both in terms of the actual cost of the hiring process and in terms of time spent by administrators interviewing and training new employees. On the other hand, if most employees plan to stay for less than a year or two, annual bonuses may be of little value to the firm.

- *Encourage increased sales*. The most obvious goal of an incentive scheme is to discourage employees from shirking, and to encourage them to make every possible sale. What may be less obvious is that there has to be a clear relation between the effort an employee puts in and the reward that she receives. This is why Module 40 was critical of profit sharing in large companies – if an individual's efforts increase company profits by $10,000 but she is one of 1,000 employees, profit sharing will give her only $10. Not much of an incentive. Unless all of the company's employees are altruists, it will have to give them a bonus that is large enough to compensate them for the extra effort they have put in.

And that bonus should come close enough to the time employees made an effort that they fully recognize the connection. If hard work in January is not rewarded until the end of December, employees probably won't have the same incentives as if they were rewarded at the end of January (even if they were planning to stay with the company until the end of the year).

Also, there is an argument for setting sales targets according to the time of year. If sales are traditionally low in February, the firm might consider setting lower targets for its employees in that month, to recognize that the same level of effort in February creates a lower level of sales. (That is, there is an argument for tying bonuses to effort as much as to output/sales, in order to keep employees interested in the firm's incentive scheme.)

- *Provide the greatest encouragement to the most profitable sales*. Imagine that if one store in a chain earns approximately $8 million per year, it will be considered to be successful – i. e. it will earn enough profit to keep headquarters satisfied. Assume also that, in the process, all of the costs of rent, heat, electricity, advertising, and salaries are covered. For example, assume that its average item sells for $50 and that the costs per item, averaged across all items sold in a year (when sales are $8 million) are $20 paid to the manufacturer of the item, $5 in wages and salaries (including non-sales staff), $10 for rent, heating, electricity, etc., and $10 for all other miscellaneous expenses (including advertising and payments for the firm's headquarters). That leaves a profit of $5 per item.

  Now, assume that the store increases sales by $1 million, without increasing the number of employees or the size of the store. Because wages and salaries, rent, heating, etc., and miscellaneous expenses won't increase (they are what economists call fixed costs), the cost *per extra unit sold* will be $0. So, on the sale of one extra item, the cost will just be the $20 payment to the manufacturer (assuming that the firm, in the long run, will have to bring in more stock to meet the increased sales) and the profit to the company will be $30!

  It is these "incremental" sales that the company really wants to encourage. This would argue, then, for a scheme (i) that paid larger bonuses after certain targets had been met; and (ii) that set those targets at the level at which the store began to cover its fixed costs. (In this example it is only when the store's sales exceed $8 million that the really big bonuses would kick in.)

- *Not encourage undesired behaviour*. One of the problems with incentive schemes is that they encourage only a limited number of employee behaviours and, therefore, may inadvertently reward behaviours that are not consistent with the firm's goals. For example, if the employees of a clothing store are rewarded strictly according to dollar value of sales, they may lack the incentive to do things like tidy up clothes that customers have put back in the wrong places; and because bonuses reward individual sales, employees might be

tempted to “steal” customers from one another. Perhaps this problem could be dealt with, to a certain extent, by tying part of the employee’s bonuses to storewide goals. Maybe the percentage bonus could increase with both the employee’s and the store’s sales. For example, if the store sells more than a $1 million in a month, individual employees receive a 1% commission on sales over $30,000; but if it sells less than $1million, they get a commission of 0.75%. [But, of course, the firm has to be careful to avoid the free rider problem previously discussed with respect to profit sharing.]

MODULE 43. WAGE DISCRIMINATION I

One of the great policy questions of our time is "Why do members of some groups – such as women, blacks, and Hispanics – earn less than equivalent members of other groups?" Critics argue that the presence of this differential is an indication that the market system does not operate efficiently. This module investigates that critique.

The analysis proceeds in three steps. The first section asks whether some of the differential can be explained by factors that are outside the workplace – by differences between groups in their natural aptitudes, and by discrimination that occurs in non-market aspects of society. The second section will investigate discrimination by employers in a competitive market. Finally, in Module 44, the analysis considers the impact of imperfect competition and of the roles of discriminatory banks and customers.

1. Non-Market Sources of Wage Differentials

Assume that members of one group are observed to earn higher wages than individuals in another group, all else being equal. Before it can be concluded that this differential arises from discriminatory behaviour by employers, it must first be determined whether the members of the first group bring a different set of characteristics to the labour market than do members of the second. Two possible sources of differences in characteristics will be considered here: *genetic* factors, such as differences in strength and verbal skills, and *systemic* factors, such as differences in the ways that boys and girls are raised.

1.1 Genetic factors

Many of the characteristics that allow one person to earn more than another– such as strength and intelligence – are, at least in part, inherited; and some wage differentials may arise from differences in these characteristics. It is clear, for example, that men, on average, are stronger than women. Hence, among individuals with high school education or less, men often earn more than women because they are better able to obtain jobs requiring physical strength, such as those in construction.

The sources of other differentials are more controversial. It is sometimes argued, for example, that men have a higher aptitude than women for quantitative, mechanical, and visual-spatial tasks; and that women have greater aptitude for verbal and social skills. Aside from the question of whether these differences can be measured accurately, it has been argued that they arise at least in part from differences in the way that boys and girls are raised. If parents, media, and schools all, intentionally or not, reinforce a societal view that boys are better at mathematical and scientific tasks than girls, and that girls are better at verbal and nurturing tasks than boys, it may not be surprising that the two sexes direct themselves to employments that take advantage of these respective aptitudes. And

if mathematical and scientific jobs pay more than verbal and nurturing ones, women will earn less than men – not because of natural ability but because of cultural expectations.

The question of whether differences between men and women (or other groups, such as blacks and whites) arise from natural ability or whether the main factor is cultural upbringing is an important line of research in the debate about sources of income differentials. The answer to that question, if there can be one, will have an important impact on public policy: if men truly do inherit different job-related characteristics than do women, then a policy that aims to treat them equally will have a much different effect than if differences between them arise from cultural factors.

1.2 Systemic discrimination

Systemic discrimination is said to occur when the "system," that is society, discriminates among groups in such a way that differences are created among them with respect to productivity in the workplace. This discrimination can be either implicit or explicit.

Implicit discrimination is the unintentional result of the application of societal norms. Traditionally: children's books portrayed boys (even when they were pigs or bears) as more adventurous than girls, mothers stayed home and fathers went to work, and girls helped make dinner while boys played sports. Girls were given dolls and skipping ropes while boys were given trucks and Lego sets. And girls were taught "home economics" in school while boys studied "industrial arts." Although there have recently been many attempts to remove these biases, some are sufficiently ingrained in our society that it is likely that at least some remnants will persist for some time to come.

Explicit discrimination arises when one group consciously chooses to exclude others from opportunities to develop the skills needed for the workplace. For example, law schools and medical schools long had policies that strongly favoured white, Christian, males, often barring students from other groups altogether. When this type of discrimination is widespread, policies that require that firms to hire on the basis of education will fail to reduce wage differentials among the affected groups. Even non-discriminatory employers may prefer to hire men who have university education over women who do not. Discrimination will be reduced by changing policies in the educational sector, not in the workplace.

2. Discrimination in the Workplace: Competitive Markets

Discrimination occurs in the workplace when employers treat workers with the same employment skills differently, based on factors such as age, sex, race, and religion. Implicitly, most people seem to think that the source of this behaviour is that: men are misogynists, so they are less likely to promote women than men; whites are racists, so they pay blacks less than whites; Christians are anti-Semitic, so they refuse to hire Jews; etc.

But economists take a more nuanced view. What would happen, they ask, if not *all* men were misogynists; if not *all* whites were racists; if not *all* Christians were anti-Semites? In those cases, wouldn't non-discriminatory employers maximise their profit by hiring from minority groups?

For example, assume that construction is a competitive industry and that firms hire only male carpenters, at $30 per hour. If well-qualified female carpenters are available who would be willing to work for $20 per hour, any firm that hired women at that wage would be able to earn significantly higher profits than its rivals and could undercut those firms' prices. Soon its rivals would have to choose between going out of business or matching the non-discriminatory firm's treatment of women. As that firm and others increased their employment of women, the demand curve for female carpenters would shift to the right, and for male carpenters to the left. If the least discriminatory firms treated male and female carpenters equally, ultimately the wages of the two groups would equalize.

[Note that, in this model, if the least discriminatory employer still had a "preference" for males over females, the wages of the two groups would not equalize. For example, if the least discriminatory employer had a 10 percent preference for males, the ultimate wage differential would be 10 percent – perhaps $28.00 for males and approximately $25.50 for females.]

Even if all employers were discriminatory, it might be possible for new non-discriminatory employers to enter the industry and hire from the minority group. If manufacturing firms hired whites in preference to blacks, for example, a black employer might be able to start a competing firm that hired blacks. A Muslim employer could hire other Muslims; a female employer could hire other females, etc. As long as minority group employers did not discriminate against other members of their own group, wages of minority workers would eventually increase to equal those of majority workers.

But at least three factors could act to inhibit this equalization:

1. If the discriminatory firms operate in a non-competitive industry, it would be difficult for new firms to enter, regardless of the availability of financing.
2. Potential minority employers may have difficulty raising financing to start a new firm if banks discriminate against them.
3. Customers from the majority group might discriminate against firms that employed a significant number of minority workers.

These factors will be considered in Module 44.

MODULE 44. WAGE DISCRIMINATION II

3. Discrimination in the Workplace: Imperfectly Competitive Markets

The model described in Section 2 of Module 43 assumed that the labour market was competitive – that is, that new firms could enter the market easily. In that case, if employers discriminated against a minority group, new firms could enter and hire workers from that group at wages below the prevailing market level. Existing firms would have to hire from the minority group or face losing business.

But if discriminatory employers operate in an imperfectly competitive industry, new firms will find it difficult to establish themselves. If the automobile industry was discriminating against black assembly-line workers, for example, it is unlikely that new firms would enter just to take advantage of the low wages available for those workers, because the costs of establishing such a firm are so large relative to the saving in wage costs.

This ability to discriminate is likely to be most pronounced when firms are the primary employers in their region, and when their workers have specialised skills that cannot easily be transferred to other firms. For example, as automobile factories are often located in relatively small towns where there are few other employers, employees who feel they have been discriminated against may not be able to leave the firm to work for non-discriminatory firms in other industries. Similarly, if the automobile industry discriminates against, say, female mechanical engineers, there may be few alternative industries to whom those employees could turn. In these cases, the marketplace may provide few remedies for those suffering from discriminatory policies.

But this leaves many sectors in which market forces *will* offer those remedies, even when the employer operates in an imperfectly competitive industry. Consider the IT industry, for example. Firms like Google and Facebook may dominate their industry, but they employ workers whose skills are transferrable to other firms, and they have their headquarters in cities like San Francisco and New York where employees have many other options if they feel they have been discriminated against. And, even if an IT firm dominates its local market, like Microsoft in Seattle, its programmers are often sufficiently geographically mobile that they could always relocate – to San Francisco, New York, or even Toronto. The result is that, although female or black employees may be willing to accept a certain amount of discrimination in return for the security offered by large, international firms, the ability of those firms to discriminate will be limited by employees' options elsewhere.

4. Discrimination by Financial Institutions

Economists' argument that it will be difficult for firms to discriminate in competitive markets requires that there be ease of entry into those markets. If, for example, the local grocery store pays more to white employees than to blacks, it might be profitable for a black entrepreneur to establish a competing grocery that hired only black workers. But if financial institutions were

discriminatory, they might refuse to lend to such individuals, or might lend only at high interest rates. Historically, at least, this type of discrimination held back Chinese business owners in Western Canada and black entrepreneurs in the Southern U.S.

5. Discrimination by Customers

Imagine that the owners of law firms, hardware stores, and coffee shops recognise that if they were to replace their white, male, Christian employees with black, female, Muslims they would reduce their costs significantly. When any of them attempt to do so, however, they discover that their customers refuse to interact with the new employees – it is the firms' *customers* who discriminate. It is this prejudice that has driven much of the discrimination against minority workers in retail and personal services (e.g. lawyers, accountants, plumbers, and hairdressers). And it has been a major driver of discrimination against black athletes. Even seventy years after the "major leagues" began to integrate in North America, there is still evidence that ticket prices are highest for teams with the greatest numbers of white players.

This bias is difficult to overcome, in large part because the customers who have the most to spend are usually drawn from the majority group. If car dealers hire only white salesmen, for example, a new competitor may be able to minimise its wage bill by hiring black salesmen. But, if most customers are whites who prefer to deal with white salesmen, the new firm will soon find that it is losing money. It is only at firms that cater primarily to black customers that we would expect to see a significant number of black lawyers or salesmen. And because black customers generally have lower incomes than do whites, the wages of workers at firms that cater to blacks will be lower than those that cater to whites.

This effect may be exacerbated if members of the minority group have been conditioned by members of the majority group to believe that minority group workers are of low quality. For example, if women become convinced that male accountants, lawyers, and financial advisors are superior to female, even female customers may act in a discriminatory manner.

5. Crowding

Some non-economists have argued that employers might restrict minority workers' access to high status jobs in order to "crowd" them into low status jobs. In this way, supply to the low status jobs will be increased and wages driven down. Retailers, for example, might deny women promotion to supervisory roles in order to maintain a supply of salespeople, cashiers, and stockroom workers.

There are two problems with this theory. First, if crowding reduces the wages of workers in low status jobs by increasing supply to those jobs, it must also increase the wages of workers in high status jobs by decreasing supply to those jobs. The theory requires that the decrease to the former group exceeds the increase to the latter; but there is no reason to believe that this will occur.

Second, if the workers that a firm refuses to hire are crowded into the workforce of a different firm, the costs of this action (higher wages) will be borne by the first firm while the benefits will be enjoyed by the second. For example, if blacks and women are denied lucrative jobs in

manufacturing in order to crowd them into low wage laboring jobs, manufacturers will experience higher wages while the employers of laborers – for example, cleaning companies and fast food restaurants –will benefit from lowered wages. It is not clear why manufacturers would agree to this policy.

But crowding might be an economically advantageous policy for majority group *employees*, rather than employers. If, for example, most factory workers are white males, it might benefit them to insist that their employers discriminate against black and female workers as that would reduce the overall supply of skilled factory workers, thereby increasing their wages. As this form of crowding would be expensive for employers, non-discriminatory employers would object to it and it would only persist if workers possessed sufficient bargaining power to enforce it.

MODULE 45. APPLICATION:

POLICIES TO REDUCE WAGE DISCRIMINATION

There are two broad approaches to reducing discrimination in the workplace. The government can introduce policies that encourage: (i) equal compensation for equal work or (ii) equal opportunities in hiring practices. As most of the literature on this issue concerns male-female wage differentials, it is the rectification of those differentials that will be the focus of this module.

1. Equal Compensation

Those who believe that the male-female earnings differential can be attributed to labour market wage discrimination have called for legislation requiring either "equal pay for *equal work*" or "equal pay for "*work of equal value*." The former policy, usually referred to as **equal pay** legislation, requires that women be paid the same wages as the men who are working in jobs that are "substantially similar" to the ones in which women are employed.

The second policy, generally referred to as **pay equity** or **comparable worth** legislation, requires that each job be evaluated according to a common set of criteria and that wages be based on those criteria – ensuring that even if men and women worked in jobs that were substantially different from one another, they would be paid equivalent wages for tasks of equivalent, or comparable, value.

1.1 Equal pay

Although most countries now have equal pay legislation, those regulations have not proven to be very effective. One reason for this is that women and men are often employed in different occupations. For example, even if female carpenters were to earn the same hourly wages as male carpenters, that would have little effect on the gap between the wages of women and men in general because so few women work in that occupation.

Furthermore, workers within a given occupation are often given wage increases as their experience, productivity, and responsibility increase. We might find, for example, that in an engineering firm, engineers start at relatively low wages in a position called something like Engineer 1. Then, as they gain experience, they might be promoted to Engineer 2, Engineer 3, etc. Pay equity would require that all workers at, say, Engineer 2 receive the same wage. But women in such a firm might well experience discrimination in the *promotion* process, with the attributes of female workers being valued, implicitly, at less than the attributes of male workers. If this is how discrimination arises, it will not be sufficient to ensure that, say, each female Engineer 4 is paid the same as her male equivalent, it will also be necessary to ensure that women have the same probability of reaching Engineer 4 as do males with similar qualifications to theirs.

1.2 Pay equity/comparable worth

Because equal pay legislation has limited effect when women and men work in different occupations, some researchers have suggested that each employer be required to establish a *job evaluation* scheme in which all jobs are compared using a common set of criteria. These criteria – which include factors such as level of formal education required, numbers of workers supervised, amount of contact with the public, value of the equipment for which the employee is responsible, and whether the work takes place indoors or outdoors - are used to place a value on each position within the workplace. Women are considered to have been discriminated against if the job evaluation scheme finds that the value of their contribution to the company exceeds the wages they are being paid.

The primary criticism of job evaluation schemes is that the construction of the valuation system is subjective. First, as it is feasible to use only a subset of all possible employment criteria, the choice of which criteria to use may have a significant effect on the valuation of the firm's numerous jobs. If, for example, most female employees are clerical workers while most male employees are warehouse workers, choice of "amount of lifting" as a job characteristic, in preference to "familiarity with Office 365," will result in male jobs being valued more highly than female. Yet there is no clear, objective method of determining which characteristics should be used and which should not.

Furthermore, even if agreement can be reached concerning the evaluation criteria, further problems will arise when attempting to attach values to those factors. For example, assume that an electrician's position requires grade 10 plus two years of technical school and three years of apprenticeship; whereas a nursing position requires grade 12 plus three years of college. Both require five years of education beyond grade 10. But should the two years of technical school plus three years of experience required of electricians be treated as being equivalent to the two extra years of high school plus three years of college required of nurses?

And how much weight should be given to the working conditions of an electrician relative to those for a nurse, or to the relative dangers of the two occupations, or to the requirement that nurses work night shifts, weekends, and holidays?

In response to these issues, it has been suggested that statistical analysis might be used to evaluate employment criteria. Given information about the characteristics of the various jobs at a firm and about the wages paid to each job, a technique called *multiple regression analysis* can be used to estimate an equation like

$$W = 4.1\ E + 2.5\ S + 0.3\ N$$

where W is the hourly wage, E is number of years of post-secondary education required, S is number of workers supervised, and N is number of hours per week of night and weekend shifts. The weights – here, 4.1, 2.5, and 0.3 – are chosen statistically in such a way as to maximise the equation's ability to predict what the hourly wage will be in each job. For example, the equation above predicts that a job that requires four years of post-secondary education, involves supervision of five workers, and averages eight hours of weekend shifts will be paid

$$W = 4.1(4) + 2.5(5) + 0.3(8)$$
$$= 16.4 + 12.5 + 2.4$$
$$= 31.3$$

The advantage of such a statistical approach is that it is objective. There is no wrangling among employees, with each group lobbying for the greatest weighting to be given to the characteristics associated with its job. The disadvantage, however, is that if the firm has been discriminating against its female employees, lower wages will have been paid to the jobs with the greatest numbers of females. The statistical analysis will "find" that the characteristics associated with those jobs receive the lowest weightings. We end up in a circular argument: women are being paid less because they are concentrated in the jobs with the least valued characteristics; and those characteristics have the lowest valuation because women, who are paid low wages, are concentrated in the jobs that have those characteristics.

2. Equal of Opportunity

Equal pay regulations are designed to deal with discrimination against minority groups once they have obtained employment. But there is considerable evidence to suggest that discrimination is present in the hiring process – women and blacks are less likely to be invited to job interviews than are white males, and they are less likely to be hired once they have been interviewed. Two types of policies have been suggested to counter this form of discrimination.

2.1 Equal opportunity

Government regulations often require that employers give equal opportunity to minority group workers in the hiring process. Although well-intentioned, these regulations prove difficult to enforce. Employers argue that it is not discrimination that leads them to hire white males – they are hiring the best qualified workers available and that just happens to be white males. To prove discrimination, the government has to show that the minority workers who were rejected were at least as productive as the white males who were hired, a task that is very difficult for a third party (the government) to determine objectively.

2.2 Affirmative action/employment equity

If equal opportunity simply means that employers have to treat equally qualified applicants equally, minority workers may be at a disadvantage if discrimination has meant that they come to the job market with lower quality education than do majority group workers. Assume, for example, that discrimination has meant that blacks and women have obtained less education, and education of lower quality, than have white males. If employers use education to predict how productive workers are likely to be, whites will be given preference in the hiring process even if employers are not prejudiced.

Affirmative action programs seek to rectify this imbalance by setting quotas for hiring minority group workers, regardless of qualifications; while *employment equity* programs require that

preference be given to minority group applicants, until a target level of representation has been met. Of the many criticisms these programs face, two are heard most commonly: one based on politics and one on economics.

The political argument is that some of the blacks and females who are hired under these programs will have lesser qualifications than will some of the white male workers whose applications are rejected. If the latter group is able to muster sufficient sympathy from voters, politicians may become reluctant to introduce these policies.

The economic argument is that, if educational qualifications are reliable indicators of worker productivity, hiring workers who lack the desired education may simply set up those individuals for failure - which prejudiced employers and fellow employees may interpret as validation of their views, leading to an increase in discrimination, rather than the intended decrease.

Note, however, that this argument relies on the assumption that schools and colleges teach skills that are valuable to employers. If education is a "signal" of worker productivity, as was discussed in Module 40, it may be that minority workers who lack that signal are just as productive as white male workers who have above-average levels of education. In that case, the minority workers who are given jobs through affirmative action may perform as well as the "preferred" white males.

This signaling argument will be particularly relevant if discrimination had acted to reduce the quality of education that minority workers received as children. For such individuals, the cost of obtaining higher education, in terms of the effort it will require, will be greater than it is for white males. In such a case, if a minority worker completed, say, grade 12 or a junior college certificate, that may signal as much talent and perseverance as would completion of, say, a bachelor's degree by a white male. Affirmative action would not lead the employer to hire unproductive workers (assuming that education is just a signal.)

G. GOVERNMENT AS PRODUCER

MODULE 46. GOVERNMENTS AS PRODUCERS OF GOODS AND SERVICES

Because resources in our economy are allocated primarily by market forces, introductory courses in Economics focus on how markets operate. What is often overlooked is that even in the most capitalist of economies, more than a third of goods and services are produced by the government: we have a *mixed economy*. By concentrating only on market production, most economics courses fail to consider how a significant percentage of our economy's resources are allocated.

The purpose of the three modules in this section is to provide an introduction to the economics of government agencies. Module 46 investigates the question of why governments in developed economies produce some goods and services and not others. Module 47 asks how government agencies decide what to produce – most importantly, how they determine how much consumers are willing to pay for government-produced goods. Finally, Module 48 applies the principal-agent problem, first introduced in Module 39, to the question of how governments motivate their employees.

Most of us take it for granted that governments provide some kinds of goods and services, like police, roads, and schools, and that the private sector provides others, like groceries, clothes, and automobiles. But most government services were at one time provided by private firms; and many services, like health care and buses, that are provided by governments in some countries are provided by the private sector in others. The question of how this division came about is a complicated one that would require books to analyse effectively. This module is designed only to introduce this topic by identifying four economic roles that government agencies might serve in an otherwise market-oriented economy.

1. Externalities

If goods generate external benefits, private firms may not be able to charge beneficiaries for the full value of those goods and underproduction may occur. In some such cases, governments may establish their own agencies to produce the goods in question. Two classic examples of such agencies are public transit and police forces.

Public transit systems, such as buses and subways, provide external benefits by removing cars from public roads, thereby reducing commuting times for car drivers. Although the latter group might be willing to pay for those benefits, it would be difficult for a private firm to collect from them, primarily because of the free rider problem. (See Module 38.) If the government provides public transit, however, it can use taxes to make the beneficiaries of reduced commuting time pay for that benefit. In this sense, then, it is seen that it may be misleading to say that "taxpayers" "subsidise" public transit. Rather, the taxes represent a "price" that car drivers (willingly) pay in return for a benefit they receive.

[Note: the externality argument helps explain why governments are more likely to provide public transit *within* cities than *between* cities: inter-city buses have a lesser effect on highway congestion

than city buses have on city streets. Hence, the failure to "internalize" external benefits has a lesser effect in the former case than in the latter.]

Police forces may also provide external benefits. For example, if a group of citizens was to hire a private security firm to reduce burglaries, muggings, and dangerous driving in their neighbourhood, it is likely that the resulting reduction in crime would benefit individuals who had not helped to pay for security. And if individuals can benefit from an action without paying for it, there may be little incentive for them to make such a contribution. Again, taxes are a price that citizens pay for the external benefit that that a public service provides.

Other cases in which government agencies act to provide externality-producing goods and services include protection of endangered species and production of armed forces, lighthouses, and fireworks displays.

2. Monopoly

In Module 25, it was argued that the cost structures of some industries are such that it may be efficient to concentrate production in one company, a monopoly, even though monopolies generally charge higher prices and produce less output than is socially desirable. Two common examples are railways and utilities, such as electricity, natural gas, and water. In some cases, monopoly prices and outputs are controlled through government regulation; but in others, those firms are operated as government agencies.

In Module 27 it was argued that, in the latter case, government may separate the delivery of the product from its production – for example, ownership of railway tracks may form a government monopoly, separate from the private companies that provide railway rolling stock; and ownership of electrical transmission lines might become a government utility, separate from the private companies that generate the electricity.

3. Tolling

In some cases, the cost of collecting revenue is large enough that private firms will not produce a good even though the benefit obtained from that good exceeds the cost of production. For example, if revenue could only be obtained from the users of urban roads by establishing toll booths at each street corner, the cost of collection would quickly exceed the net benefit from those roads. A similar argument can be made with respect to neighbourhood parks, such as those that typically have a children's playground. In these cases, governments might use their powers of taxation to raise the "tolls" necessary for construction and maintenance of the desired facilities.

In the case of roads, this argument may be weakened in two circumstances. First, if drivers remain on one road for long distances, such as when travelling from one city to another on a major highway, it may be possible to control access to that road using only one or two toll booths. If this reduced the cost of collection below the net benefit from the road, it would not be necessary for the government to replace private firms as providers of highways.

Second, as electronic devices become less expensive, it may become cost effective to install sensors on urban streets that would record each driver's use of the road system. Firms could use the results of those records to charge drivers on a street-by-street basis. But not only would this raise significant privacy concerns, it would also create problems of monopoly power if owners of roads along particularly valuable routes were able to raise prices by restricting access.

4. Redistribution of Income

One of the most common arguments for government provision of goods and services is that this intervention is required to ensure that low income citizens receive basic necessities. If, in the absence of government involvement, the poorest members of society could not access minimally acceptable levels of these goods, the government might step in to produce and distribute them itself. For example, governments in developed countries provide education for all children up to the age of eighteen, and most provide health care to children and low-income adults.

But governments that wish to ensure that all citizens receive basic necessities do not always need to do so by producing those goods and services themselves. For example, western governments rarely produce or distribute food or clothing, and many have only limited programs for providing housing for the poor. Instead, they generally prefer to provide low-income citizens with income supplements, which the poor can use to purchase the necessary commodities from private firms.

The economic rationale for this policy is that individuals are generally in a much better position than is the government to decide what types of food, clothing, transportation, and entertainment are best suited for their families. (Interestingly, in this light, one situation in which governments do provide food is in long-term care facilities; and one of the most common complaints about those facilities concerns the quality and variety of the food that is offered.)

MODULE 47. GOVERNMENT CHOICES OF WHAT TO PRODUCE

Once the government has established agencies to produce goods and services - such as policing, education, water supply, national parks, and public transit – it has to decide what each of those agencies is going to produce, and how much. Which areas should police patrol most intensively? How many students should there be in each elementary school classroom? How should a city's wastewater be disposed of? Should private hotels, restaurants, and ski hills be allowed in national parks? Should special bus lanes be provided on major roads, to allow speedier service?

If these agencies were operating as private firms, they would use prices to measure both the value that consumers placed on their goods and the social cost of taking resources from other sectors to produce those goods. But if government agencies do not charge "consumers" for their products (like police and education), or charge only a subsidized price (like public transit or national parks), how will they measure the value of their products? This module discusses four general techniques that are commonly used to answer this question.

1. Elections

1.1 General election

The clearest way that citizens make their views known to the government is through their voting behaviour at elections. If voters don't like the current government's policies on education, public transit, national parks, or any of a myriad of other issues, they can vote for an opposition party at the next election.

But this is like buying a house by choosing between the first two your realtor shows you. You may prefer house A to house B, but that does not mean that A has all of the options you would be willing to pay for if you had been given the choice.

Similarly, you may prefer Party X to Party Y, but that does not mean that you agree with most, or even any, of X's policies. Given that government agencies make literally millions, if not billions, of decisions in the time between any two elections, your single vote, for either X or Y, will not provide information about your preferences for anything but a tiny subset of those decisions. Your vote might indicate, for example, that you would like the government to spend more money on policing, or that you would prefer that fewer immigrants be allowed into the country; but it provides only limited information about the types of detailed questions raised in the first paragraph of this module.

[Note also that, because the chance that an individual citizen's vote will have an impact on any election is virtually zero – one vote only affects the outcome if the vote would otherwise be tied - the benefit to any citizen from deciding carefully which party to vote for is also virtually zero. (You only benefit from deciding to vote for Party X instead of Party Y if your vote affects whether X wins the election.) But that means that the information about voter preferences that is provided

to government agencies by voting behaviour is very limited. Contrast this with a market-based economy, where the individual purchaser of, say, a house or car obtains a significant benefit from information collected about that purchase, information that can be used by the seller to provide a product that matches the buyer's preferences.]

1.2 Referenda

A type of election that meets some of the problems raised concerning general elections is a *referendum*. If voters are asked, for example, whether they think the government should spent $10 billion on a new rapid transit line, or $100 million on the purchase of tablet computers for the school system, the result is clear. Unlike a general election, there are no other issues that confound the interpretation of the vote.

Nevertheless, referenda are very imprecise instruments. A referendum that asks only whether voters would support the construction of a new rapid transit line does not provide information about voter preferences concerning that line's length or route, on whether it should be put underground as it passes through the city centre, or whether people who live near the proposed route should contribute more to the cost of its construction than should other citizens, among a whole host of other issues.

Furthermore, the referendum process provides only a limited opportunity for participants to suggest, or negotiate, alternative methods of achieving the desired outcome. Perhaps voters would have preferred an increase in the number of buses, or a widening of major streets; but if those are not offered on the referendum, they may vote for the "second best" option, the new rapid transit line, rather than have no change at all.

Also, as referenda are very expensive to run, it only makes (economic) sense to use them to decide major issues.

2. Consumer Surveys

A direct way to measure consumers' preferences is to conduct surveys. Three types of surveys will be considered here: public opinion polls, willingness to pay, and choice experiments.

2.1 Opinion polls

If the government is considering ways to increase access to downtown, it could, for example, conduct an *opinion poll* of a random selection of citizens to see whether they preferred wider roadways, more buses, or a new rapid transit line. And if they preferred rapid transit, they could be asked whether they think it should start, say, ten kilometres from the city centre, or twenty; whether the stations should be one kilometre apart or two; whether the line through the city centre should run underground or over ground through the city centre; etc.

Although this approach may be attractive, it is also full of potential hazards, most of which arise when respondents lack adequate information concerning the implications of their answers. Three types of information are of particular importance: price, opportunity cost, and impact.

Price: It would not make sense to ask consumers whether they would like a nicer car or bigger house without specifying what the price of those items will be. For the same reason, before a survey of voter preferences can provide useful information, it must indicate what the cost to citizens would be if the proposal they agreed to was put into effect. Survey participants might well say "yes" to the question "do we need more elementary schools?" but "no" to the question "should we spend $400 million building new elementary schools?" For example, it often appears that survey participants are more likely to say that they are concerned about climate change than they are to say that they would like to raise taxes to pay for policies to reduce carbon emissions.

Importantly, any survey must be careful to specify "price" in a manner that is meaningful to respondents. For example, voters in a city of a million might balk at a proposal if they are told that it will cost $1 billion; but accept the same proposal if they are told that the cost will be amortized over twenty-five years, causing annual taxes to increase by $50 per person.

Opportunity cost: That a government action carries no monetary price does not imply that it has no cost. As most decisions to adopt one policy imply that some other policy is *not* adopted, the cost of the selected policy is the opportunity lost to adopt its alternatives. If a school board decides to teach reading using a phonetic system, it gives up the opportunity to use the whole language approach. If a city decides to set aside bike lanes on city centre streets, the money cost may pale compared to the congestion costs imposed on car drivers. For surveys about either of these issues (and most other public policy issues) to provide meaningful information about voter preferences, those surveyed must be made aware of the opportunity costs of each option presented to them.

Impact: Perhaps the most difficult aspect of policy to convey to voters is potential impact. For questions about police patrols, high school class sizes, or rapid transit lines to be meaningful, respondents must be informed about the anticipated impact of those policies. What effect will additional police patrols have on public safety? Is there evidence that class sizes are correlated with student learning? Would failure to build a new rapid transit line cause firms to flee the city centre and, if so, what would the impact be on economic growth?

Summary: That these issues must be taken into account in order to build meaningful voter surveys does not mean that those surveys cannot be useful indicators of public opinion. It does mean, however, that those surveys will often be expensive to construct and administer, making them less useful as a method of determining day-to-day decisions.

2.2 Willingness to pay

Instead of asking survey participants whether they would like the government to produce a given good, given the estimated cost, economists often suggest that survey respondents be asked to indicate how much they would be *willing to pay* for that good. For example, the survey might ask each respondent: "In a referendum, would you vote to build a new rapid transit line if annual taxes

had to be increased by $X?" By varying the value of $X that was presented to respondents, the survey would be able to trace out a "demand" curve for the project. For example, respondents might be divided into four equal groups who were presented with values for $X of $40, $30, $20, and $10, respectively. If ten percent of the group who were presented with a $40 price tag replied that they would vote for that option, 40 percent of those presented with $30 said they were willing to vote for it, 60 percent of the third group would vote for $20, and 90 percent of the fourth would vote for $10, it might be concluded that 50 percent of citizens would vote for the proposal if the cost was an increase in annual taxes of approximately $25.

There are numerous criticisms of the willingness to pay approach, two of which are particularly compelling. First, the technique is only well suited for yes/no questions, like "should the government build a new rapid transit line: yes or no?" But most decisions that government agencies make require decisions among numerous possible levels of provision. For example, the question concerning school class size is not whether the number of students per class should be decreased; it concerns the class size citizens would prefer. It might be useful to know that they would be willing to pay an increase in taxes of $10 per year to decrease class sizes from 30 to 28, but that does not tell the Department of Education whether 28 is the preferred class size – they might prefer an increase of $15 with a class size of 27. A more complex method of eliciting preferences will be required.

Second, as citizens have little experience purchasing government-provided goods and services, like policing and education, it is often very difficult for them to estimate accurately what their willingness to pay might be. Survey respondents might be able to rely on their own past experiences to guess how much they would be willing to pay for a new model of car or smart phone. But, as they will have had little experience "purchasing" police services or protection of endangered species, their estimates of willingness to pay may be highly inaccurate.

As an example, environmental agencies have often conducted surveys to determine how much citizens would be willing to pay to protect endangered species. In the United States, the answer to this question is often in the range of $10 to $20 per year per species. But there are over a thousand endangered species in the United States. Extrapolating the figure for one species to the total seems to imply that respondents would be willing to pay between $10,000 and $20,000 per year for endangered species protection, numbers that are simply not credible.

2.3 Choice experiments

To overcome the problem that willingness to pay surveys allow for only "yes/no" answers, economists and statisticians have developed a technique known as a *choice experiment*. This approach is an extension of the willingness to pay technique. But now, instead of being given varying options for the price they are willing to pay for a given project, respondents are given varying combinations of prices and project characteristics.

In the example of class size, given above, one set of respondents might be asked to choose between two options: (i) paying $10 per year to decrease class sizes to 28 and (ii) paying $20 per year to decrease class sizes to 27; another group of respondents might then be asked to choose between

$10 to reduce class size to 29 and $30 to reduce class size to 26. By varying the dollar cost/class size combinations offered to alternative respondents, it is possible to use statistical analysis to obtain an estimate of the willingness to pay for alternative levels of government service.

By increasing the number of characteristics, choice experiments can obtain increasingly sophisticated measures of consumer preferences. For example, in the case of a new rapid transit line, survey respondents might be asked to choose between options each of which combined a different combination of projected tax increase, total length of the line, distance between stations, and surface versus subsurface route through the city centre.

The primary drawbacks to the choice experiment technique are that it needs very large sample sizes (in order to offer sufficient numbers of options) and that its interpretation requires the use of highly sophisticated statistical techniques. As both of these add significantly to the cost, choice experiments are not often used.

3. Special Interest Groups

One possible measure of consumer demand for government services might be the intensity of requests/demands from groups of like-minded individuals, often called *special interest groups*. For example, the Parent-Teacher Association (PTA) might bring a petition from its members for a reduction in class size; a downtown business association might press for a new rapid transit line to be routed underground; and a consortium of environmental groups might demand restrictions on commercial use within national parks.

But the intensity of these demands is a very poor measure of the value of the projects that are being requested. First, such demands are often not accompanied by offers to pay for them. Special interest groups simply assume that "the government" will pay - government projects are treated as if they were free, creating a free rider problem. To say that you would like to have something if it is provided to you for free is not an indication that you believe that that project offers a net benefit to society.

Second, each request by one special interest group is often met by similar requests by other groups that something different be done. Although the downtown business group may prefer an underground route for the transit line, the taxpayer's association may prefer a cheaper, over ground route; and whereas the environmental consortium may demand a ban on hotels and ski hills in parks, business groups and skiers may press for an extension of those amenities. Once there are two or more groups demanding inconsistent policies, government agencies are left in the almost impossible situation of determining which groups' preferences are stronger than the others'. Should the transit route go underground because the benefits to the businessmen's association exceeds the costs to taxpayers? And how should the park service compare the benefits that skiers obtain from additional ski hills against the benefits that environmental activists obtain from untrammeled wilderness?

4. Public Participation

One approach to resolving conflicts among different interest groups might be to bring the various groups together and require that they negotiate a policy that is acceptable to all of them. Essentially, resources would be allocated, not by buying and selling resources in the market, but by bartering those resources – by "trading" the rights to one resource for the rights to another. The taxpayer's association might agree to a small increase in taxes in return for an agreement from the downtown business association that the transit route would go underground for only a few blocks, or that the route would proceed above ground but with underpasses for major streets. And environmentalists might agree to accept a new ski hill if skiers agreed to place that hill in a "second best" location that imposed only minor effects on wildlife.

Even in optimal situations, however, this proposal has proven difficult to implement because trade can only occur if each party "owns" a right that is desired by the other. For example, assume that an over ground transit route has already been built through the city centre. If the business association would prefer to reroute it underground, it is not clear what it could offer to the taxpayer's group in return for it agreeing to that. Similarly, if a ski hill already exists in a national park, it is not clear what environmentalists could offer to the operator in return for its agreement to move or change its operation.

MODULE 48. GOVERNMENT AS EMPLOYER

Once the government establishes an agency to produce goods and services for it, it becomes an employer; and, as employer, it needs to align the incentives that it offers its employees with the goals of the government: the principal-agent problem (see Modules 41 and 42). In this module, a distinction will be made between the incentives given to senior administrators (for example, the presidents and vice-presidents of universities) and those given to lower level employees (for example, professors and office staff).

1. Administrators

1.1 Economics of bureaucracy

The study of how senior government administrators behave is referred to (by economists) as the *economics of bureaucracy*, after the title of one of the first books to consider this issue, William Niskanen's *Bureaucracy and Representative Government*. This section introduces a simplified version of his model.

Assume that the government has established a single agency to produce a particular service – for example, a public transit system, a school board, a national park service, or a university. Because that agency is the sole provider of this service, it is in effect a monopoly. The "demand curve" facing this monopoly plots average government expenditure per unit of output against numbers of units of output.

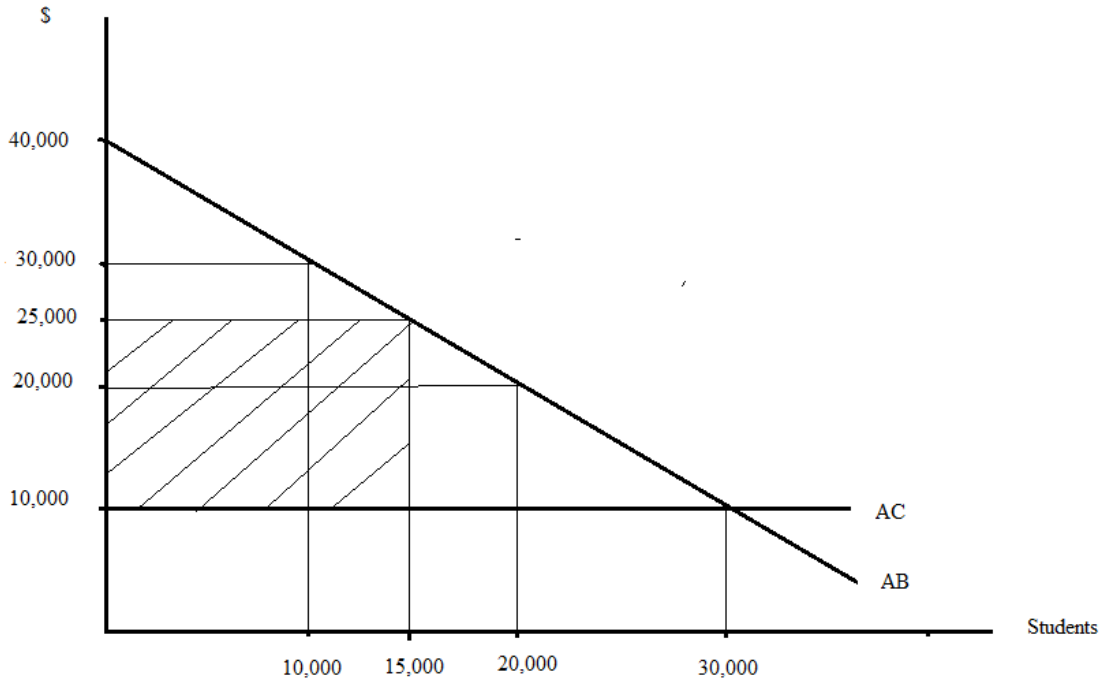
Assume, for example, that Figure 16 represents the cost and benefit information for a single university (or university board that is responsible for all of the universities that are funded by the government). Curve AB represents the maximum budget *per student* that the government is willing to devote to this university, at each number of students. It indicates that, if the average cost of educating each student is $30,000, the government will be willing to fund a university enrolment of 10,000 students. If the university offers to educate students at a cost of $20,000 each, the government will wish to expand university enrolment to 20,000; and if the cost of education is $10,000, the government will agree to enrolment of 30,000. If the budget expended per student is thought of as the "price" of that student's education, then the AB curve in Figure 16 represents the government's demand curve for university students.

Curve AC in Figure 16 represents the average cost of educating each student. For simplicity, it has been assumed that this cost, which is composed primarily of the opportunity costs of buildings and employees, is horizontal; that is, it does not change with the number of students. In this case, average cost has been assumed to be constant at $10,000, implying that AC is horizontal at that figure.

Given Figure 16, Niskanen asks: how can we expect the senior administrators at this university to act? To answer this question, he makes two assumptions. First, he notes that, as the university is the sole supplier to the government, it has a monopoly in university services. Following from this,

he assumes that the university administrators are free to charge the government any of the "prices" on curve AB.

Figure 16 Maximisation of Budget Surplus



His second assumption is that administrators, like most other individuals in society, act as if they are self-interested; and that this implies that they will choose the price-student combination that maximises the difference between total revenue (measured by the size of the university's budget) and total cost. In Figure 16, you can see that this will not occur at 30,000 students because, at that level, average budget equals average cost – the difference is zero.

Rather, the maximising rule will require that the university choose a lower level of enrolment. One such level is 20,000, at which the price (average budget) per student is $20,000 and cost per student is $10,000, creating a *budget surplus* of $200 million (= ($20,000 - $10,000) x 20,000).

With a little experimentation, it can be seen that the enrolment level that *maximises* the budget surplus is 15,000, at which the price per student is $25,000 and the surplus is $225 million (= ($25,000 - $10,000) x 15,000). This surplus is depicted as the shaded area in Figure 16.

The reason that administrators might act to maximise the budget surplus is that, as that sum is not necessary for the "production" of education, it can be directed to purposes that benefit the administrators themselves. There are at least three such uses.

Maximise personal consumption: The budget surplus can be used to increase the personal consumption levels enjoyed by senior administrators. This may imply increases in personal income; but it can also arise from improved fringe benefits, such as pensions and vacation time, from generous travel allowances, for example travel to conferences by first class air fares at five-star hotels in desirable tourist destinations, and from luxury office space and furnishings.

Minimise personal workload: Some of the budget surplus could be used to reduce the administrator's workload, for example by hiring workers to take on many of the tasks that would otherwise have been the administrator's responsibility.

Reduce labour strife: As one of the most stressful aspects of the senior administrator's job is dealing with demands from lower level employees – for higher wages, greater fringe benefits, longer vacations, improved working conditions, etc. – some of the budget surplus might be used to reduce this stress by making concessions to those employees. Thus, it will not only be administrators who will have better pay and working conditions than their private sector counterparts, but so will lower level employees.

1.2 Government response

If the model of bureaucracy presented here provided an accurate description of government administrators' behaviour, it would be expected that the government would take actions to restrain that behaviour. Four common responses from the government are: encourage competition among agencies, contract out some services to the private sector, benchmark government employees' compensation packages against those available in the private market, and introduce employee incentive plans.

Competition: In the model of bureaucracy, administrators are able to maximise their budget surpluses because their agencies are assumed to act as monopolies. To reduce this ability, the government could create competing agencies. It might, for example, establish more than one university and encourage them to compete with one another, in terms of both cost and quality of service.

Implicitly, this policy is used by many school boards. Traditionally, as school students were only allowed to attend the public school associated with their geographic community, each school had a monopoly in its district. Recently, however, many school boards have agreed to allow students to transfer to any school they wished (within the area served by the board), with the budget for students "following" them from one school to another. For example, if the average expenditure per student is $7,000, when a student moves from School 1 to School 2, School 1's budget is reduced by $7,000 and School 2's is increased by that amount. In this way, schools are encouraged to compete for students by offering improved quality of instruction.

[But the ability of governments to create competition is limited by the fact that most government agencies are able to obtain significant economies of scale, (such as are available to utilities and public transit), which would be lost if agencies were to be divided into smaller units.]

Contracting out: The costs charged by a government agency can be constrained if private firms compete, on the basis of price and quality, for the right to provide those functions. Highways and government buildings, for example, are commonly built by private construction companies, and janitorial and food services in public hospitals are often provided by private contractors. NASA's decision to invite private firms (including Boeing, SpaceX, Blue Origin, and Dynetics) to compete for the contract to build and launch a space shuttle is an example of contracting out.

Benchmarking: The salaries paid by government agencies can be constrained if it is possible to compare them to salaries in the private sector. This, of course, requires that there are comparable workers in the two sectors, which is often not the case. For example, in countries with a nationalised health service, there will be very few doctors, nurses, and laboratory technicians working outside that service; and few countries will have private markets for elementary school teachers or classics scholars.

Employee incentive plans: Some abuses might be avoided if the government could establish incentive schemes that induced government employees to adopt behaviours consistent with maximization of the government's goals. These plans are discussed in the next section.

2. The Principal-Agent Problem in Government Employment

Governments face the same principal-agent problems as are faced by private firms – that their employee's goals are different from those of the employer, and that the employer may face difficulty measuring the effort expended by its employees. Hence, it would be expected that government agencies and private firms would employ similar incentives to encourage their employees to work hard towards the employer's goals. And, to a large part, that is what is observed.

But government agencies face a problem that is not shared with the private sector. Whereas the private firm's goal may be clear – to maximise profits – the goals of the government agency are much less easily defined. For the reasons raised in Module 47, governments cannot easily measure what voter preferences are among the huge number of choices available to them. And if an agency is not clear what its goals are, it will have great difficulty providing incentives to employees to meet those goals.

What are "the goals" of a school board, for example? Is it to "produce" as many students as possible, at as small a cost as possible? Is it to produce students who are able to think and write critically; or is it to provide workers to fill underserved occupations in the labour force? Should teachers provide social and psychological mentoring? How important is physical activity? To what extent are school institutions for discouraging racial and sexual discrimination? Whereas private schools can "answer" these questions by observing which characteristics parents are willing to pay the most for, public schools have fewer, direct measures of their success and, therefore, less ability to determine whether their teachers and principals have achieved the (implicit) goals society expects of them.