INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)” If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insuire you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
PAUL, David Terry, 1943-
A PRIMER IN ECONOMICS.
The Ohio State University, D.A., 1973
Economics, general

University Microfilms, A XEROX Company, Ann Arbor, Michigan

Copyright © 1973
by David Terry Paul
All Rights Reserved
A Primer In Economics
Professional Project

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Arts in the Graduate School of The Ohio State University.

By
David Terry Paul, B.B.A., M.B.A., M.A.

The Ohio State University
1973
# A Primer in Economics

## Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vita</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>iii</td>
</tr>
<tr>
<td>1</td>
<td>Scarcity and Alternative Cost</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Comparative Advantage, Specialization and Trade</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Demand</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>Cost and Supply: Profit Maximization and Competition</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>Cost and Supply: Marginal Cost and the Supply Curve</td>
<td>43</td>
</tr>
<tr>
<td>6</td>
<td>Applications of Supply and Demand: Shortages</td>
<td>53</td>
</tr>
<tr>
<td>7</td>
<td>Some Simple Ways Not to Cure the Farm Problem</td>
<td>65</td>
</tr>
<tr>
<td>8</td>
<td>International Trade—Specialization by Countries</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>Inflation—A General Rise in Prices</td>
<td>91</td>
</tr>
<tr>
<td>10</td>
<td>The Meaning of Profit Maximization</td>
<td>107</td>
</tr>
<tr>
<td>11</td>
<td>Competition and Economic Efficiency</td>
<td>122</td>
</tr>
<tr>
<td>12</td>
<td>Competition and Discrimination I</td>
<td>129</td>
</tr>
<tr>
<td>13</td>
<td>Competition and Discrimination II</td>
<td>140</td>
</tr>
<tr>
<td>14</td>
<td>The Economics of Pollution</td>
<td>153</td>
</tr>
<tr>
<td>15</td>
<td>Seemingly Non-Competitive Conditions in Markets: Quality, Price Differences, and Advertising.</td>
<td>161</td>
</tr>
<tr>
<td>16</td>
<td>Monopoly--Markets Insulated from Competition</td>
<td>176</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>17</td>
<td>The Market for Labor as a Factor of Production</td>
<td>189</td>
</tr>
<tr>
<td>18</td>
<td>Work Disincentives in Labor Markets</td>
<td>203</td>
</tr>
<tr>
<td>19</td>
<td>The Theory of Investment, as Applied to Human Beings</td>
<td>214</td>
</tr>
<tr>
<td>20</td>
<td>Introduction to Macroeconomics</td>
<td>232</td>
</tr>
<tr>
<td>21</td>
<td>The Determination of the Aggregate Price Level, Income, and the Rate of Interest</td>
<td>241</td>
</tr>
<tr>
<td>22</td>
<td>Macroeconomic Policy: Theory and Practice</td>
<td>252</td>
</tr>
</tbody>
</table>
May 18, 1943

1965

B.B.A., University of Cincinnati, Cincinnati, Ohio

1966

M.B.A., University of Cincinnati, Cincinnati, Ohio

1966-71

Teaching Assistant, Department of Economics, The Ohio State University, Columbus, Ohio

1971

M.A., The Ohio State University Columbus, Ohio

1971-73

Lecturer, Department of Economics, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Economic Theory
Labor Economics
Economic History
Soviet Economy
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Production Possibility Frontier for Turnips and Pot</td>
<td>8A</td>
</tr>
<tr>
<td>2</td>
<td>Shifts in the Production Possibility Frontier for Turnips and Pot</td>
<td>10A</td>
</tr>
<tr>
<td>3</td>
<td>Production Possibility Frontiers for Appropriate and Inappropriate Specialization</td>
<td>15A</td>
</tr>
<tr>
<td>4</td>
<td>A Demand Curve</td>
<td>24A</td>
</tr>
<tr>
<td>5</td>
<td>Observations of Quantities Demanded</td>
<td>24A</td>
</tr>
<tr>
<td>6</td>
<td>A Typical Marginal Cost Curve</td>
<td>36A</td>
</tr>
<tr>
<td>7</td>
<td>Several Hypothetical Marginal Cost Curves</td>
<td>37A</td>
</tr>
<tr>
<td>8</td>
<td>The Production Possibility Frontier for Turnips and All Other Goods</td>
<td>43A</td>
</tr>
<tr>
<td>9</td>
<td>The Two General Shapes of Marginal Cost Curves</td>
<td>45A</td>
</tr>
<tr>
<td>10</td>
<td>Continuously Diminishing Marginal Cost</td>
<td>45A</td>
</tr>
<tr>
<td>11</td>
<td>Demand and Marginal Revenue Curves for the Competitive Firm</td>
<td>47A</td>
</tr>
<tr>
<td>12</td>
<td>Demand and Marginal Revenue Curves for the Non-Competitive Seller</td>
<td>47A</td>
</tr>
<tr>
<td>13</td>
<td>Marginal Cost and Marginal Revenue Intersection</td>
<td>48A</td>
</tr>
<tr>
<td>14</td>
<td>Demand and Supply Conditions for the Industry and the Firm</td>
<td>50A</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>15</td>
<td>A Shift in Supply Caused by Technological Change</td>
<td>52A</td>
</tr>
<tr>
<td>16</td>
<td>Supply, Demand, and Three Alternative Prices</td>
<td>54A</td>
</tr>
<tr>
<td>17</td>
<td>The Effects of Temporary Shifts in Demand Upon Prices and Quantities</td>
<td>55A</td>
</tr>
<tr>
<td>18</td>
<td>Marginal Cost Curves for Three Levels of Production Efficiency</td>
<td>67A</td>
</tr>
<tr>
<td>19</td>
<td>The Industry Supply Curve, Apportioned by Type of Producer</td>
<td>68A</td>
</tr>
<tr>
<td>20</td>
<td>Market for Textiles - U.S. Domestic Market Only</td>
<td>77A</td>
</tr>
<tr>
<td>21</td>
<td>U.S. Supply and Demand Functions for Textiles to and from the Rest of the World</td>
<td>78A</td>
</tr>
<tr>
<td>22</td>
<td>The Japanese Textile Market</td>
<td>78B</td>
</tr>
<tr>
<td>23</td>
<td>The U.S. and Japanese Markets for Textiles</td>
<td>78C</td>
</tr>
<tr>
<td>24</td>
<td>The Effects of Interference With Trade</td>
<td>87A</td>
</tr>
<tr>
<td>25</td>
<td>The Effect of a Quota Upon Trade</td>
<td>89A</td>
</tr>
<tr>
<td>26</td>
<td>Prices, Costs, and Shifts in Demand</td>
<td>98A</td>
</tr>
<tr>
<td>27</td>
<td>Various Levels of Profit and Output at a Point in Time</td>
<td>110A</td>
</tr>
<tr>
<td>28</td>
<td>The Effect of Time on Profitable Levels of Output</td>
<td>110A</td>
</tr>
<tr>
<td>29</td>
<td>The Total Revenue and Total Cost Curves</td>
<td>117A</td>
</tr>
<tr>
<td>30</td>
<td>The Marginal Revenue and Marginal Cost Curves</td>
<td>118A</td>
</tr>
<tr>
<td>31</td>
<td>MR, MC, TR, TC Curves Shown Together</td>
<td>118B</td>
</tr>
<tr>
<td>32</td>
<td>Two Ways of Finding the Profit Maximizing Output</td>
<td>120A</td>
</tr>
<tr>
<td>33</td>
<td>The Efficient Allocation of Output Between Two Firms</td>
<td>126A</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>34</td>
<td>The Demand for Segregated and Integrated Housing</td>
<td>130A</td>
</tr>
<tr>
<td>35</td>
<td>The Effects of Adequately Enforced Fair Housing Laws</td>
<td>133A</td>
</tr>
<tr>
<td>36</td>
<td>Profitable Investment in Employees by Firms</td>
<td>144A</td>
</tr>
<tr>
<td>37</td>
<td>The Effects of Effectively Enforced Fair Employment Laws</td>
<td>147A</td>
</tr>
<tr>
<td>38</td>
<td>The Appropriate Allocation of Production with Pollution as a Byproduct</td>
<td>158A</td>
</tr>
<tr>
<td>39</td>
<td>Supply and Demand Revisited</td>
<td>161A</td>
</tr>
<tr>
<td>40</td>
<td>Two Kinds of Demand Curves</td>
<td>161A</td>
</tr>
<tr>
<td>41</td>
<td>The Contrast Between the Competitive and Non-Competitive Equilibrium Positions</td>
<td>176A</td>
</tr>
<tr>
<td>42</td>
<td>Continuously Declining Marginal Cost</td>
<td>180A</td>
</tr>
<tr>
<td>43</td>
<td>The &quot;Backward Bending&quot; Supply Curve of Labor</td>
<td>193A</td>
</tr>
<tr>
<td>44</td>
<td>The Choice of Income Versus Leisure</td>
<td>193A</td>
</tr>
<tr>
<td>45</td>
<td>Result of Gain in Non-Wage Income</td>
<td>197A</td>
</tr>
<tr>
<td>46</td>
<td>Result of Increase in Wage Rate</td>
<td>197A</td>
</tr>
<tr>
<td>47</td>
<td>Female Wages Relative to Male Wages</td>
<td>202A</td>
</tr>
<tr>
<td>48</td>
<td>Wage-Hour Path for Male Workers</td>
<td>202A</td>
</tr>
<tr>
<td>49</td>
<td>The Demand for Investment</td>
<td>228A</td>
</tr>
<tr>
<td>50</td>
<td>The Market for Loanable Funds</td>
<td>235A</td>
</tr>
<tr>
<td>51</td>
<td>Loanable Funds and Commodity Markets at Equilibrium</td>
<td>237A</td>
</tr>
<tr>
<td>52</td>
<td>Supply of Loanable Funds from Commercial Banks</td>
<td>239A</td>
</tr>
<tr>
<td>53</td>
<td>Commodity Market Equilibrium</td>
<td>244A</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>54</td>
<td>Simultaneous Equilibrium of Loanable Funds and Commodities Markets</td>
<td>245A</td>
</tr>
<tr>
<td>55</td>
<td>A Reduction in Investment Demand</td>
<td>248A</td>
</tr>
<tr>
<td>56</td>
<td>Re-Attainment of Equilibrium</td>
<td>250A</td>
</tr>
<tr>
<td>57</td>
<td>The Keynesian View of the Effect of an Increase in Government Spending</td>
<td>258A</td>
</tr>
<tr>
<td>58</td>
<td>The Neo-Classical View of an Increase in Government Spending</td>
<td>260A</td>
</tr>
<tr>
<td>59</td>
<td>Macroeconomic Conditions and Policy Actions 1929-33</td>
<td>264A</td>
</tr>
<tr>
<td>60</td>
<td>Macroeconomic Conditions and Policy Actions 1965-69</td>
<td>268A</td>
</tr>
</tbody>
</table>
INTRODUCTION

This text differs from most "principles of economics" texts in several important respects. It is designed to provide a concise treatment of the essential elements of economic theory appropriate for a one-term introductory course. It was designed to allow an instructor the maximum amount of flexibility in his teaching, especially the option of assigning additional readings, emphasizing those areas he would choose. It was anticipated that this text would be a modestly priced (perhaps paperback) primer in economic theory.

This text treats microeconomic theory and its applications at much greater length than their macroeconomic counterparts. This emphasis is not meant to imply that the former area is intrinsically more important than the latter. Rather, this choice was made for a number of other reasons. The first is the belief that microeconomic theory provides the basic "tool kit" for all of economic analysis. The microeconomic foundation of macroeconomic theory has become more apparent with recent developments in the literature and the macro section of this text reflects that orientation. The second reason for the microeconomic emphasis is our belief that the level of "economic literacy" concerning microeconomic issues is considerably lower than for macroeconomic. While there seems to be a certain public understanding of many macroeconomic concepts such as national income, unemployment, and fiscal policy, an appreciation of the relevance and importance
of the microeconomic aspects of many public and practical issues is seriously lacking. The text stresses these applications. Finally, we believe that microeconomic theory, once it is presented, makes more intuitive sense than macroeconomic theory and that treating the former prior to that latter is more effective from a purely pedagogical standpoint.

The professional reader will note that there is also a strong emphasis upon competitive economic theory and that the author views the American economy as essentially competitive. In the last analysis, this question is, of course, an empirical rather than a theoretical one. This issue is confronted directly in Chapters 15-16. But the competitive theory forms the basis for discussion of every chapter.

The text does provide more than simply the essential elements of economic theory, however. The format is designed to present an integration of economic theory and applications of the theory. Microeconomic theory (essentially supply and demand) is first presented in Chapters 1-5, followed by four chapters in which the theory is applied to specific problem areas. Again this format is repeated with the presentation of the economic theory of the firm (chapter 10-11) and further applications (discrimination and pollution in chapters 12-14). Even within “theory” chapters, the explanation relies upon examples and within “applications” chapters there is a stress upon the underlying theory. Finally, the microeconomics section concludes with an examination of some aspects
of labor markets in chapters 17-19. The same format of jux-
taposing theory and applications is repeated in the chapters
on macroeconomic theory (20-22). The focus here is upon a
macroeconomic model that is consistent with either the Keynesian
or Monetarist positions and an examination of recent policy
actions and results.

This text was written under the supervision of Dr. Belton
Fleisher and Dr. Kenneth Kopecky both of the Ohio State Uni-
versity Department of Economics. It is no exaggeration to
state that without their constant aid and encouragement this
project would never have been completed. I hope that this
project is worthy of their help.
Chapter 1. Scarcity and Alternative Cost

Scarcity and the Three Economic Questions

Scarcity is the fundamental condition of economic life; therefore, choices have to be made among desirable alternatives. This process of choice is a matter of allocation, the allocation of scarce resources among alternative uses. For example, a housewife must decide how to spend (allocate) her family's limited income among a choice of many alternative commodities. Most of us, at least, cannot have as much of everything as we would like and must decide among different things, all of which we would like to have. We shall see that these choices depend upon the prices of different goods, and the buyers' incomes and tastes.

The entire body of economic theory rests upon this concept of scarcity. Because available resources are scarce relative to the many alternative uses to which we might put them, choices among these alternatives must be made. Economics is sometimes called the "science of choice," in fact, a recent text in economics incorporated that definition in its title.

There are really three kinds of choices that any economic system must resolve. These choices are three economic questions which arise in a world characterized by scarcity. These questions are a direct consequence of scarcity and must be answered in some way, regardless of the institutions and ideology of an economy. Institutions and ideology may influence the answers
but the questions are so fundamental that they transcend any particular set of answers. These three economic questions are:

1) What goods (and in what quantities) will be produced?
2) How will the chosen goods be produced?
3) Who shall receive the goods produced?

The "what" question is probably the most obvious consequence of scarcity. Since resources are limited relative to wants, choices must be made among alternative uses. Available resources are insufficient to produce all of everything people want so, somehow, decisions must be made to decide what goods (and in what quantities) shall and shall not be produced.

The "how" question is relevant only under conditions of economic scarcity. Given such an environment where resources are limited, it behooves an economy to use those methods which best utilize the available resources. Since it is impossible to produce "everything," it is obviously desirable to produce as much as is possible. There is almost always more than one way to produce any output. The best way is that method which produces the most. Another definition of the best method is that way which makes the cost the least. This latter way of viewing the problem comes close to the popular connotation of "economizing". We shall see that these two perspectives are generally equivalent. They both emphasize the desirability of finding the most appropriate method of production.

The "who" question has traditionally been of greatest concern to individuals. Whose wants will be satisfied and
to what degree? How big a "slice of the economic pie" will each receive? As long as recourses remain relatively scarce, some method of apportioning among individuals available output must be employed.

You should see that these questions are interrelated and that they all are consequences of the fundamental condition of scarcity. For instance, the answer to "what" relates to "who" since individuals' wants differ. Similarly, "how much" is produced depends upon "how" production is undertaken, and so on. We have only discussed these three questions in a general way. The rest of the text, and indeed the entire study of economics, is an exploration of some aspects of these three questions.

**Turnips and Pot: A Problem of Scarcity**

Let us consider the problem of scarcity in a way that is at once simpler and more complex than the way we normally face it as consumers of goods and services. Imagine a small self-sufficient economy where only two goods are produced and consumed, turnips and pot. What shall be produced and how? (We shall ignore for the time being the problem of deciding who shall receive what is produced.) Since there are only two goods in this example to choose between, the choice may seem easier than the ones we all make. The choice is complicated, however, by the fact that technical production decisions must also be made. That is, simultaneously, the "what" and "how" questions
must be answered. Both kinds of decisions are manifestations of the underlying economic condition, scarcity.

To simplify our problem, suppose that both turnips and pot require the application of two scarce resources, labor and land. Labor is scarce because no one wants to work all the time and even if someone did, there are only so many hours in a day. We will assume that both turnips and pot require the same amount of labor input per unit (bushel) of output. Turnips and pot differ, however, in the amounts of land they require to produce a unit of output. Land is also a limited resource and comes in two types that differ in their capacities to produce pot and turnips. This information is summarized in the table below.

<table>
<thead>
<tr>
<th>TYPE OF LAND</th>
<th>QUANTITY AVAILABLE</th>
<th>TURNIP OUTPUT PER ACRE OF LAND</th>
<th>POT OUTPUT PER ACRE OF LAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;A&quot;</td>
<td>10 acres</td>
<td>3 bushels</td>
<td>4 bushels</td>
</tr>
<tr>
<td>&quot;B&quot;</td>
<td>10 acres</td>
<td>2 bushels</td>
<td>2 bushels</td>
</tr>
</tbody>
</table>

The problem of deciding how to produce our pot and/or turnips (once we have decided what we want) boils down to a matter of deciding how to use our twenty acres or "what to produce where?". This decision is made on the basis of costs or prices. At first you may not see exactly where cost or price is involved in this problem. Once it is brought up, however, a glance at the table probably suggests to you that
a bushel of turnips "costs" 1/3 an acre on type A land since one acre yields three bushels of turnips and that pot "costs" 1/2 an acre of land per bushel on type B land and so on. The object is to produce our commodities at the lowest cost (or equivalently, to maximize output from a given amount of cost).

Looking at costs in this intuitive way, you would say that both turnips and pot are "cheaper" on type A land. Although this is perhaps the most natural way to look at things, it turns out to be the wrong way when we seek to do our best in coping with the problems of scarcity. To see why this is true, suppose that we had decided to produce exactly six bushels of turnips and devote the rest of our resources to the production of pot. "Obviously" the cheapest way to produce 6 bushels of turnips is to use 2 acres of type A land, leaving us with 18 remaining acres of land for pot, you might be inclined to say. But look at the table below and you will see that if you had grown the 6 bushels of turnips on 3 acres of type B land instead, you would have had more pot to go along with your turnips. Clearly, the "right" answer must be to grow the turnips on B land, but why?

The key to understanding this turn of events is determining how the 17 acres devoted to pot production when turnips are grown on B land yield 54 bushels of pot, 2 bushels more than the 18 acres left for pot production when turnips are grown on A land. Consider again the first table. Each acre
If Turnips Grown on Type A Land | If Turnips Grown Type B Land

<table>
<thead>
<tr>
<th>Acres of A land devoted to Turnip Production and Output of Turnips</th>
<th>2 acres, 6 bushels of turnips</th>
<th>0 acres, 0 bushels of turnips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acres of A land devoted to Pot Production and Output of Pot</td>
<td>8 acres, 32 bushels of pot</td>
<td>10 acres, 40 bushels of pot</td>
</tr>
<tr>
<td>Acres of B land devoted to Turnip Production and Output of Turnips</td>
<td>0 acres, 0 bushels</td>
<td>3 acres, 6 bushels of turnips</td>
</tr>
<tr>
<td>Acres of B land devoted to Pot Production and Output of Pot</td>
<td>10 acres, 20 bushels of pot</td>
<td>7 acres, 14 bushels of pot</td>
</tr>
<tr>
<td>TOTAL OUTPUT</td>
<td>6 BUSHELS OF TURNIPS, 52 BUSHELS OF POT</td>
<td>6 BUSHELS OF TURNIPS, 54 BUSHELS OF POT</td>
</tr>
</tbody>
</table>

of B land will yield 2 bushels of pot or 2 bushels of turnips. Using one half an acre of land to produce one bushel of turnips "costs" one bushel of pot since that is the alternative use of the land. By the same reasoning, a bushel of turnips "costs" $\frac{4}{3}$ of a bushel of pot if it is grown on A land. You can now see how the 17 acres of land remaining when turnips were grown on B land produced more pot than the 18 acres remaining when turnips were grown on A land. The cheapest (most efficient) way to produce turnips is to use B land as long as it is available. As an exercise, you should convince yourself that the cheapest way to produce pot (in terms of the turnips it "costs") is to grow it on A land as long as any is available. The true cost or price of something is what you give up to get it.
Economists call this way of looking at cost "opportunity cost." This second and more subtle, but correct, method of measuring cost focuses upon how much of the final product (turnips or pot) you must give up to get the other. The line of reasoning that may seem more natural measures cost in terms of land given up in production. You might have thought thus: Since growing the turnips on A land costs one acre less than on B land, it must be cheaper to use A land. The reason why this thinking leads to the wrong answer is that it measures cost in terms of the units of the resource instead of the units of the final goods. After all, it is the final goods we consume, not the resource itself. Still, you may be puzzled as to why your intuition led you astray. Are not measurements of resource units important in allocation? After all it is the resources that are scarce, is it not? The answer is, of course, "yes" to both questions but this plausible line of reasoning omits the crucial fact that the units of the resource land are not identical, in particular, they differ in their capacities to produce the final goods. Thus, a bushel of turnips grown on A land had a different (higher) cost in terms of pot than an identical bushel grown on B land. Don't be discouraged if this new manner of reasoning and determining costs still seems strange and difficult at this point. Judging from what they said and did, many of history's most celebrated politicians and statesmen never understood it at all, but you are probably beyond that point already.
The Production Possibility Frontier

We will summarize the information on this production decision and move toward a solution to the problem of determining exactly what combination of pot and/or turnips we want to produce (there is an infinite number of possibilities) by introducing a new concept. This is the production-possibility frontier, drawn in figure 1. The production-possibility frontier defines the maximum amounts of each good that could be produced along with a given amount of the other good, given our resources and technology. "Technology" is simply our stock of knowledge about how to combine resources into final goods.

Each point on the frontier is technologically efficient, that is, it minimizes the cost of a given output. Take for instance our combination of 6 bushels of turnips and 54 bushels of pot indicated by point A in figure 1. There is a more costly way to produce the same output. You could produce 6 bushels of turnips and 54 bushels of pot even if you grew the turnips on type A land, provided that 11 acres of type B land were available. But that, of course, changes the problem. You should see that the effect of increasing the amount of land available would be to shift the production possibility frontier upward so that the maximum amount of pot attainable, given the requirement of 6 bushels of turnips would be 56 bushels. Given the present conditions, one could produce less than at the technologically efficient point, however, as we have seen--
Figure 1
Production Possibility Frontier for Turnips and Pot
a point such as B. Often in economics, as in the above example, minimization of cost and maximization of output are done simultaneously, or are really two different ways of looking at the same process. Each point on the production possibility frontier maximizes the outputs, given the inputs. But each point also represents the minimum input needed to produce a given output. Points within the frontier are inferior to ones upon it since there are many points on the frontier that represent more of both of the goods than points such as B without requiring more resources. The shaded area between A and C includes such points. Points A and C are also preferrable to B.

Note that the slope of the production possibility frontier changes after reaching the point where 40 bushels of pot and twenty bushels of turnips are produced. If you do not immediately see why this is as it should be, review the tables at the beginning of the previous section.

Each point on the production possibility frontier is the result of using the best (i.e. the least costly) methods of production. Each point is "technologically efficient" since there is no better way to secure the given output. Alternatively, there is no greater output of turnips (for specified quantities of pot) that can be produced with the available amounts of land and labor. We have not yet specified how an economy would choose which of these technologically efficient points is best in terms of its tastes for pot and turnips. This would require a kind of "social welfare" curve indicating
the order of preference among the many technologically efficient points along the production possibility frontier of figure 1.

Extensions of the Production Possibility Concept

The production-possibility frontier is a "menu" of choices, each of which is technologically efficient. A society in deciding what to produce simply chooses that combination lying along the frontier that it finds most satisfactory. If the entire curve shifted outward, the economy would be better off since for each choice on the initial menu, a better one is now attainable, i.e. one that offers more of at least one of the goods with at least the same amount of the other. Similarly, an inward shift of the entire curve unambiguously makes the society worse off. The production possibility frontier shifts because of changes in the amounts of resources available and changes in technology. When we speak of a "growing economy" we mean one whose P-P frontier is moving rightward.

To test your understanding of what the P-P-F means, figure 2 where in addition to the previously drawn production possibility curve of figure 1 labeled as I-I are drawn two new curves labeled M-M and N-N. These new curves represent shifts in production possibilities where the capacity to produce one good has increased at the same time the capacity to produce the other has decreased. The economy could now produce more turnips if it devoted all of its resources to turnip produc-
Figure 2
Shifts in the Production Possibility Frontier for Turnips and Pot
tion as shown by the points at which M and N intersect the X-axis. But the maximum output of pot now attainable is less than previously. Is the economy better or worse off? If point A was the optimum combination along the first curve, you should see that the new frontier M-M offers a range of choices between points B and C that are clearly preferrable and the economy is better off. Along N no such choice is available, but a new range of choices from D to E is now possible. Whether the economy is better or worse off depends upon whether or not any of these points is preferred to A. Finally, remember that the P-P-F depicts only the capacity to produce outputs. Two of the central problems in economics are to bring any economic system to this frontier and to consider measures to expand the frontier. Both are aspects of the "how" question.
Chapter 2. Comparative Advantage, Specialization, and Trade.

INTRODUCTION

In the study of economics, certain fundamental ideas recur to explain similar or parallel phenomena. Just as we demonstrated the economically rational manner of coping with scarcity in the previous case, where there were 2 goods and 2 kinds of land, we can use the same principles to analyze a similar problem of scarcity. Suppose again that there are 2 goods but instead of 2 kinds of land, suppose that there are 2 kinds of workers, differing from each other in their capacities to produce the goods. Since these 2 individuals are each consumers interested in their own welfare, as opposed to the "social" welfare we analyzed in the previous problem, this case is not exactly like the first. Despite the new twists, however, a very fundamental notion underlies both examples, the principle of alternative cost, or comparative advantage. This principle sheds light upon two pervasive economic phenomena, specialization and trade and is so general that it can be applied to examples involving 2 individuals or to entire countries, and to economic units in between.

Another Case of Turnips, Pot, and Scarcity

Suppose there are two individuals, Joe and Pete who have different capacities to produce, per unit of time spent at work, as shown in the table below. Joe is "better" at producing both turnips and pot than Pete and conventional reason-
ing probably suggests that he can produce both goods more cheaply than Pete. But, as in the case of the two grades of land, the cost of producing a good should be measured in terms of the alternative output foregone and this measure of alternative cost reveals a relationship parallel to the previous case.

<table>
<thead>
<tr>
<th>Production-Possibility Combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>PETE</strong></td>
</tr>
<tr>
<td>4 Bu. T, 0 Bu. P</td>
</tr>
<tr>
<td>3 Bu. T, 1/2 Bu. P</td>
</tr>
<tr>
<td>2 Bu. T, 1 Bu. P</td>
</tr>
<tr>
<td>1 Bu. T, 1 1/2 Bu. P</td>
</tr>
<tr>
<td>0 Bu. T, 2 Bu. P</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>JOE</strong></td>
</tr>
<tr>
<td>20 Bu. T, 0 Bu. P</td>
</tr>
<tr>
<td>15 Bu. T, 1 Bu. P</td>
</tr>
<tr>
<td>10 Bu. T, 2 Bu. P</td>
</tr>
<tr>
<td>5 Bu. T, 3 Bu. P</td>
</tr>
<tr>
<td>0 Bu. T, 4 Bu. P</td>
</tr>
</tbody>
</table>

While Pete sacrifices 2 bushels of turnips for each bushel of pot he produces, Joe must give up 5 bushels of turnips per bushel of pot obtained. Similarly, Pete's cost of turnips is 1/2 bushel of pot while Joe's cost is only 1/5 of a bushel of pot per bushel of turnips produced. Some general principles relevant to these examples should be noted explicitly. The first is that cost is appropriately reckoned in terms of output foregone rather than the absolute measures of cost one might be inclined to use. Second, in cases where factors of
production (inputs into the process of production such as land and labor) differ in their relative capacities to produce different outputs, it is mathematically impossible for one factor to produce each product more cheaply when cost is measured in alternative output foregone. In the simple, 2 factor-2 good cases we have considered you can see that the costs of producing one good for a factor is simply the reciprocal of the cost of producing the other good for that factor. If one factor produces one of the goods at a lower (higher) alternative cost, it produces the other at a higher (lower) alternative cost in comparison with the other factor of production. In the world, there are, of course, billions of individuals, resources, and final goods. Still, the underlying conditions are the same. Any factor of production must have this kind of cost advantage in producing some goods in comparison with some other factors. These alternative cost differences are described by the principle of comparative advantage.

The principle of comparative advantage stresses these differences in the relative or alternative costs of producing goods. Pete's cost of turnips is relatively greater than Joe's since the former must give up 1/2 a bushel of pot, as we have seen, whereas the latter sacrifices only 1/5 of a bushel for it. The customary manner of measuring costs in terms of units of inputs per unit of output, is a measurement
of absolute cost. While absolute costs are important, we have seen that they can easily mislead; cost reckoned in foregone alternative output is generally a better guide to optimal allocation of resources. In other words, relative cost, which reflects foregone or alternative output, is a better guide for output maximization (with given inputs) and/or cost minimization (with given outputs).

Analysis of the idea of comparative advantage leads us to the principle of specialization. Individuals or lands or any resources should be used to produce those goods that they produce best, i.e. at the lowest relative costs. We saw this for the case of land and can demonstrate it for our two individuals in their joint production-possibility frontier in figure 3. The upper curve shows the maximum amounts of each good that can be produced by Joe and Pete together, for given amounts of the other good. These maximum amounts are attainable only if there is appropriate specialization in production.

The end points of the P-P-F are simply defined as the maximum amounts of the goods attainable, when all resources are devoted exclusively to the production of one or the other good. For combinations of both goods, however, 2 different P-P-F's exist; the upper and clearly preferable one is the result of appropriate specialization, i.e. it results when the individuals specialize in the "right" product. The right product is the one for which the individual has the lower relative cost. The wrong pattern of specialization yields
Figure 3
Production Possibility Frontiers for Appropriate and Inappropriate Specialization
the inferior results in terms of total output available for consumption.

Beginning from the point of production of turnips exclusively (24 T, 0 P) consider 2 different ways of increasing the output of pot to 1 bushel. If Joe is to produce this unit of pot, turnip production will fall by 5 bushels; if Pete devotes enough effort to produce that first bushel of pot, total production of turnips will fall by only 2 bushels. The same relationships hold for the second bushel of pot. Clearly, if the pot and turnips are both to be produced, it will be better to allow Joe to produce all the pot unless the optimal output of pot and turnips is so great that more pot than Joe could produce is desired. If more than 2 bushels of pot are desired, Joe should produce only pot and Pete a combination of turnips and pot. The turning points in the production possibility curves indicate where production of the good is begun or ended by the second producer. As an exercise you should 1) compute the slopes of 2 curves, 2) estimate where P-P-F would lie if both producers produced some of both goods at all times. Note, finally, that the cost of inappropriate specialization becomes less as the distance between the two curves narrows. The explanation is intuitive; as the gain from specialization diminishes, the cost of inappropriate specialization also declines.
The Benefits From Specialization

Although appropriate specialization in production expands potential output, you may have wondered whether each individual benefits from specialization. Suppose that you prefer to consume (a) good(s) other than that for which you have a comparative advantage. Even if Pete is the champion pot producer (in the sense of relative cost), there is no guarantee that he would want to consume a large amount of pot or any pot at all, for that matter. Is specialization worthwhile only if there is a congruence of tastes and cost advantage? The answer is no, and the reason is the possibility of trade between Joe and Pete. Specialization coupled with trade will generally be preferrable for both producers to the condition of autarky (self-sufficient production and consumption, i.e. no trade). On his own, before trade, Pete finds that a bushel of turnips costs him 1/2 a bushel of pot. He would be better off to produce some pot and exchange it for a bushel of turnips with someone like Joe, if the offer price of pot is less than 1/2 a bushel of turnips. Since a bushel of turnips only costs Joe 1/5 of a bushel of pot, a trade would appear to be advantageous for both of them, i.e. there is a range of prices for turnips (and pot) for which both would benefit through specialization in production and trade. Exactly what the trade price will be depends upon several factors including their tastes and bargaining skills. For trade to be worthwhile the price of turnips must be somewhere between 1/5 and 1/2
bushels of pot and (equivalently) the price of a bushel of pot must lie somewhere between 2 and 5 bushels of turnips. The exact quantity of pot and turnips traded also depends upon their tastes and the "exchange rate" (the price) of turnips and pot.

The principle of comparative advantage which this simple example illustrates is one of the most fundamental in economics. Even if you supposed that the columns of the table of production-possibility combinations of this chapter referred to entire countries, instead of to individuals, the principle would remain intact. Differences in the relative costs of producing different commodities, i.e. the existence of comparative advantage, give rise to mutually beneficial trade possibilities—between individuals, countries, or intermediate units. Just as is the case for trade between individuals, one of the trading countries might have an absolute cost advantage or disadvantage in the production of both commodities. But again, the possibility of mutually advantageous trade depends solely upon the basis of differences in relative costs, not absolute costs. Both partners can benefit when each specializes in the commodities for which it has a comparative advantage (i.e. lower relative cost) and trades its own speciality for that of its partner.
Chapter 3. Demand

INTRODUCTION

In the last chapter, trade in goods was explained as an activity complementary to specialization in production. Trade and specialization are pervasive in the study of economic life, both past and present. Individual economic self-sufficiency is truly rare except where men are prevented by physical isolation from engaging in specialization and trade. In this chapter we begin to develop a fuller explanation of trade among individuals by examining the behavior of buyers of commodities.

Money and Markets

Our first observation is that individuals generally conduct trade not directly in goods, but in an intermediary medium of exchange called money. Money is found in virtually every economic system, even the fairly primitive, suggesting that it must be very useful. An economic system without this medium is a "barter economy." In a barter economy, where trade is conducted directly in goods, individuals must seek others with a "mutual coincidence of wants", i.e. someone who wants what one has produced and also has something acceptable in trade. The near universality of money as an institution suggests that a barter system must be an inferior means of conducting trade, that it is less burdensome to conduct trade indirectly through the medium of money than directly in the goods themselves. Seen in this sense, money is simply a con-
venience, a useful but only supplementary institution which serves to facilitate the basic economic activities of specialization and trade. There is much to this view of money, that it is "only a veil," a lubricant to keep the wheels of the economy smoothly turning. Later in our study we will see that the institution of money has other important consequences.

Most trades involving goods and money take place in markets, arenas where buyers and sellers interact. A market may be in a well-defined place, like a stock exchange, where buyers and sellers engage in continual contact and offers are easily ascertained. In other markets, contact between buyers and sellers is more irregular and information more difficult to obtain. These and other differences among markets will attract our interest but, fundamentally, all markets operate to perform one vital function, the determination of prices and this transcends any dissimilarities. To understand this process of price determination we will first consider the behavior of buyers, a topic traditionally termed demand. The purposes of this chapter are 1) to derive the "law of demand," and 2) to demonstrate its pervasiveness by examining some common phenomena.

The Demand Relationship

We begin by making a number of assumptions about the behavior of buyers. These assumptions simplify our problem which is to focus upon how buyers respond to different prices for a good. These assumptions can, and will, be relaxed to
make the analysis more general but initially they allow us to attack and understand parts of the problem one at a time. We assume that:

1) individuals make buying decisions for a specified time period, one that is sufficient to acquire "enough" information upon which to make their decisions

2) within this time period, individuals have a given amount of income available for making purchases

3) individuals' tastes and preferences are given and constant for the period and are "reasonable", therefore not necessarily representative of abnormal or pathological buying behavior

4) prices of other goods are given and constant for the time period

5) individuals seek to maximize the utility, i.e. to gain the most value or satisfaction from consumption that they can, given all these assumptions

6) individuals' tastes are such that substitution among goods without changing satisfaction is possible, that the same amounts of satisfaction can be achieved by consuming different combinations of goods.

Once we have given exact specifications to these six conditions, there is a definite quantity individual(s) would like to buy at any price for the good in question. Beginning from one price-quantity combination, we must consider how a change in the price, holding our assumptions constant,
affects the quantity desired. The set of all these points, pairs of prices and desired quantities, is the demand relationship.

**A Demand Experiment**

Suppose that a typical college student has an income of $100 a week which he spends entirely on combinations of two goods, joints of pot and fifths of scotch. This student clearly wants the "highest" possible level of satisfaction from consumption. Suppose, too, that initially when confronted with prices of $5 and $1, for fifths and joints respectively, he decides that the combination of ten fifths and fifty joints is the best he can buy. Now suppose that the price of joints falls to $0.50 per unit, perhaps the result of an improved harvest of pot (the effects of which work itself out in many intermediate markets between here and Acapulco). It is reasonable to expect the student to now consume more joints. This expectation arises because of two effects of the price change we can analyze separately. First, because he still has the same $100, the price decline enables him to buy more of all goods than previously. This increase in his "purchasing power" we will call the "income effect." Most people consume more of most of what they are already consuming when such a change arises. Secondly, the relative cost of consuming joints versus fifths has changed in favor of the former, since there is now a lower price of pot. Since his tastes for pot and scotch
are unchanged, this would induce an individual to increase purchases of pot, now instead of scotch. This latter effect, called the substitution effect and the aforementioned income effect arise from every price change. Together, they virtually ensure the fact that the quantity desired by buyers varies inversely with its price. You will find it instructive to follow the same reasoning to predict the effects of an increase in the price of pot upon the amount of pot desired.

Conceptually, we can separate the income and substitution effects in the above example by assuming that, at the same time the price of pot fell, our student’s purchasing power also fell because of a reduction in allowance to (say) $70. If the prices of pot and scotch were unchanged, he would clearly be worse off. The question is now whether his situation is better, worse, or the same given both the price and income changes. It is certain that our buyer will increase his consumption of pot relative to scotch because of the change in their relative prices and the law of substitution. Suppose that he now chooses the combination of 7 fifths of scotch and 70 joints as the best now attainable for him. Notice that he adjusts to the price change by substituting pot for scotch. Suppose further that our individual judges this combination to be exactly as satisfying, no more or less, than his initial choice of 10 fifths and 50 joints. Of course, we have arbitrarily assumed satisfaction is just equal to the first combination. The point is that there is some combination of a
price and income reduction which enable the individual to exactly maintain his level of satisfaction from consuming pot and scotch. Now we can identify the 20 joint increase as the substitution effect of the price change. If now we "returned" the $30 in income we would measure the income effect of the price change. The additional joints consumed (say, 30 more) is our estimate of the income effect, the amount by which consumption rises because the price change has made him "better off". Our estimate is crude and might have been made in slightly different ways. The important point that you should grasp is this distinction between the income and substitution effects which are, of course, inseparable in real world price changes.

The Demand Curve and the Law of Demand

By considering many alternative prices, we can specify the entire demand relationship. This series of pairs of prices and desired amounts can be depicted graphically in a demand curve as in figure 4. The demand curve shown implicitly holds at constant values all our assumed conditions. If you (or an inquisitive economist) attempt to establish actual demand relationship, you will find a relationship more complex and messier than the simple one shown. Instead of all observations lying along a straight line, or any one line, they probably would be arranged around the line running (perhaps) generally downward from the upper left corner to the lower right corner.
Figure 4
A Demand Curve

Figure 5
Observations of Quantities Demanded
as in figure 5 where each X marks an observed price and quantity. The difference between the two diagrams is that the latter draws upon "raw" real world observations where values we assumed constant in our six conditions are free to vary. Statistical techniques have been developed to refine such observations into the neater relationship of the first diagram. These two figures should demonstrate to you of the effect of our assumptions and in what sense the theory "simplifies" reality.

The demand relationship is sometimes stated as the "law of demand" which holds that the price and the quantity demanded vary inversely. You might be concerned that most individuals probably could not specify exactly how many units of goods they would consume at various prices. If individuals do not "know" their own demand curves, how can someone else attempt to predict their behavior? How can we be so sure of the "law"? Reliable predictions are still possible partly because we are usually interested in the "average" reactions of a number of buyers. Economists feel that buyers who follow "reasonable" rules to maximize their satisfaction will behave in a manner consistent with the law of demand. Of course, you should not accept such assertions without reservation and maintain an open and slightly skeptical attitude about your study of economics, and everything else.
Indirect Verification of the Law of Demand

The fact that demand curves are not directly observable leads us to some interesting verifications of the demand relationship. In the two examples that follow, the demand relationship can be inferred from observed behavior. The responses to changes in price are explicable and predictable, once you get the hang of it. Often, as in the examples that follow, alternative explanations consistent with the observations suggest themselves. It is especially easy to cite changes in tastes or attitudes as an explanation. Such a claim is usually difficult to refute since reliable measures of such subjective variables are rarely obtained. Economists are accustomed therefore to seeking explanations in terms of variables that can be measured—prices, income, output, etc. rather than the somewhat more speculative variables. This, of course, does not imply that measurable variables are always the only or even the most important factors in any economic problem. But an assertion that cannot be disproved, however interesting, seems less worthy of the specialized skills of researchers than ones that can be found to be powerful or wanting.

First, consider the relative proportions of Toyotas and Buicks among the automobiles registered in Columbus, Ohio and Honolulu, Hawaii. The proportion of Buicks to Toyotas is higher in Columbus than in Honolulu where the opposite is true.
Suppose for example, that in Columbus 15% and 10% of the cars registered are Buicks and Toyotas, respectively while in Honolulu the proportions are 30% Toyotas and 5% Buicks. (The argument that follows would be the same though slightly more complicated if the proportions of Buicks and Toyotas were [say] 10% and 15% in Columbus and 5% and 40% in Honolulu). Several explanations are possible. One might suppose that a difference in tastes reflecting more of an oriental culture would favor Toyotas in Hawaii as opposed to Columbus. Another possibility is that the costs of operating a car are so much greater in Hawaii as to favor smaller cars. Another suggestion is the influence of differences in the level of income between Columbus and Honolulu. Clearly, however, there is a strong case to be made for the influence of a relative price effect, viz. Toyotas are relatively (as contrasted to Buicks) more expensive in Columbus than in Honolulu. This relative price effect is consistent with the law of demand which holds that the desired quantity to be purchased varies inversely with the relative price of the commodity.

The price of these automobiles is determined mainly by two factors—the costs of manufacture and transportation. Toyotas' and Buicks' costs of manufacture are the same regardless of where they are sold but their respective costs of transportation differ greatly depending upon whether they are purchased in Columbus or Hawaii. Let us denote the higher transportation costs of Buicks sold in Hawaii as opposed to
Buicks sold in Columbus, with the symbol BT. Similarly since Toyotas sold in Hawaii, as opposed to Columbus are labeled Tt. The price of a Toyota in Columbus reflects its cost of manufacture plus TT, similar reasoning yields the price of Buick in Hawaii, and so on. Consider the following fractions:

in Columbus:

\[
\frac{PB_c}{PT_c} = \frac{\text{Cost of manufacture of Buick} + Bt}{\text{Cost of manufacture of Toyota} + Tt}
\]

in Honolulu

\[
\frac{PB_h}{PT_h} = \frac{\text{Cost of Manufacture of Buick} + BT}{\text{Cost of Manufacture of Toyota} + Tt}
\]

Since the numerator of (2) exceeds that of (1) while the denominator of (1) exceeds that of (2) the price of a Buick is higher, relative to a Toyota, in Honolulu while the reverse is true in Columbus. To make the effect of transportation upon relative prices, and thus upon demand, more clear, assume values for the variables as follows: cost of manufacture of Buick and Toyota are $4,100 and $1,700, respectively; transportation costs are BT=$650, Bt=$100, TT=$400, Tt=$200. Therefore,

\[
\frac{PB_c}{PT_c} = \frac{4,200}{2,100} = 2.0 \quad \frac{PB_h}{PT_h} = \frac{4,750}{1,900} = 2.5
\]

Buicks cost about twice as much as Toyotas in Columbus but two and a half times as much in Honolulu. So, when car buyers in Columbus and Honolulu compare Toyotas and Buicks,
one would expect the relative registration proportions to differ as we initially stated. This example is consistent with the law of demand but does not "prove" it. Furthermore, it does not prove that any other explanations are unimportant or even less important than relative prices. The example only seeks to depict what economists believe is an important and consistent relationship between price and quantity demanded.

As a second illustration of the law of demand, consider the fact that crime rates are generally higher in densely populated inner city areas than in suburbs. Numerous explanations probably occur to you, e.g. that standards of behavior are different at different socio-economic levels, etc. One explanation consistent with greater crime rates in urban "ghettos" is offered by the law of demand. Again, we do not claim that price alone explains the phenomenon, only that it is at least a part of the explanation.

The economic explanation of these different crime rates stresses that the cost or price of committing crime is higher in the suburbs than in the inner city. This cost reflects 1) the opportunity cost of conviction, and 2) the probability of getting caught. Both factors operate to reduce the cost of committing a crime in the inner city. If convicted, time spent in prison is more costly for the resident of the suburbs because his labor market alternatives are more attractive. The difference in wages or incomes of our hypothetical residents of the ghetto and suburbia can be taken as an approxi-
mation of their opportunity costs. Ghetto residents' wages are smaller because of less education, discrimination, and other factors which tend to reduce their labor market alternatives vis-a-vis suburban residents. In addition, the probability of conviction is lower in the ghetto because of numerous factors including the density of population, lack of open spaces, lower density of police protection. Whether or not a life of crime will "pay", depends partly upon the costs associated with it and those costs are smaller in the ghetto than in the suburb. Therefore, according to the law of downward sloping demand we would expect higher crime rates in the ghetto.

A Restatement of the Law of Demand

The demand relationship holds that the quantity of a commodity that buyers demand per time period varies inversely with its price relative to other goods, all other things equal. Other factors can alter the desired quantity, of course, but we wish to focus upon the relation between price and the quantity demanded. In principle, we can enumerate these other factors to be "held constant" (incomes, tastes, prices of other goods) as we did initially is specifying certain assumptions in the beginning of this chapter. In practice, it is sometimes a problem to "isolate out" the demand relationship from these other effects. Statistical and analytical techniques have been developed to allow us to focus upon this fundamental relationship between the price and quantity demanded, however.
The result is that economists feel more sure of the validity of the law of demand than any other concept in economic theory.

**Utility Maximization**

The law of demand leads us to a general explanation of buyer behavior called "utility maximization." We assume that households and other buying units attempt to gain the greatest attainable level of satisfaction or "utility" from the commodities they purchase. Since buyers are subject to some given amount of buying capacity, the goal of utility maximization is approached by allocating spending between different commodities on the basis of their prices and amounts of utility they provide. An increase or decrease in the price of some commodity alters the buyer's possibilities for utility maximization, other things (incomes, tastes, price of other goods) held constant.

An increase in the price of a commodity reduces the quantity desired by buyers since it alters the possibilities of utility maximization in a way which induces buyers to substitute other commodities for the now more expensive one. Similarly, a reduction in the price of a commodity suggests that utility can be increased by buying more of the now less expensive commodity. Thus, utility maximization is entirely consistent with the law of demand.

Naturally, some buyers are more sensitive than others to price changes and the degree of change in desired quantity varies according to many factors. These factors include the
tastes of buyers, the availability of substitutes, the initial prices of commodities, and the time period of analysis. Nevertheless, whether or not the sensitivity of the quantity desired to price changes is great or small, economists believe that such a relationship characterizes all buyer behavior. Thus, the market demand for all commodities bears an inverse relationship to their own prices.
Chapter 4. Cost and Supply: Profit Maximization and
Competition

INTRODUCTION

In the last chapter we began a theory of markets, the places where buyers and sellers interact, exchanging goods and services for money payments. Although money is a useful medium for conducting these transactions, economic activity, even in a society as complex as ours, should be recognized as ultimately consisting of trade in goods and services for other goods and services. The impersonal nature of most markets might lead us to describe the economic activity of individuals as "bartering (albeit with money) with the economy."

We will now consider the behavior of sellers in markets, the suppliers of goods and services. The concept of supply is in many ways parallel to the theory of demand. Supply and demand measure the willingness of sellers and buyers, respectively, to participate in markets at different prices. We assume that both buyers and sellers attempt to maximize their material well-being ("profits" for sellers, "utility" for consumers) and base our theories of demand and supply upon this assumption. Together, the interaction of these participants from the two sides of the market determines price and resource allocation.

The economic entities that supply goods and services are called businesses or firms. Each firm attempts to maximize its net income, or "profit" and thus most favorably increase
the wealth of its owner(s). The owners desire to increase their wealth because they want to increase their utility in consumption, and additional wealth increases their opportunities to consume. In a way we have come full circle, for everyone who receives an income is a "firm" in a sense, even if all he sells is his own labor services. Customarily, however, we think of firms as entities which "hire" resources and combine them to produce goods and services to sell in the market.

Before going on with our discussion of business firms, we should note that a useful way to consider household behavior is to think of households as "firms" that are engaged in the production of utility. Each household attempts to maximize its utility subject to several constraints: the prices of the commodities it buys and sells, the quantities of various resources it owns, etc. One of the key resources households possess is time since it enters into the family decisions regarding both its income and consumption levels. The family typically sells some of the time of its members in the form of labor services to the market but also "spends" part of its time in consumption (in conjunction with tangible commodities). Each family attempts to allocate the time and other resources it has at its command in such a way as to maximize its utility. Since families differ in their tastes and resource endowments, we observe a wide range of levels of consumption and income among families. It is useful for understanding
these differences to imagine that each family behaves as a firm whose goal is utility maximization.

Despite this similarity between firms and individuals, you may be concerned that there are some important conflicts between the interests of firms and consumers. In particular, if firms, as we usually use the term, are out to maximize the profit they derive from the sale of goods and services to families that seek to maximize their utility from the consumption of those goods, are not the gains of one group secured at the expense of the other? Quite simply, are the interests of buyers and sellers not fundamentally in opposition? Surprisingly, the apparent conflicts between buyers and sellers are resolved to produce a broad harmony of interests. Understanding how this process works will take you a long way toward understanding the entire operation of a market economy. Proponents and opponents of the market economy as an economic system probably base their positions largely upon their divergent views of how the apparent conflict between buyers and sellers in such an economic system is resolved.

Profit Maximization

Let us investigate just what profit maximization involves. A part, but only a part, of the firms' task in profit maximization is to assure the lowest possible cost of production for any particular quantity and quality of output—i.e., cost minimization. To minimize production costs, firms hire re-
sources and attempt to combine them in such a way that the
cost per unit of output is as low as possible. Firms are
generally not limited by their owners' wealth in securing
resources. Borrowing is a means by which firms may choose
a scale of production that does not depend on their owners'
level of spending power. The firm may vary the combination
of factors it hires in attempting to minimize price. In
changing the array of resources it employs to produce output,
the firm's flexibility is limited by any fixed commitments
("fixed cost") it may have made in the past extending into
the present and future. For example, firms usually have a
cost commitment in existing plant and equipment. The firm's
capacity to vary its combination of resources increases as
time goes by and old commitments expire. Ultimately, the
firm is limited in its capacity to vary its combination of
resources by costs arising from the sheer size of the enter­
prise. Even with modern technology, the costs associated
with managing an enterprise (information, co-ordination, etc.)
eventually begin to rise faster than output itself. We can
summarize the cost functions of virtually all firms (those
few exceptions form a special study in themselves) in a curve
called the marginal cost of output function, drawn in figure 6.

This curve shows the cost of producing one more unit of
output at various levels of output. In practice, the shape
of the marginal cost curve may vary somewhat between firms or
Figure 6
A Typical Marginal Cost Curve
in the same firm at different times, but the crucial characteristic is its slope. Eventually, the marginal cost curves of virtually all firms begin to rise. In figure 7 we show several more hypothetical marginal cost curves.

Firms eventually find that expanding output can only be accomplished by larger and larger additional amounts of resources per unit of additional output, i.e. at higher marginal costs. This proposition is called the law of diminishing returns, and is among the most pervasive of economic and technological phenomena. A simple illustration of the idea of diminishing returns is given by the "problem" of growing the world's food supply in a flower pot or perhaps 1 acre of land. If it were not for the law of diminishing returns, we could simply continue to add successive units of labor, equipment, fertilizer, etc., with no decline in the additional output obtained from successive additions of resource inputs added to the fixed amount of land. But, of course, you cannot accomplish this trick. Soon, you find that to increase output more additional inputs are required or per unit of additional output. Each additional unit of output would cost more in terms of additional resources. In fact, if you continue to add more resources to the fixed supply of land, eventually, a reduction in output (negative returns) might occur, since the resources might become redundant and "get in each other's way."

Firms attempt to avoid diminishing marginal returns by varying as many resources as possible when expanding output,
Figure 7:
Several Hypothetical Marginal Cost Curves
by increasing the scale while at the same time employing some ideal array or combination of resources. This is, as we have said, more feasible, the longer the time period over which adjustments can take place. But ultimately, the firm's capacity to produce is limited by the task of managing it. Of course, the firm can survive in the face of diminishing returns, because all the firm's competitors face the same basic problem. Conversely, the presence in the economy of a multitude of firms is perhaps the most persuasive casual evidence that diminishing returns is a force of crucial economic importance.

**Competition and Profit Maximization**

The notion of a multitude of competing firms is of the utmost importance in understanding how the apparently conflicting motives of firms and consumers are reconciled. This is because the outcome of profit maximization is dependent upon the nature of competition among firms for the purchases of buyers. To the extent that firms compete with each other, profits become more difficult to secure. One basic way that a firm competes is by producing at lower costs (for given quality of output) than its rivals. Firms compete for both resources (whose owners are also seeking to maximize their incomes) and customers; producing at lower costs gives the firm a competitive advantage on both fronts. In this competition, the race is to the swift, who tend to grow at the expense of the laggards who, at best, fall behind and, often disappear entirely.
The search for the cost-minimizing method(s) of production has two aspects. The first is pursuing the goal within the framework of existing knowledge and methods, what we have called "technology". The second aspect is the development of innovations or "breakthroughs" which set new standards for cost minimizers to attempt. You might think that this classification of two ways to cost minimization is a distinction without a difference; however, the second route, cost minimization through change is inherently more risky. It may be easier and safer to "follow the leader." On the other hand, following rather than attempting to lead may also be costly, so costly that firms who only follow fall so far behind as to drop out of the race. Consumers benefit from competition among the firms, because each firm, seeking to make the most favorable offer to buyers, spurs the others on to improve the offers they extend. This is the market system at its best, directed by "the invisible hand of competition" to give people more of what they want. To some extent, how competitively sellers behave and how much risk they are willing to bear depend upon incentives (e.g. patents, and tax treatment) and their attitudes. But how competitively sellers behave also depends on how attractive alternative kinds of behavior appear to them.

Do Firms Really Compete?

By this point, you may have become a little impatient and skeptical of our description of firms' competitive behavior.
Will it not be easier and more profitable for firms to agree not to compete too strenuously with each other? We have all heard stories about fabulous new products which sellers suppress because they believe that such innovations would eventually reduce their profits. For example, it is alleged that rubber companies could produce a tire that lasted much longer than present models, but they realize that sales would fall and they would be better off with existing conditions.

Whether the behavior of firms is more like the competitive model where each, in trying to "keep the customer satisfied", spurs others to do the same to the ultimate benefit of buyers, or whether firms' behavior is better described as a conspiracy against competition is not a question that can be answered theoretically (although many people make such assertions). Economists believe that firms' behavior falls on a continuum, where some are more competitive and some more collusive than others. Economists who favor the market system stress the difficulties of successful collusion which may force firms to behave competitively regardless of their desires not to do so. First of all, successful collusion requires the full participation of all sellers. But each seller individually has a strong incentive to break the compact, provided he believes that others would continue their non-competitive behavior. Each seller realizes this, so any such agreements may attempt to provide some kind of enforcement power. These powers are unenforceable at law in the U.S.
unlike some other countries—indeed they are generally considered evidence of criminal behavior—so formal arrangements to prevent competition tend to be unstable. Still, some would charge that informal restrictions on competition may last for relatively long periods. Clearly, some industries are dominated by a small number of firms, in fact, this is a common (although misleading) stereotype of American industries. Perhaps the greatest threat to collusion is the difficulty of keeping out new entrants attracted by what successful collusion can accomplish, viz. large profits. Economic history is replete with examples of new firms crossing old industry lines.

It is instructive to consider some kinds of behavior cited as "proof" of collusion among sellers in markets. An example already mentioned is the technologically possible dramatic innovation kept off the market. Assuming that some such innovation exists, reasons other than suppression of competition might explain its non-appearance on the market. The purchase price might be prohibitively high. Other costs such as the long commitments for very durable goods consumers must make might make them unattractive. Another frequently cited example of collusion is the similarity of price offers, infrequency of price changes, and general lack of price competition among sellers. While these allegedly common forms of behavior may be consistent with a degree of collusion, they can also be consistent with a degree of competition, once other costs are considered. In the final analysis, judgment on this issue is
difficult and only partly based upon objective evidence. There are reputable economists who view the American market economy entirely differently from each other as to the issue of competition and collusion. Since even advanced study cannot resolve this question, we can only hope to scratch the surface in our initial treatment.
INTRODUCTION

In this chapter we will continue with the theory of supply begun in Chapter 4. The supply of goods and services consists of the various quantities offered for sale at different prices. Supply is provided by firms who compete with each other and seek to maximize their profits. Profit is the difference between the firm's revenue from sales minus its cost of production—Profit = Total Revenue - Total Cost.

Part of the competition firms must meet is the competition for the resources used in the production of goods and services. Most resources are adaptable for many uses (kinds of production) and this fact, combined with the conditions of economic scarcity and competition among firms, forces firms to seek the most efficient (least costly) means of using resources. In the production possibility (P.P.) curve in figure 8 we depict the effects of these conditions, using as an example, the competition for resources faced by our friend the turnip firm, in relation to all other industries.

The Production-Possibility Curve and Marginal Cost

The P.P. curve depicts all the possible combinations of output of turnips, on the one hand, and goods that could be produced by all other firms, on the other. The vertical axis is labeled R.L.O.—"resources left over" for the production of all the goods and services of these other firms. The maxi-
Figure 8
The Production Possibility Frontier for Turnips and All Other Goods
mum output of this particular turnip firm is \( T_m \), shown on the horizontal axis. If this firm produced no turnips, \( R_m \) resources would be available for producing the output of other firms. To produce the amount of turnips, \( T_1 \), less resources are available for other firms to use. The value of the resources used by this turnip firm is measured by the distance \( R_m - R_1 \). Similarly, we can determine the total opportunity cost (foregone resources) of producing any amount of turnips up to \( T_m \). Of more interest, however, is a different kind of cost called the marginal cost. Marginal cost is the addition to total cost incurred in the production of one more unit of output. Suppose that \( T_2 \) is one more unit (say 1 million bushels) of turnips more than \( T_1 \). The marginal cost of the \( T_2 \)th unit of turnips is the distance between \( R_1 \) and \( R_2 \) and is measured by the slope of a chord between the points \( A \) and \( B \) on the P.P. curve. For very small distances along the X-axis (very small changes in turnip output), the slope of the P.P. curve is the marginal cost of turnips measured in $ worth of resources foregone. Note that this slope changes dramatically along the P.P. curve. At first, the MC of turnips is relatively low compared to the continuously increasing steepness which indicates the turnips are becoming more costly in terms of the alternative goods given up. In the previous chapter we discussed why the MC of turnip firms behaved this way citing the law of diminishing returns. The MC of output rises because some and, ultimately,
one resource remain fixed in supply to the firm. For the turnip industry as a whole, as opposed to individual turnip firms, the MC curve rises because all resources are not equally suited to the production of turnips. Starting with those resources best suited, the MC of turnips is initially lowest but it rises as other resources better suited to other uses, and thus more costly for turnip production, are employed.

There are two general shapes to MC curves noted in figure 9. The important characteristic shared by both is their rising slope which appears sooner and later, respectively. The latter curve implies that over some range of output the firm's marginal costs fall. For almost all firms this does not continue indefinitely as in figure 10. If such MC were common, we would expect to see most markets occupied solely by a single seller since a solitary seller's cost of production for any output would necessarily be less than that incurred by more than the single seller who would operate further "down" the MC curve. Since we usually do not observe such single firm domination, we infer that MC of most sellers eventually rises.

**Marginal Revenue**

To complete the development of our theory of supply we need only introduce one new concept, marginal revenue, which is the addition to sales receipts gained from the sale of one more unit. The combination of marginal revenue (MR) and marginal cost in an analytical framework called "marginal ana-
Figure 9
The Two General Shapes of Marginal Cost Curves

Figure 10
Continuously Diminishing Marginal Cost
lysis" enables us to predict the quantities sellers would offer at various prices, i.e. the supply curve. The theory is built upon our assumption that sellers seek to maximize their profits within a competitive environment. Although the theory is developed for predicting the behavior of competitive firms you will subsequently see that, with slight modification, it is applicable to all kinds of markets, competitive and otherwise.

Marginal revenue is the amount by which a firm's total revenue (dollar sales) change when one more unit is sold. If you think that the MR is simply the price at which the last unit is sold, a natural first reaction, see the 2 tables below. They depict parts of the demand schedules for 2 different kind of firms.

<table>
<thead>
<tr>
<th>Q</th>
<th>P</th>
<th>TR</th>
<th>MR</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>10</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>48</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>49</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q</th>
<th>P</th>
<th>TR</th>
<th>MR</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>10</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>70</td>
<td>10</td>
</tr>
</tbody>
</table>

(1) (2)

The firms in the first table is faced with a downward sloping demand for its output, the price and quantity demanded vary inversely. To sell additional units, the price must be lowered, not only for the last unit, but for all previous units. The MR associated with any price is always less than
the price for such a firm. The MR lies below the firm's demand curve since the addition to total revenue is always less than the price. The effect upon total revenue is a result of 2 opposing forces: an increase of \( P \) for the last unit sold and a decrease due to the reduction in price of the previous units required to sell one more unit. MR could become negative. You will note that the MR of the 4th unit is not given because of incomplete information. To determine the MR of the 4th unit, the total revenue of 3 units must be known.

The second table depicts the conditions of demand for a competitive seller. A competitive seller is defined as one who can sell as much or as little as he wants at the market price without having any effect upon the price. This definition of competition may strike you as a bit strange and, indeed, the notion of competition conveys something to economists that is different than its every day connotation. The demand curve for the competitive seller is shown in figure 11 which corresponds to the table 1. The figure 12 corresponds to table 2 and depicts a non-competitive seller.

"Competition" suggests personal rivalry between sellers to most people. Economists think of competitive markets as ones in which there are many sellers, each insignificant relative to the market as a whole. Each seller's relative anonymity makes such competition impersonal. This gives rise to the horizontal demand curve in figure 12 where the firm's MR is equal to its price, since no relatively small firm needs
Figure 11
Demand and Marginal Revenue Curves for the Competitive Firm

Figure 12
Demand and Marginal Revenue Curves for the Non-Competitive Seller
to reduce its price to increase its sales. Although there are few markets for which these conditions are satisfied perfectly, the theory turns out to be a useful basis for predicting the supply offered by many actual firms. Note that while the demand curve for the firm is horizontal, this does not imply that the industry demand for the good is not inversely related to its price. The demand for the good (the demand for the output of all firms taken together) is consistent with the law of demand. A competitive firm is so insignificant relative to the entire market that its own demand curve is horizontal. What is true for individual firms is not true of them all, taken together.

**The Profit-Maximizing Output**

In figure 13, the MC and MR curves of one competitive firm are plotted against quantity. Indicated on the horizontal axis are three points $Q_1$, $Q_2$, $Q_3$ where, respectively, the MR exceeds, is equal to, and is less than MC. Our task is to show at which of these points the firm will find the best output to produce and sell, if it decides to sell any at all.

Since the competitor has no control over price, his decision boils down to choosing the best output, given the market price, or producing none at all. The price is given, he can "take it or leave it." Consider a point like $Q_1$. Here the last unit produced added more to the firm's revenue than to its cost, i.e. $MR > MC$. Producing this unit would make the
Figure 13
Marginal Cost and Marginal Revenue Intersection
firm better off than producing one less unit since it would have a positive effect upon profit \((TR - TC)\) since it adds more to TR than to TC. Remember, however, we still are not sure that the firm would prefer to produce at all. But what is true of \(Q_1\) is also true of all points up to \(Q_2\). Since each of these units adds more to revenue than to cost, the firm would be successively better off as it proceeded from a point like \(Q_1\) to \(Q_2\). Now consider a point like \(Q_3\) where the MC exceeds the MR. Each such unit produced would reduce the firm's profits. The firm would be better off to reduce its output to \(Q_2\) since not producing each unit down to \(Q_2\) reduced cost more than revenue. Therefore, the firm's best output to produce, if it produces at all, is the point where \(MR = MC\), i.e. this is the output that represents its maximum profit from production. It follows that the amount the firm would offer for sale at any price is shown by the point on its MC associated with the price. That is, the MC curve is the supply curve of a competitive firm, provided the firm finds it profitable to produce some output. To convince yourself that this is so, imagine the horizontal demand curve rising for the competitive firm. Its best output will always be given by the intersection of the MR and MC curves. But each such intersection point (i.e. the points along the supply curve) is given by the MC curve.
Price Determination

The industry supply curve, our chief concern, is simply the horizontal summation of all the firms' supply curves. One simply adds the quantities offered for sale by all the firms at the various prices to derive it. The industry and firm supply and demand curves are in figure 14. The quantity scale on the graph for the industry is much larger than for the firm. You can imagine that the supply and demand conditions for the competitive firm have been "blown up" to facilitate our analysis.

The price to which the market tends is \( P_E \). At \( P_E \) both buyers and sellers of the good are satisfied, i.e. both choose the same quantity \( Q_E \) as the one they would like to buy and sell respectively. Only at \( P_E \) is there this agreement. Of course, buyers would be even happier with a lower price and sellers would prefer a higher one. Still \( P_E \) is the only price that satisfies both in the sense that the quantity desired by each is equal. Such a point \( P_E \) is the equilibrium price and \( Q_E \) is the equilibrium quantity given the supply and demand curves \( S_1 \) and \( D_1 \).

If this market is allowed to determine \( P \) "by itself" it will tend toward \( P_E \). Consider a price such as \( P_A \) above the equilibrium. At \( P_A \) the quantity buyers choose to buy is less than the quantity sellers would like to sell. The difference between the two is measured by the distance between the curves at that price and might colloquially be called a "surplus".
Figure 14
Demand and Supply Conditions for the Industry and the Firm
or "glut on the market." If the price is free to vary, this situation is unlikely to persist. Rather, some of the sellers unsatisfied at $P_A$, i.e. unable to sell all they want, find it preferable to offer lower prices at which they can sell part of their "surplus." This will tend to reduce the price. This tendency will continue until such time that no frustrated sellers are left, viz. until price equals $P_E$. An interesting and important question we have yet to consider is how long this process will take. We will say only that the time required varies between markets but that eventually the market will approach $P_E$.

The equilibrium price is like a magnet, the market is attracted toward it whether it is above or below it at some moment. You will find it instructive to consider how the market reacts when the price is at a point like $P_B$ where the quantity demanded exceeds the quantity supplied. Such a situation is commonly described as a "shortage." One way it might arise is through legislation preventing the price charged from exceeding $P_B$. After all, everyone prefers a lower price except producers who are likely to be far outnumbered by consumers at the ballot box. Can the laws of supply and demand be repealed by formal legislation? It has been attempted countless times. Formal legislation may alter the observed price, but the excess demand at $P_B$ has a number of ways of manifesting itself. Unsatisfied buyers generally express
their frustration and this force may give rise to phenomena such as "black markets," discrimination between the "excess number of buyers," "under the table arrangements," etc. To supplement price controls, governments sometimes institute rationing and deal out punishment to price control offenders. The point is that the excess demand expresses itself in a variety of ways. The quantity desired by buyers exceeds the quantity sellers offer for sale at $P_B$. Usually a price change (to $P_E$) would reconcile this discrepancy. If this kind of adjustment is precluded, some other must take its place.

**Shifts in Supply and Demand**

Consider the effect of a new technological development, say an improved hybrid turnip, which shifts the production possibility curve as in figure 15.

At any output the MC of turnips has been reduced which shifts the supply of turnips rightward. This effect, in turn, creates a "surplus" at the old equilibrium price. A new equilibrium determined by the intersection of the new supply curve and the old demand curve (assumed to be unchanged) will be the eventual outcome. Thus, the benefits of lower costs are passed along to consumers, even though the sellers are seeking to maximize their profits, because of competition. As an exercise you should consider the effect upon the industry equilibrium price and quantity of an increase in demand for turnips. Such an increase in demand could arise from the three basic demand factors we discussed in Chapter 3, viz. changes in tastes, incomes, or prices of other goods.
Figure 15
A Shift in Supply Caused by Technological Change
Chapter 6. Applications of Supply and Demand: Shortages

INTRODUCTION

In the last several chapters we have introduced the concepts of supply and demand to explain the behavior of sellers and buyers participating in markets. It is no exaggeration, but it is a deceptive oversimplification, to state that every question in economics boils down to a matter of supply and demand. We will now begin to apply our supply-demand apparatus in an attempt to explain how an economic system answers its perennial questions of "what," "how," and "for whom." Regardless of how far one proceeds in economic analysis, this quest and the conceptual framework remain the same. Moreover, despite the fact that our discussion is couched within the institutional framework of a "free enterprise" economic system, most economists would claim the theory is broadly applicable for "planned" economies. Beneath the welter of institutional differences, any economic system must "solve" the same fundamental problems. Scarcity and its implications are not avoided by the choice of the "right" economic system. Of course, proponents of one system or another generally feel that their systems come up with "better" answers to the questions of what, how, and for whom.

In this chapter we shall investigate the operations of supply and demand in a somewhat indirect fashion. In the main, we will consider examples where the price mechanism is inhibited by attempts to prevent its customary operation. A consideration
of how supply and demand forces manifest themselves, even under these special circumstances, will give an appreciation of their essential character.

The supply and demand curves show the quantities, per unit of time, that sellers and buyers would be willing to offer and purchase, respectively, at various prices. Thus in figure 16, at $P_A$ the quantities that are desired by buyers and sellers are $Q_1$ and $Q_2$ and similarly at $P_B$. (The fact that $Q_1$ and $Q_2$ represent supply and demand at both prices should not be taken as a general condition; rather this was deliberately chosen for convenience and ease of exposition.) At both $P_A$ and $P_B$ there is a divergence between the quantities supplied and demanded measured by the distance $Q_2$ minus $Q_1$. At $P_A$, the quantity supplied exceeds the quantity demanded; a condition familiarly termed a "surplus." At $P_B$, the quantity demanded exceeds the quantity supplied, commonly described as a "shortage." More precise descriptions of the conditions of supply and demand at $P_A$ and $P_B$ are "excess supply" and "excess demand," respectively. An excess supply or demand prevails at all prices except $P_E$, given these particular conditions of supply and demand. Our theory predicts that market forces tend to push the price toward equilibrium positions such as $P_E$, or to put it another way, that conditions of excess supply or excess demand should be only temporary if market forces are allowed to operate.
Figure 16
Supply, Demand, and Three Alternative Prices
**Buffer Stocks**

First, we will refine our theory to take account of a process we have not yet considered. In figure 17 $D_1$ and $S$ describe the initial conditions of demand and supply. Suppose that the equilibrium price of $P_E$ prevails. Now, a rather naive interpretation of our model holds that if demand even temporarily shifts to $D_2$ or $D_3$ that we should expect a prompt change in the price to the new equilibrium values of $P_A$ and $P_B$, respectively. Otherwise, this view holds, a shortage would exist in the case of an increase in demand to $D_2$ (the distance $Q_A$ minus $Q_E$) or a surplus would be created by a decrease in demand to $D_3$ (measured by the distance $Q_E$ minus $Q_3$).

As a matter of fact, most markets do not operate in the very simple manner just described, viz. responding immediately to every shift in demand with a new equilibrium price. Consider, for example, the retail market for beer. The demand for beer can be expected to fluctuate from week to week even if it remains stable over a number of weeks. A beer retailer might find that he sold an average of (say) 1,000 cases of beer per week over a year's time and that the average was about the same for each month as in the preceding year. From week to week, however, the demand might fluctuate between 200 cases and 3,000 cases depending on factors such as the weather, televised sports, holidays, student vacations, etc. The effects of these factors are roughly predictable, so retailers of beer
Figure 17
The Effects of Temporary Shifts in Demand Upon Prices and Quantities
and other goods can and do vary their orders to accommodate expected fluctuations in demand. However, since the variations in sales cannot be predicted perfectly and customers appreciate a high degree of certainty that beer will be available if they take the trouble to go to the retail store, its genial proprietor will generally find it profitable to carry an inventory of beer to protect himself and his customers against fluctuations in demand and to increase the probability that customers will not be disappointed and consider more reliable sources of supply. Exactly how large an inventory he should carry depends upon a number of costs such as storage, delivery, etc., not to mention the fact that the beer is perishable (in the sense that its quality may change with the time spent in inventory before being sold). The point is that most sellers will hold inventories rather than let the price rise and fall with every apparent change in demand or let shortages occur whenever unexpected increases in demand are experienced. Beer is available to customers at more or less customary prices, rather than prices that fluctuate frequently. Customers appreciate this degree of certainty about price as well as availability since such stability makes planning and efficient shopping less difficult and costly for them. Of course, prices do occasionally change, sometimes as a result of shifts in demand. Temporary changes in demand for many goods will not change their prices, however, rather the impact is felt in inventory levels that function something like shock absorber to "smooth out"
bumps." The inventories are buffer stocks against demand fluctuations.

An interesting example of how inventories operate and the costs incurred in maintaining inventories is given by the food retailing co-operatives that have sprung up in many college communities. These co-ops claim to offer lower prices than conventional retailers. One means of offering lower prices is to reduce costs. A co-op might require, for example, that all individual customer orders be placed and prepaid before the co-op itself makes the purchases on behalf of all customers in the aggregate. In this case, the costs of storage and maintaining stocks of goods available for sale are sharply reduced. If net costs of good are reduced and prices reduced accordingly, it is because certain costly services are no longer being performed for buyers, not simply because grasping "middlemen" have been "avoided."

The Energy Crisis

To see a relationship between inventories and shortages consider the "energy crisis" that was so widely publicized in the 1972-1973 winter season. We constantly heard that supplies of various fuels were running low and that draconian measures and/or sharp price increases would be required to accommodate the excess demand. Can economic theory lend some insight into this crisis?

Sellers of fuels, like most suppliers, maintain inventories (e.g. of natural gas) or excess capacity (e.g. to ge-
nerate electricity) against fluctuations in demand. Normally, these buffers would be sufficient to provide assurance that fuels would be available to meet unusually high energy demand. Why, in the rather mild winter of 1973, should supplies have run "dangerously" low?

A large part of the answer seems to have been the effects of a Federal policy limiting the price of natural gas. This maximum price was established several years ago presumably with the "good intention" of keeping natural gas prices lower than they would be otherwise. However, during the winter in question, supply and demand conditions similar to those of figure 17 where \( P_B \) is the legal price. At this price the quantity demanded exceeds the quantity supplied and inventories are being depleted. The price \( P_B \) is such that suppliers of natural gas do not have a strong enough incentive to provide as much as is currently being consumed. At a higher price the quantity demanded would be less and the quantity supplied somewhat greater, and inventories would not be so "dangerously low" and falling. At a sufficiently high price, inventories would not be any lower than usual, because current consumption would be more or less offset by current production. The "crisis" is simply a situation of excess demand made chronic by price legislation preventing the market mechanism from operating.

The energy crisis is complicated by several additional factors. First is the fact that the demand for natural gas has increased greatly in recent years, probably more than
producers expected, because people have come to appreciate more its clean-burning characteristics. With the concern over pollution, "tastes" have changed in a way to increase the demand for natural gas. Second, the supply of natural gas beyond the current inventories can be expanded only through new discoveries. There is no doubt that a great deal of natural gas lies untapped, but supply will be expanded from this source only if sufficient incentives are given to producers to engage in exploration. Even then, augmenting the supply of natural gas takes more time than many commodities. In addition, one cannot rule out the possibility that domestic producers of natural gas may have intentionally restricted output to aggravate the shortage of natural gas in an attempt to bring the issue to public attention and increase pressure to remove the maximum price on natural gas. Still, there is strong support for the idea that specific price controls on natural gas production, coupled with the controls of "phase I" and "phase II" anti-inflation program of 1971 and 1972, have been the principal causes of the gas shortage. This suggests that the solution to the problem lies in restoring the ability of the price mechanism to equate the quantity of gas demanded and supplied. Price increases are required to assume the appropriate reallocation of resources in this area. Continuation of gas price regulation as in the past policies, will only perpetuate and probably aggravate the shortage.
Other Attempts at Price Control

Government attempts to regulate prices are an old story in economic history and offer some useful insights into how the price mechanism operates. Often the attempt is made to control all prices at once rather than only individual items like natural gas. No one likes higher prices, except sellers, of course, and they only want higher prices for the things they sell, so such attempts to control prices are usually politically popular. Economists generally agree that inflation (an increase in the average level of all prices) is undesirable although for reasons that are distinctly different than those commonly held by the man on the street. Moreover, despite the agreement regarding the undesirability of inflation, economists tend to be skeptical of methods of direct price control typically favored by politicians and voters. This skepticism of the effectiveness and even the desirability of controls is based upon the economists explanation of the causes of changes in the price level.

We have seen that an individual price tends to rise when the demand for the good increases relative to the supply. Roughly the same process operates to cause the price level (an average of all prices) to rise and fall. Inflation is caused by an increase in the demand for all goods (on the average) relative to their supply. Laws which attempt to control prices by simply stating that they should maintain some particular value (usually the same price as previously)
may be politically expedient but fail to treat the cause of the problem. Rising prices are caused by rising demands and are only reflections of the underlying conditions causing demand to increase. Just as one would not consider the breaking of a thermometer an effective means of reducing the heat in a room, economists generally do not consider price controls a fundamental cure for inflation. Regardless of how forcefully controls are enforced, they cannot alter the existence of excess demands for goods although they can change the ways in which this pressure manifests itself. For example, sellers may attempt to alter the terms of a transaction by reducing quality or services offered instead of raising price directly. Because buyers are willing to pay higher prices (openly or disguised), i.e. there is excess demand at the given prices, sellers have the incentive and capacity to change them. Furthermore, price controls are generally applied uniformly regardless of the conditions of excess demand or supply for particular goods. Shortages for those goods with the greatest excess demands become most acute and usually lead to abandonment or at least relaxation of controls in those areas, to the impositioning of rationing, and/or to the development of black markets.

Another example of interference with the price mechanism is given by governments' attempts to fix the price of borrowed funds, "the" interest rate. (Actually, there are many different interest rates for different kinds of loans and borrowers.
It makes some sense to speak of "the" interest rate because the various rates tend to move together.

At any given time, some individuals find that their current funds available for purchase of goods and services is different than the amount they would need to purchase those commodities. Individuals whose "need" is greater than their current means, are generally willing to sacrifice something to obtain more funds. Others whose "need" is less, might be willing to lend those in the first category the temporary use of their own purchasing power under the right conditions. These conditions are summarized in the interest rate which is a price paid for the use of borrowed funds. Lenders usually require compensation, not only for delaying their own use of the funds, but also for risks and costs they incur in making loans—the risk of default, changes in the price level, etc. We can discuss this problem in terms of the supply and demand functions for borrowed funds (actually there are many markets for particular kinds of loans). These demand and supply schedules relate the price and quantity of borrowed funds in the same way other supply and demand curves relate prices and quantities.

Historically, many people have felt, despite the willingness of borrowers to pay interest, that for lenders to receive interest on loans was unjust and unfair. This feeling may have reflected concern that in a slowly changing economy, the demand for loans often represented unusual circumstances of dire need where, lenders enjoyed a very strong bargaining
position vis-a-vis borrowers. Due to such concerns (often expressed as religious dogma) "usury" laws, regulating interest rates, have been common. Regulation has attempted to establish maximum allowable rates, presumably to protect borrowers against unscrupulously grasping lenders. Usury was once considered to be the charging of any interest; today it is usually thought of as charging an excessive rate.

Such laws, no matter how well intentioned, create the same difficulty that all price controls, general and specific, generate, namely, excess demand. The willingness for borrowers to pay interest indicates that borrowed funds must be providing some useful service. The price serves to allocate loanable funds among competing users and uses. If lenders find that they cannot charge a rate above X%, but that there is an excess demand for loans at that rate, they are likely to use other means of allocating these funds among competing users. These methods including raising the price indirectly as in "real estate points", refusing to lend to high risk borrowers, and discriminating among borrowers on basis of personal characteristics such as age or race. Thus, paradoxically, the potential borrowers with greatest desire to find loans are likely to find it most difficult to borrow at the legal maximum rate; they may turn to extra-legal lenders such as the neighborhood "loan shark." In such circumstances it seems clear that usury laws are a major cause of the sur-
vival of loan sharks— who are often associated with organized crime— which seems an unlikely intention of the law.

We have seen that a good deal of insight into the price mechanism is gained by observing the consequences of attempting to interfere with its operation. A second lesson lies in some of the unforeseen consequences, often unfortunate, which seem only to confirm the old adage about where the path of good intentions sometimes leads.
Chapter 7. "Some Simple Ways Not to Cure the Farm Problem"

INTRODUCTION

We have seen the effects of several attempts to interfere with the price mechanism. The market forces of supply and demand have their ways of upsetting the best intentioned efforts of governments to "improve" upon their performance. In this and the following chapters we will consider several of these attempts, beginning with a consideration of the "farm problem." In each of these cases you should ask the following questions:

1) What is the effect of the intervention upon the allocation of resources?

2) Who really benefits and who suffers from these programs? Are these parties the same ones as the proponents of the programs think (or say) they are?

A Statement of the Problem

Attempts to improve upon market performance in agriculture are ostensibly motivated by a dissatisfaction with some aspects of the income distribution of producers. The farm problem is not that too little is produced (ironically, it is almost the opposite) or that the price is too high (throughout the 20th century the price of farm goods has generally fallen relative to other prices). Rather, the problem is that many farmers earn incomes well below the average of the nation as a whole. Incomes of many farmers are so low that they have sought other jobs and occupations with the result that the number of farms,
especially the traditional family farm, has dwindled. Many individuals, especially farmers, feel that the family farm is an institution that should be preserved. Various "solutions" to the farm problem have been proposed and put into practice but they generally attempt to deal with these two interrelated concerns: the low income of many farmers and the threatened extinction of the family farm.

**Why Many Farmers Earn Only Low Incomes**

In the competition among producers in a market economy, the highest incomes tend to go to those producers who best satisfy buyers' demands. For some particular commodity desired by consumers, "best" means "most efficiently" or most cheaply for a given quality. For most agricultural goods, whose qualities are fairly standardized, those producers who can produce the most output for a given amount of resources will earn the highest incomes. Others, whose production per unit of resource is less, naturally tend to have lower incomes. Unfortunately for those concerned, family farms are generally less efficient producers of most agricultural goods with the result that their incomes tend to be low relative to other farm producers and to other occupations, in general. Family farmers simply tend to be less efficient in converting resources into commodities than most other producers. The competitive performance of family farms has worsened in recent decades. One way to depict this situation is to categorize farmers into three groups on the basis on their efficiency
in production. Let us establish three degrees of efficiency, A, B, C, where group A includes the most efficient and group B the next most efficient farmers. Family farmers are highly overrepresented among group C farmers.

Exactly why family farms are generally least efficient is largely beyond our story although some of the reasons are related to their small size, relative to other farm producers. The direction of causation is not that simple, however, since the better farmers tend to have larger enterprises because they are successful and less efficient farmers tend to remain small or even "shrink." Large size is not an unmixed blessing in all forms of agriculture or other forms of production. In many lines, relatively small firms succeed handsomely and compete successfully with larger rivals.

Figure 18 reveals the effects of marked differences in efficiency between our three groups of farm producers. In the figure 18 we have graphed representative marginal cost curves for each type. Group A producers are the least cost producers which is simply another way of saying that they are the most efficient. Groups B and C follow in terms of cost of production. You can see that a given quantity like \( Q_1 \) could be produced only at successively higher costs by groups A, B and C.

Another way of presenting the same information is to construct an industry supply function, specifying what portions of the supply come from groups A, B, and C. We can derive the
Figure 18
Marginal Cost Curves for Three Levels of Production Efficiency
industry supply curve from the graph of the MC functions as in figure 19. For example, at prices below $MC_0$, no output is forthcoming since even the most efficient producers cannot meet their costs of production at such low prices. Between prices equal to $MC_0$ and $MC_1$, the entire output is supplied by group A producers since the price is still too low for groups B or C to profitably operate. At prices equal to $MC_2$ and $MC_3$, respectively, producers from groups B and C will enter the market. The industry supply will represent the contributions of all three groups at sufficiently high prices. Conversely, at lower prices the higher cost producers cannot participate, i.e. compete, because they are too inefficient. Of course, nothing prevents the least efficient producers from trying to compete with low cost producers. The result is virtually foreordained, however; Unless price is favorably high, the least efficient farmers must struggle to eke out a very meagre income. Things would be all right if only the price was sufficiently high. And if wishes were horses, beggars would ride. But at this point, governments have often seen fit to come to the rescue of the small farmer.

**Proposed Solutions to the Farm Problem**

Almost all proposed solutions to the farm problem begin with the recognition that even the least efficient farmers can survive and perhaps enjoy a modest prosperity if the price is sufficiently high. Of course, the issue is seldom stated
Figure 19
The Industry Supply Curve, Apportioned by Type of Producer
in such cold terms, rather the intent is to promote "fair" prices. Fair prices are understood to be prices high enough to accomplish the same purpose.

In figure 19 we can analyze the effects of policies that raise the price of agricultural output to producers. The supply curve in the figure is derived from the MC curves of figure 18. Note that the supply curve is divided into three parts according to the group(s) contributing to production. For example, the supply curve up to $Q_1$ is supplied entirely by group A producers since price $P_1$ corresponds to $MC_1$ in the previous graph. Between $Q_1$ and $Q_2$ output is supplied by producers of A and B since the price ranges between $P_1$ and $P_2$ which correspond to $MC_1$ and $MC_2$. Only above $P_2$ do group C producers get in the act.

To better appreciate what the solution must accomplish, let us first consider the situation before intervention. Suppose that the market demand curve was equal to $D^*$. The market price and quantity would be $P^*$ and $Q^*$ respectively. We can roughly calculate the profits of producers at $P^*$ by first noting that the total revenue is given by the rectangle bounded by the axes and perpendicular lines originating from $P^*$ and $Q^*$ as indicated on the figure 19. We state (without proof at this point) that the relevant cost of this output is given by the area under the supply curve up to $Q^*$. Producers' profits are equal to the remaining (upper) part of
the rectangle, i.e. the total revenue minus the cost. These profits are earned mainly by group A producers since group B producers are producing little and group C not at all at this point.

A sympathetic legislator might consider that if price were raised to \( P^{**} \) or higher, group C producers could enter the market profitably. We can see this by observing the difference between the portion of this new total revenue rectangle lying above the supply curve in the area beyond \( Q_2 \) as indicated by the darkened triangle. Assuming that all the output beyond \( Q_2 \) was supplied by C producers, this area would measure their potential profit from producing which is dependent upon measures which increase the market price beyond \( P_2 \).

However, type C producers are not the only beneficiaries of such a policy. As a matter of fact, they are not even the major gainers. Consider the effect of raising price to \( P^{**} \) upon the profits of type A and B producers. Their profit is again the area between the TR rectangle and the area under the supply curve. Since the costs for type A and B farmers remain the same, you can see that these profits have increased considerably more than the profits of group C. Further increases in the price of output have the same effect. But, of course, family farmers are spared the task of finding a new job and the family farm is protected from "extinction."

All that is left to decide is exactly how the price is to be raised to its "fair" level. One method is to simply
establish a minimum price at F**, a legal price above the market equilibrium price. All farmers benefit as described previously although the benefits accruing to the intended group C farmers are much less than to their less needy, more efficient competitors. Passing a law, however, does not repeal market forces, and quickly another problem arises. At P** the quantity of produced far exceeds the quantity demanded given by the demand curve, D*. If something is not done to combat this surplus, i.e. the excess supply at P** (measured by the arrow in figure 19, farmers unable to sell much of their production at P** will offer it at lower prices and the same problem we began with will reappear. The government can easily dispose of this problem by making sure that the demand "really" is great enough. The simplest way would be for the government to assure farmers that it would buy any unsold production, the "surplus," at the end of the season. This straightforward plan has the disadvantage of candor, since it openly reveals that the government is willing to buy what consumers consider to be overpriced output. It looks too much like the government is doing farmers a special favor. To avoid this embarrassment, the policy has often been stated that the government will "lend" farmers an amount equal to their surplus production valued at P** per unit. The government then takes delivery of the output as "security" for the loan. It is understood that if the market price fails to reach P** on its own, farmers will simply "default" on their loans, i.e.
the loan will never be repaid and the government will be left holding the bag; albeit filled with grain.

As you can imagine, the government might soon accumulate an embarrassing stock pile of surplus commodities under either of the above programs. Again, the appearance of favoritism would become obvious. To avert this situation, solutions are sought to effectively raise the price to \( P^{**} \) while somehow averting the surplus. One way to accomplish this would be to promise producers a price of \( P^{**} \) by giving producers the difference between \( P^{**} \) and the market price they would in fact receive if they sold \( Q^{**} \) in the market place. This subsidy measured by the vertical arrow at \( Q^{**} \) would be paid producers for each unit they sold. While this effectively raises the price and eliminates the surplus, since the entire \( Q^{**} \) is sold, the "handout" aspect of this program is more obvious than ever. An ingenious solution evolved at this point, however, which appeared to accomplish the objective of fair prices without accumulating surpluses or an open handout. Producers were offered the alternative of receiving a payment from the government for not producing, for withdrawing land from production. This program was promoted as a means of soil conservation, holding land in a "soilbank" as the program was popularly called. It was hoped that production would be reduced sufficiently to raise the price to something like \( P^{**} \), i.e. to shift the supply curve sufficiently to the left to make \( P^{**} \) the new market equilibrium. This is shown by the intersection of \( D^{*} \)
and $S_2$ in figure 19. To reduce production to this level in the face of the alternative price $P^{**}$ producers in groups A and B would have to be offered what amounted to bribes about equal to what they would expect to make by producing and selling "surplus" output for stockpiling in government warehouses. Again, you can see that such a program favors the large efficient producers. The payment is, after all, roughly in proportion to the capacity (size) of the producer. Actually, even this clever plan fell through because it over-estimated the amount that production would be reduced as the soil was "retired" from production. Farmers tended to remove only their worst land from production and to farm the remainder more intensively with the result that the old program of surplus again arose.

Who Pays for the Farm Programs?

We have seen that the main beneficiaries of these programs have been the larger and more efficient producers and, to a small extent, the ostensible objects of the policy, viz. the family farms. Who pays? Depending upon whether the program raises the actual market price directly or only raises the price received by farmers indirectly, the cost of these programs is borne by consumers or taxpayers. Either way, the losers of these schemes are mainly everyone except the farm groups. The distribution of the burden may vary somewhat depending upon where different individuals fall into
tax categories or how much they consume of particular goods, but the cost is shared by a very large group that included almost everyone to some degree.

Are Such Programs Justified?

An economist can point out the effects of these various proposals but whether or not they are justified is a matter of one's individual values. Of course, greater clarity as to the real effects of the program probably does affect how people feel about these proposals.

We have seen that out of a concern for the plight of poor farmers and the threat to an old and dear institution various remedies have been suggested. Each of these remedies, in fact, tends to benefit unintended (?) groups more than family farmers although all the proposals have the potential of accomplishing the stated objectives.

One may wonder whether a concern for the financial plight of poor farmers might not be addressed in some more efficient way, i.e. one that benefits the needy without burdening the general public with subsidies to "less deserving", more wealthy recipients. Some would suggest that the farm problem will not be permanently solved unless the market is allowed to operate, that whether or not one remains in farming should be determined by the efficiency of the producers. In the meantime, one could aid poor farmers because they are poor (as one would aid anyone poor) not because they are farmers.
These measured might include direct income subsidies, re-training or otherwise aiding occupational opportunities, etc. This approach might well doom the family farm to extinction or, at least, allow its decline to continue. Here, once again, values come to play and we must determine whether preservation of this institution is worth the cost. But always remember that there is a real cost of such a policy and someone, maybe you, must bear that cost.
Chapter 8. "International Trade—Specialization by Countries"

INTRODUCTION

This chapter applies a number of the basic principles of economics to explain some aspects of international trade in goods and services. Like the trading already discussed, international trade arises and bestows benefits because it allows specialization in production. We shall be particularly interested in the benefits and how they are affected by measures that impede international trade. Our concern is determining how trade and interference with trade affect the allocation of resources.

A Simple Model of International Trade

It is not necessary to explain the phenomenon of international trade in terms of the many countries and millions of goods and services that comprise actual trade flows. The essential aspects of international trade can be discussed with a model of the world that is highly simplified. Advanced treatments of the subject differ mainly in details rather than fundamentals. Accordingly, we shall suppose that there are but two countries and one commodity involved in trade. These are America, Japan and textiles, respectively.

First, we shall consider each country's supply and demand functions for textiles. Relative domestic conditions, i.e. how one country's supply and demand functions compare to that of the other country will form the basis of our explanation.
of international trade. In figure 20 the supply and demand conditions are given for America, assuming all production and consumption of textiles is limited to the U.S. market. Without international trade, American producers can sell only to American buyers and the latter can purchase only from domestic sellers. There is nothing unusual about these supply and demand functions and they would lead, in the absence of trade, to equilibrium values $P_0$ and $Q_0$. Prices above $P_0$ would cause an excess supply, prices below $P_0$ an excess demand. Of course, all prices other than $P_0$ would be temporary in a free market, given these conditions of supply and demand. Imagine, however, that the price could somehow be maintained at levels above and below $P_0$. Gaps between domestic production and consumption would arise, progressively larger the farther one moved away from $P_0$ in either direction. Above $P_0$ domestic production of textiles would exceed domestic consumption; below $P_0$ the opposite conditions would hold.

If we drop the assumption that buyers and sellers of textiles are restricted to the domestic market, we can conceive of the "excess supplies" above $P_0$ and the "excess demands" below $P_0$ as indicating the supply function of American producers for textiles export and the demand function of American producers for textiles export and the demand function of American consumers for textile imports, respectively. At prices above $P_0$, American producers would want to produce more textiles than domestic consumers would buy. Indeed, the existence of other
Figure 20
Market for Textiles - U.S. Domestic Market Only
markets is the only thing, short of some kind of interference, that could allow the price to be different from $P_0$. Similarly, American buyers could only satisfy their demand for textiles at prices below $P_0$ if other foreign producers made up the "shortage" between domestic production and desired consumption. We depict these American supply and demand functions for exports and imports in figure 21.

This same procedure can be repeated for our world's other country, Japan. This is done and summarized in figure 22. In figure 23 the supply and demand conditions of both countries are brought together. Note that the domestic equilibrium price in Japan is $P_1$, the one that would prevail if trade was limited solely to the Japanese domestic market. $P_1$ is lower than the corresponding price in America. Similarly, the Japanese export and import functions "originate" at the lower price $P_1$.

Ignoring the costs of transportation, the equilibrium price of textiles would become $P^*$ if trade was established between America and Japan. At that price the quantity of textiles exported from Japan would equal the quantity of textiles imported by America. In other words, the differences between domestic consumption and production of textiles in America ("deficit") and Japan ("surplus") are made up by the import and export of this same quantity of textiles.
Figure 21
U.S. Supply and Demand Functions for Textiles To and From the Rest of the World
Figure 22
The Japanese Textile Market
Figure 23
The U.S. and Japanese Markets for Textiles
Benefits and Losses Associated with International Trade

By contrasting the equilibrium positions of the U.S. and Japan before and after international trade, it is apparent that two groups have benefitted from the opportunity of trade and two groups have lost. The "winners" are American consumers and Japanese producers of textiles. The former now purchase a larger amount of textiles at a lower price than previously while the latter sell a larger amount at a higher price than initially. But U.S. producers of textiles sell a smaller quantity of textiles at a lower price than previously and Japanese consumers have a smaller amount of textiles available to them at a higher price than before. Several questions now arise, viz.

1. Can one estimate the relative gains and losses from trade?
2. If we can, are the gains larger than the losses or vice-versa?
3. Can trade exist if one country acts only as the buyers and the other only as the seller?

To these questions we shall respond with non-rigorous but essentially correct intuitive answers.

The gains from trade are generally greater than the losses attendant to trade. The winners' (American consumers and Japanese producers) gains exceed the losers' (American producers and Japanese consumers) losses because of the resource
reallocation trade brings. The shift in world textile production to Japan from the U.S. increases the total amount of textiles produced. This result is an example of the effect of specialization upon total production we have already seen (chapters 1-2) in the cases of two different types of land and two different types of producers.\footnote{We have ignored part of this process of specialization and reallocation which results from American resources being shifted from textiles to other uses and the corresponding shift of Japanese resources from such pursuits to the production of textiles. Including these factors would complicate the analysis but would not change the conclusion.} Appropriate specialization in production can generally increase total output to the potential net benefit of all concerned. Since the winners' gains exceed the losers' losses, the latter could (hypothetically) always be "compensated," still leaving consumers better off by the net increase in output brought about by specialization. In this sense, the gains from trade are demonstrably greater than the losses.

Of course, the winners are usually under no obligation or compunction to compensate the losers in international trade. Even if the losers recognized the overall gain from trade, they could be expected to oppose trade from self-interest because it is sure to make \textit{them} worse off. Domestic producers would find
their assets less valuable since their market and profits would shrink in the face of more efficient foreign competition. Note that not all American producers are driven out of the market by the entry of Japanese textiles, only the less efficient. This is the usual case. Resource owners who supply factors of production to the American textile industry will be worse off, too. Most noticeable will be the effects upon American workers in the textile industry. Employment will fall, wages might be unfavorably affected, and some workers may be forced to find other jobs or even experience unemployment. Clearly, producers and workers alike will have good reason to oppose the entry of foreign competition.

Interference With Trade

Opposition to foreign competition has historically been effective in restricting the scope of international trade. Despite the fact that the overall effects of trade are beneficial to all nations, producers are generally better organized to work their will than the mass of consumers. (Efficient producers in export industries favor free trade, but, of course, cannot affect the policies of countries to which they would export.) Each import-competing producer will be vitally affected by trade and will devote much effort to combatting it. However, each consumer's benefit from trade in a particular item is usually small since he consumes many commodities. Thus, the beneficiaries from trade are seldom
as well organized to promote their interests. The result has been government interference with trade throughout economic history in the form of tariffs and quotas.

A tariff is a tax upon goods produced outside the area that has the effect of raising the price that foreign producers must charge in the domestic industry. Obviously, it then becomes easier for domestic producers to meet foreign competition. Sufficiently high tariffs can effectively eliminate foreign competition altogether. In practice, tariff levels have varied greatly, sometimes amounting to only trivial interference, sometimes seriously threatening or even eliminating trade. The absence of trade restrictions, i.e. free trade has been an unusual circumstance, clearly an exception to the rule in economic history. Another reason sometimes given for tariffs is tax revenue for government. Of course, domestic sales taxes or other taxes could raise this revenue without interfering with trade. For a detailed graphical treatment of the effects of interference with trade see the appendix to this chapter.

The second basic instrument of trade interference is the quota. A quota establishes the maximum amount of a good that can be imported. Like tariffs, quotas can vary in their effect upon trade, depending in this case upon how low the quota allowance is set relative to the free trade level of imports. Sometimes quotas and tariffs are used together to interfere with trade or, as it is more likely to be described,
to offer protection for domestic producers. Such a combination is unnecessary, although sometimes useful for producers, since either device is quite sufficient to obstruct trade to any degree desired. By the way, economists find quotas even more objectionable in principle than tariffs, since tariffs at least generate some government revenue.

Our simple example probably has given you cause to wonder whether trade could continue as we describe with only one commodity produced by one country for export to the other country. In general, countries like individuals, can trade only to the extent that they have goods to offer in return. In practice, international trades like their domestic counterparts are conducted through the medium of money rather than a direct barter arrangement of goods for goods. Just as individuals in the same country find money a convenient tool for effecting trades, individuals residing in different countries use the various forms of international money in facilitating international trade and specialization.

A number of technical problems arise because each country has its own currency and the mechanisms used to trade currencies sometimes breaks down. Such situations, which are often sensationalized in newsreporting, are called "balance of payments crises". Without going into great detail, we can explain balance of payments problems by again considering the analogy of an individual consumer. To receive goods from others, an individual must produce goods which he sells se-
curing a money income which he in turn uses to purchase goods produced by other "specialists." For some time he could make purchases by relying upon accumulated money balances, other forms of wealth he had acquired, or credit extended by sellers all the while consuming without producing goods for exchange. Eventually these non-income sources of purchasing power may run out. Countries, like individuals can behave similarly, although this usually takes the form of producing some goods in exchange for even more imports. Balance of payments crises arise when, like individuals, countries are eventually forced to bring their expenditures into line with their basic means of purchasing power. Countries like individuals usually find this an uncomfortable process. Furthermore, the process is often complicated by political factors since the method chosen to reconcile the balance of payment problem can create a whole new group of winners and losers, those who have held various currencies.

At this point, another question about trade may have occurred to you. Can a country always manage to sell abroad to finance its imports? After all, we all know that the per unit labor costs of foreign producers are less per worker than the costs of American producers. Again, we will offer a non-rigorous but valid answer.

As in all cases of specialization, even if one producer could produce all goods at a lower absolute cost, he would find it advantageous to concentrate upon those goods in which
he was particularly superior, i.e. the goods in which his relative costs were lowest. Remember the example in Chapter 2 of Joe and Pete. The same principle applies in the case of specialization between countries. Even if one country had the absolute advantage in the production of all goods, it would find it to its advantage to concentrate upon the production of only some goods, viz. the ones for which its comparative advantage was greatest. It would allow other countries to produce the goods in which they (the countries with higher absolute costs) have a comparative advantage. In this way, all countries can participate in trade at their mutual gain regardless of the distribution of absolute costs among them.

By the way, you should not accept popular versions of the labor cost argument at face value. It is often alleged that American workers and firms cannot compete with foreign workers because the latter earn such low wages. Such competition is said to be "unfair," presumably because it threatens the standard of living of American workers. Throughout the 20th century American workers have earned wages higher than their foreign counterparts, even when the U.S. was selling much more abroad than it bought in return. The same was true of 19th century England when it was the world's leading exporter. In both cases lower wage economies were "forced" to erect tariff barriers to protect their own producers. These examples should suggest the essential fallacy under "low wage-unfair
competition" arguments for trade protection. The argument is based upon differences in absolute costs rather than relative costs. Free trade benefits the participating countries as a whole but, as explained earlier, firms and workers in industries facing more efficient foreign competition can be harmfully affected. Of course, they attempt to paint the picture in somewhat different colors, implying that unfair competition is a threat to all citizens of a country.
APPENDIX: A GRAPHICAL DEPICTION OF INTERFERENCE WITH INTERNATIONAL TRADE

In this section we will demonstrate the effects of restrictions upon international trade. Such interference makes some groups better and worse off. In other words, the presence or absence of free trade creates winners and losers. Consider the domestic and international markets shown in figure 24 which is the same as figure 23. As previously explained in the absence of trade restrictions, the price of textiles would be the same in both markets, $P^*$, (ignoring costs of transportation). America is the importing country so that domestic consumption ($Q_1$) exceeds domestic production ($Q_2$). Clearly, U.S. consumers benefit from international trade while domestic producers are worse off. Just the opposite is true in Japan, the exporting country.

Suppose that free trade is abrogated by the imposition of a tariff on imported Japanese textiles. Beneficiaries of this action will be American producers and Japanese consumers while the losers are U.S. consumers and Japanese producers. The size of these effects depend upon the size of the tariff. The greatest costs and benefits result when the tariff reaches a size sufficient to eliminate all trade. Further increases in the tariff have no perceptable effects.

The effect of the tariff is shown by the upward shift in the supply curve in figure 24. The new supply curve of
Figure 24
The Effects of Interference with Trade
Japanese textile exports to the U.S. is the curve $S + T$ which lies above the old supply curve by the amount of the tariff. (For simplicity we have assumed that the tariff is the same regardless of the price of textiles although tariffs are often applied as a percentage of the price). The new equilibrium price is $P_t$ which lies between $P^*$ and $P_o$. Higher tariffs will raise the price toward $P_o$ when exports would cease. Note that $P_t$ lies above $P^*$ by an amount less than the tariff itself. This "sharing" of the cost of the tariff is the general case. The price received by the seller ($P_t - T$) is less than $P^*$ but not by the full amount of the tariff. You will find it a useful exercise to consider how changes in the shapes of the supply and demand curves for exports redistribute the costs of the tariff between these two groups.

There is a reallocation of resources resulting from the imposition of the tariff which is shown by the changes in the quantity of textiles produced and sold in the U.S. and the quantities of Japanese textiles sold in the U.S. and Japan. U.S. production increases from $Q_2$ to $Q_3$ while Japanese exports to the U.S. fall from a number equal to $Q_1 - Q_2$ to a number equal to $Q_4 - Q_3$. Japanese domestic consumption of textiles rises from $Q_5 - Q_6$. Again we can see that the tariff has benefited American textile producers and Japanese consumers at the expense of Japanese textile producers and American consumers. The losers' losses exceed the winners' gains, at least in monetary terms, because Japanese producers could
have produced the American output induced by the tariff at a lower cost.

The effect of a quota on the quantity of Japanese textile imports and the price of textiles in the American market is very similar to the case of tariff restriction. A quota, you will recall, is an absolute legal limit on quantity of goods imported. Like a tariff, a quota has its maximum effects when trade is entirely eliminated (quota = 0). However, a quota might have no effect upon trade if the quota was set at a quantity greater than the desired quantity of imports. In between, where the quota establishes a limit upon which imports which is less than the amount that would be desired, its effects are the same as a tariff which ranges from some positive amount up to a tariff sufficiently high to eliminate all trade. Consider figure 25. We have derived the demand and supply functions for Japanese exports in the same manner as previously and will concentrate upon the effects of the quota on the U.S. market. In the absence of trade restrictions the desired amount of imports of Japanese textiles to the U.S. would be equal to the distance between \( Q_1 \) and \( Q_2 \) in the U.S. market which is equal to the quantity \( Q_3 \) in the world market diagram. Suppose that a quota is established limiting imports of Japanese textiles to \( Q_4 \). The effective supply curve of textiles for the U.S. market is now shown by \( S_q \) for which the quantity of the textiles available exceeds the domestically produced supply by the amount of imported
Figure 25
The Effect of a Quota Upon Trade
textiles \( Q_q \). The horizontal distance between \( S \) and \( S_q \) is the quantity \( Q_q \) at all prices. The new equilibrium will be at \( P_q \) as shown. The only difference between the quota and a tariff which limit the amount of exports to the same extent is that the latter yields government revenue which presumably benefits citizens while the quota (if given "free" to some licensee, as is typical) is just a windfall for some private individual.
Chapter 9. "Inflation—A General Rise in Prices"

INTRODUCTION

So far we have been concerned with prices and outputs in individual markets, that part of microeconomics dealing with what we call partial equilibrium. We can now extend our view to consider changes in the level of all prices together. The level of prices is customarily computed as an average of the prices in many markets. An increase or decrease in the level of prices, i.e. inflation or deflation, does not imply that all prices are rising or falling by the same amount or even that all prices are changing in the same direction. The rate of inflation or deflation is simply the average rate of change of prices, usually expressed as the percentage rate of change per year. Since changes in prices affect the economic well-being of buyers and sellers, concern over the behavior of the price level is an old story in economic history. "Concern" is usually a short step away from imputing responsibility and blame for the change in the price level. When this has occurred, governments have often taken action to "improve" conditions, usually by attempting to reduce the rate of change of the price level. Incidentally, although inflation has been of much greater concern in recent years than deflation, don't get the impression that price level changes are a one-way street. The U.S. has experienced several periods of considerable deflation in the last century, the most recent associated with the Great Depression. Our
theory will be equally applicable to periods when the price level is falling as when it is rising, although we will concentrate upon the latter.

**What Governments Usually Do About Inflation**

Although a complete explanation of government macroeconomic policy must be delayed until we have explored some additional aspects of economic theory, we can already evaluate one important aspect of that policy, namely price controls. This policy action, in its various forms, has become the conditioned response of governments when the behavior of the price level as determined in free markets has become "unreasonable." The price level is ordered to "behave", usually by a government fiat suspending all price changes and providing for penalties for violation of the order. Unfortunately, this straightforward approach to the problem which is generally seen as the "common sense" solution by politicians and citizens alike, is based on the fallacy of mistaking the symptom of the problem for its cause. Price control policy in itself is seldom, if ever, an adequate or effective means of combatting inflation. We shall soon see why, but first let us briefly consider why changes in the price level are sometimes undesirable.

**The Costs of Inflation**

It seems self-evident that inflation is bad because most people think of the upward movement of prices as an isolated phenomenon. One typically reasons that "if prices rose less
rapidly, or not at all, I would be better off, providing that my income in dollars remained the same." But incomes measured in money terms tend to move in the same direction as changes in the price level. Therefore, it is unrealistic to imagine that the price level could change independently of the average level of money incomes or that the behavior of money wages would be unaffected by a change in the price level.

However, what is generally true on the average, is not by any means always true for each individual or family. While some incomes are rising as fast or even faster than prices, some incomes lag behind the rise in prices so inflation makes these people worse off than they would otherwise have been. Such individuals claim (truthfully) that inflation has made them worse off. However, this common complaint is not true for most people that are wont to make it. A kind of statistical illusion is at work. Individual prices and wages generally rise intermittently and sporadically. Therefore, some price increases are occurring all the time during a period of inflation. It often seems that because the price level is rising continuously that rising prices outpace one's income (which typically rises only periodically). It is easy to imagine that one's income has, at best, "caught up" with the price changes that have occurred since the last change in income.

The appropriate computation to make to determine whether incomes or prices are "winning the race" during a period of inflation (or deflation for that matter) is to compare rele-
vant percentage rates of change rather than to rely upon "impressions," however strongly felt. During periods of economic growth, which includes almost all the years of this century in the U.S., overall incomes have in fact grown faster than prices. Most people have experienced improved economic well-being, even during inflations. This does not imply that the inflation made them better off, but only that it has not been harmful in the way people usually think of the problem. Of course, how people feel about a problem often is more important than how "experts," (economists in this case) view the situations. This is especially true when a political solution is sought, since the laymen always far outnumber the experts. The continued popularity of price controls as a method of curing inflation is a classic case of this phenomenon.

Having dispensed with the common but mistaken objection to inflation, we must acknowledge the reasons why inflation is harmful to economic systems. To begin with, both firms and consumers try to base their plans on anticipations about the future course of events, including the future course of prices. Most of us prefer certainty to uncertainty about the future, and we are willing to devote real resources to help make plans accordingly. One of the clearest costs of inflation is that an economic system with a price level subject to unpredictable changes will cause firms and consumers to need to spend more resources to make effective plans than one in which the future course of prices is known with certainty.
This uncertainty requires an input of time and other resources just to allow businessmen and consumers to attempt to predict future prices. Even with this effort, wrong and wasteful decisions are made because price behavior can turn out differently than expected.

A second valid argument against inflation is that while it has no adverse effects on the average level of real incomes apart from the effect just noted, it can arbitrarily redistribute income among different groups. For example, two parties to a contract often agree to some kind of exchange that will take place over time. One party might borrow a sum of money from the other and agree to pay it back along with interest in several installments. If subsequently prices rise more than the parties anticipated, the borrower will find the contract more favorable and the lender less so than originally anticipated. The dollars repaid are worth less than the parties believed they would be when they made the contract.

Many individuals make such commitments to receive or pay sums or services in the future. It behooves them to make as good an estimate of future prices as it does considering any part of the bargain they strike for the reason just explained. Of course, some people will make better estimates than others because they are more informed, or luckier, or smarter, or for some other reason.

The concern here is for those who make incorrect decisions and become worse off. Anyone receiving a fixed income
who underestimated future inflation is made worse off. Since everyone is equally subject to these risks and should exercise the same prudence before making arrangements that extend into the future, you might legitimately ask why we should be particularly concerned when this kind of decision turns out badly for some people. The reason probably lies in a belief that those who are most adversely affected by unanticipated inflation are likely to have relatively low incomes under any circumstances and thus require some protection from the vagaries of the price level. We tend to see this especially in the case of the elderly poor who rely mainly on welfare and social security payments for their support. Thus, many people conclude that the unforeseen effects of price changes upon economic well-being are undesirable and that reducing the variability of future price movements would be desirable.

A third reason for preferring less fluctuation in the price level is introduced when exchange between countries that use different currencies takes place. Although trade is fundamentally an exchange of the commodities one country produces for those of another, (as explained in the previous chapter) trade is conducted through the medium of exchange of money. A unit of currency (dollar, lire, etc.) becomes worth less in terms of a particular country’s goods to any holder, foreign or domestic, if the price level of that country rises. Therefore it behooves one to anticipate the price changes of any country whose currency one holds. Since changes in the price
level of different countries are not entirely predictable, this uncertainty may inhibit trade and its benefits. Exactly how large a problem unanticipated inflation becomes to international trade depends partly on the monetary system traders establish. Unfortunately, the prevailing system sometimes allows unforeseen rates of change of prices to cause problems, which is a major reason many have proposed another system. However, given the present international monetary system, another reason to favor less erratic price behavior arises.

Understanding the Process of Inflation

The popularity of price controls as an effective means of "improving" the behavior of the price level stems largely from a misunderstanding of the process of inflation. An increase in the average of prices comes from the same basic source as an increase in a single price, an excess of demand. The price level rises when conditions of excess demand prevail in most markets. While we have yet to develop a theory that explains the determination of the level of demand for all markets together, we can state that if the level of demand for all markets grows relative to the supply in these markets, then excess demands will appear, leading to rising prices in individual markets. If prices are allowed to increase, the average level of prices will rise, reflecting the underlying demand conditions in the markets.

As simple as is the process of inflation described above, casual observers in the markets typically do not perceive it
as operating in this manner. Rather, it usually appears that prices rise in response to increases in costs rather than increases in demand. This view is understandable given the perspective of the businessman or other participant in a single market. It is dangerously misleading, however, for understanding the overall process of inflation. Our first task is to reconcile what economists believe is the real process of inflation from what inflation seems to be from the view of an individual buyer or seller.

**Mistaking the Cause of Inflation**

Many sellers determine prices on some kind of "mark-up" basis. The price is set as some predetermined percentage of the cost to the seller. Sellers usually have some percentage they customarily apply as a "rule of thumb." This mark-up may vary in the light of short-run conditions or may be permanently changed in the light of new experiences or prospects but we shall take it as given for the example that follows.

Suppose we consider a retail seller of dresses who has some how already determined his mark-up policy. How does inflation manifest itself in the markets in which he acts? Three markets in the chain of distribution of cotton dresses are shown in figure 26. Initially each market is in equilibrium at its own price $P^*_1$. Note that the price is different in each market. The $P^*_1$ of the retailer is related to the $P^*_1$ of the dress manufacturer through the mark-up. The cotton cloth price,
Figure 26
Prices, Costs, and Shifts in Demand
P_h helps determine the prices at which manufacturers are willing to supply dresses to retailers. Suppose that this equilibrium is upset by an increase in the demand for cotton dresses at the retail level, as shown by the shift in demand from $D_1$ to $D_2$.

Since the retailer established his price as a mark-up of wholesale price, the initial effect of this demand shift is simply an increase in sales at the initial price. But if retailers in general are experiencing the same increase in demand, total orders placed with dress manufacturers will increase as retailers seek to replenish their depleted inventories. This secondary shift in demand is shown by the shift in curves from $D_1$ to $D_2$ at the level of the dress manufacturer. For simplicity, we will assume that the demand for cotton is unchanged because the increased demand for cotton dresses is just offset by a decrease in the demand for cotton elsewhere.

Now because his costs rise in producing the increased output, i.e. his MC curve is upward sloping, the manufacturer finds it necessary to increase the price he charges the retailer. The retailer, in turn, increases his selling price which is the old percentage mark-up applied to a new higher cost.

If you were to ask either the retailer or dress manufacturer what caused them to raise their prices, it is unlikely that they would cite the increase in demand. Both of them would probably state that the increased costs they experienced forced them to raise their prices. Strictly speaking, their observations are accurate, but taking them at face value as
an explanation of the cause of inflation is simply wrong. The initiating event which ignited the price increases, albeit through adjustments to cost changes, was the increase in demand. Yet from the perspective of most observers, standing in a single market, demand is often harder to observe directly than the increase in costs. This natural yet incomplete perception is probably the main reason most people mistake the causes of inflation and approve of price controls as a reasonable and effective weapon against inflation.

**How Inflation Gets Started**

We have described how an increase in the demand for one good leads to an increase in its price for given conditions of supply. To explain how the price level, an average of many prices, can rise, we must assume that the overall level of demand for goods and services has risen. During such a period the situation we just described in the dress market becomes general. Participants in various individual markets are likely to perceive the increase in prices as a response to increases in costs. However, they are mistaken. It is demand which has caused sales to rise. In turn, inventories are depleted, costs of production rise, and finally prices tend to rise. The chain of events linking an increase in demand to an increase in price may take considerable time to unfold.
Consider the effect of inflation, once under way, on the decisions of the participants in individual markets. They probably form their expectations about future prices on the basis of past, especially recent past, experiences. The initial rise in prices was probably unanticipated since it deviated from the immediate past. Once begun, it may well have accelerated somewhat, i.e. the rate of increase of prices may have itself increased. Only with time will individuals incorporate these new experiences into their thinking, in the process revising their expectations about future prices. The longer the period of time during which the process of inflation, especially an accelerating inflation, is experienced, the more likely it is that people will come to expect more of the same in the future. This condition is popularly described as "inflationary psychology." Technically, we say that buyers and sellers tend to have "adaptive expectations," that is they tend to adjust, or adapt, their expectations of the future to their recent experience.

Once future inflation has become widely anticipated, buyers and sellers plan their purchases and new orders on the basis of prices which they expect to increase over time, and in the typical inflationary situation they are not disappointed, at least not immediately. The importance of this adjustment of expectations to our understanding of inflation is this: when a reversal in the basic processes which caused inflation (overall demand) takes place, it takes some time before buyers and
sellers fully adjust to it. Before the time they adjust to the slower rate of growth of demand, they will again suffer the effects of mistaken anticipations. While this learning process is going on, buyers will tend to order more than they can sell, because they mistakenly expect prices to continue to rise at the recently experienced rate of inflation. Similarly, sellers of labor, as well as employers, expect to be able to charge, and pay, higher wages than conditions actually turn out to warrant. The result of these mistaken anticipations in thus attempting to stop inflation, once it gets started, is painful and costly. Once buyers and sellers begin to adjust their expectations to the slower rate of inflation, the unintended increases in inventories tend to fall, as does unemployment. In the meantime, it can be politically very difficult for the government to attempt to control inflation in the face of the adverse effects on businesses and workers brought on by anti-inflationary policies.

Recent Experience With Inflation

One reason the Nixon administration resorted to price controls in 1971 after long conspicuously resisting the clamor for controls was the "apparent" failure of the policy most economists prescribed to reduce the rate of inflation. As we have described, economists see the problem as one of excess demand. Elimination of the excess demand was recommended as the appropriate cure for inflation. Beginning in 1967 the
government took a series of actions widely publicized as taking the economic medicine economists suggested. Apart from whether or not this medicine was really taken as prescribed, or exactly which prescription was the most appropriate one of the several proposed by different economists, we must ask why inflation appeared to be so stubbornly resistant to traditional cures.

The excess demands which accumulated in the late 1960's had been mounting for a number of years. Prices had risen in most markets although not always immediately, for reasons we have already described. The general public by this time had come to expect inflation and had adjusted their buying and selling habits accordingly. Thus, the imposition of a moderate effort to reduce total demand in the economy brought about some relief from inflation but also, it caused unemployment to rise. Policies which had the effect of reducing excess demands were not reflected immediately in a sharply reduced rate of inflation, because price behavior was affected by the momentum (anticipations) built up in the experience of past excess demand conditions.

Those economists who most strongly accept the above explanation of economic affairs 1967-1971 note that the rate of inflation was reduced significantly during the period before "Phase I" was introduced, before direct controls became effective. The rate of inflation peaked at 6.2% in early 1970 and fell to 4.3% before controls. After the controls were applied the official rate fell further to about 3% in 1972.
How Price Controls "Control" Inflation

Price controls aim at stopping inflation by stopping the increases in costs which are believed to be the cause of rising prices. We have seen that a general increase in demands relative to supplies in markets will cause excess demands and upward pressure on prices. These increasing prices are only symptomatic of underlying market conditions. Such an overall increase in demand relative to supply was felt during the late 1960's in the U.S. Like many periods of inflation, this one was associated with the increased expenditures of conducting a war. Wartime need not cause inflation although the two generally are associated because of a reluctance on the parts of individuals and/or governments to reduce other forms of spending by an amount sufficient to prevent excess demands from becoming widespread. As usual, the rise in the price level, which had begun to accelerate, was perceived as a general response to higher costs. Calls for price controls began and ultimately were answered.

Economists, on the whole, are dubious about price controls as effective policy measures for the reasons we have elaborated. This skepticism is often expressed in sarcastic analogies which attempt to dramatize the fact that controls attack the symptom rather than the cause of inflation. For example, controls are likened to the breaking of a room thermometer as a means of reducing the temperature. Nonetheless,
controls continue to find favor across the political spectrum. The most recent experience of controls in the U.S. from late 1971 to early 1973 has probably strengthened this faith because the rate of inflation tapered off slightly during Phases I and II. Critics of price controls charge that such a reduction was well under way before the controls because excess demands had generally subsided. Phase III, which appears to relax the control mechanism may well see a reversal of a falling rate of inflation, critics continue, since excess demands appear to be re-establishing themselves. Supporters of controls would probably see a return to greater inflation as confirming their beliefs since Phase III controls have been made less stringent.

Furthermore, most economists would conclude that price controls are harmful as well as ineffective measures to fight inflation. This harm arises because controls are likely to affect relative prices even though the absolute level of prices is largely unaffected. Controls distort the pattern of relative prices, giving rise to "shortages" and obstructing the allocations of resources. These effects are more apparent the longer controls are in effect and the stronger are the forces of demand. Price controls as applied during Phases I-II probably had minor allocation effects (outside a few notable exceptions such as timber) because overall demand forces were slowing and the control program was in effect for only a brief time. Some economists find price controls objectionable on other grounds. They feel that price controls constitu-
tute an unwarranted interference of the government into private decisions. In addition, price controls are inevitably applied more stringently in some markets than others. This may seem arbitrary and unfair.
Chapter 10. "The Meaning of Profit Maximization"

INTRODUCTION

In the last two chapters we have applied the economic theory we developed earlier. We now return to the theory of supply to explain the behavior of business firms. Economic theory assumes that business firms attempt to maximize their profits. The notion of profit maximization conjures up a picture of a miserly capitalist grubbing for every extra cent he can, by fair means or foul. As a description of firms' behavior, this image fits some enterprises better than others. Nonetheless, economists have found that a theory of firms based upon the assumption that firms behave as if they were seeking to maximize their profits leads to more accurate predictions than alternative theories. This chapter seeks to explain the concept of profit maximization.

Before we discuss the theory of the firm and profit maximization, it would be useful to note what the theory we are developing is not. In particular, despite the connotation of profit maximization, the theory of the firm does not purport to be a practical guide for the operations of a firm. While a study of economics can give one insight into the behavior of firms, the theory is primarily designed to predict what firms do rather than how they do it. Similarly, while a familiarity with formal economic theory can be useful to a businessman, such knowledge is neither a necessary or sufficient condition for success in operating a firm. Students
often find the above disclaimer surprising. After all, "the firm" has such an important role in economic theory that it would seem that a study of economics would naturally prepare one to manage a firm. Actually, the study of various aspects of firms' operations such as marketing, management, accounting, etc. as typically offered in a business administration curriculum, is the appropriate formal training for understanding the problems of running a firm. That the study of economics is valuable for such training is evidenced by the fact that courses in economics are usually required in business administration programs. But a study of economics alone does not make a business manager.

To understand why the theory of the firm is not a direct practical guide for business management, let us consider what the theory actually is. The theory of the firm posited in microeconomic theory visualizes the firm's situation as follows:

1) the firm is dealing in an ideal world, i.e. a world of certainty and complete information

2) the firm has a single objective, viz. the maximization of profits

3) the only variable subject to the firm's discretion is output and through it the firm seeks to maximize its profits.

Those familiar with operations research will recognize that this set of assumptions closely resemble an operations research problem where profit maximization is the specified objective and output is the instrumental variable. Those unfamiliar
with operations research need not be concerned with this analogy.

While advanced treatments of the theory of the firm refine these assumptions somewhat, allowing for more "realism" and detail, it is this simple form of the theory which we shall employ. At this level of abstraction, the theory of the firm is less than a practical guide for a businessman. But to repeat, the theory seeks to predict what firms do in response to changes in prices, costs, and other market phenomena. How businessmen operate their firms from day to day and how they decide and actually undertake the courses of action predicted by the theory of the firm is beyond the scope of our study. One can safely say, however, that some businessmen are more adept at making the right decisions than others. These decision makers tend to prosper and grow, leading by example toward the maximization of profits. Competition takes care of the rest, in more ways than one.

**An Application of The Theory of Profit-Maximization**

To demonstrate the workings of the theory let us apply it to a familiar market, the retail pizza pie market. The important features of this market are as follows:

1) different firms produce very similar products

2) firms operate in a competitive environment, that is, one where entry to the market is open

3) the entry of new firms and/or revised output decisions of old firms affect the market price
Although nowhere in this description is profit maximization explicitly mentioned, the theory predicts that behavior with profit maximization will follow from these conditions. In other words, firms need not be aware of profit maximization for the theory to successfully predict their behavior. The theory predicts only that firms behave as if they sought to maximize profits.

Suppose that we began our look at the pizza market by determining the weekly output and profits of the various firms in operation at some moment. Assuming this information was forthcoming, we depict the situation of a representative firm in figure 27 where the vertical axis measures profit and the horizontal axis measures output. Each point represents the firm's profit at a given level of output.

We would expect to find a scatter of points indicating that 1) different levels of profit were earned by firms operating at the same output 2) the same levels of profit are earned by different firms operating at different levels of output.

As we shall see, these conditions will not hold in the long run. At the present our interest is in those firms operating near the curve II. This curve indicates the highest profits earned at the various outputs. This curve shall provide the mechanism for the adjustments firms make that lead them to profit maximizing behavior.
Figure 27
Various Levels of Profit and Output at a Point in Time

Figure 28
The Effect of Time on Profitable Levels of Output
Clearly, at this moment in time, some firms are doing better than others in earning profits. For example, the firms A and B are just "breaking even," i.e. their revenues are just equal to their costs (which include the opportunity cost of resources supplied to the firm by its owners). While other firms also have zero profits like A and B, it is clear that no firm smaller than A or larger than B even does this well. In other words, to earn a profit, given present market conditions, firms must produce an output between that of A and B.

Some firms are earning positive profits which means that the resources they are employing are earning more than they could elsewhere. But similar resources are being employed elsewhere in different pursuits. Therefore, given some time, we would expect new firms to enter the market. In addition, existing firms are likely to adjust their outputs to more profitable levels. We would expect firms to gravitate toward the quantities between the outputs of A and B. One suspects that the main reason that old and new firms alike move to this area is that they observe that here is where the profits are to be had. But we cannot assume what we have set out to prove, viz. that firms act like profit maximizers. An alternative explanation, not inconsistent with profit maximization but not dependent upon it either, is that firms that fail to operate within the profitable range tend to disappear from the market in competition with the others. Survival is a selective pro-
cess, those astute or lucky enough to make the right decisions, i.e. profitable ones, tend to outlive their less fortunate competitors.

The effect of firms' adjustments in output is to shift our curve indicating the greatest profits attainable at various outputs toward curve II-IT and eventually to curve III-III in figure 28.

The effect of the entry of new firms and the adjustment of old firms toward the more profitable levels of output is to make profits lower at every level of output. This is shown in the contrast between curves I and II. The reason for diminishing profits at all output levels is that the competition is getting keener. Surviving firms are those that succeed in operating at a profit. Competition weeds out firms that fail to behave like profit maximizers regardless of whether the firms consciously have the maximization of profit as a goal. The curve II is simply an intermediate position on the way toward the equilibrium for this competitive market where all firms produce output M and just break even. M is the best size for the firm given the overall demand for pizza (which we have held constant throughout the process of adjustment). That M is the best output is evidenced by the fact that it is the only output at which firms can survive and that the greatest profits were earned by firms operating at M all during the period of market adjustment.
The theory of the firm is really "the theory of the surviving firm." Firms that survive behave as if they were seeking to maximize their profits. We do not attempt to describe the firms' behavior in any detail but can safely predict that the survivors act like profit maximizers.

Explaining a Discordant Note on Profit Maximization

We should reconcile our theory with the casual observations and the testimony of firms themselves that indicate that firms do not seek to maximize profits. Such observations and testimony should not be accepted at face value for, as we have seen, whether or not firms seem to or think they are attempting to maximize profits, competition will ensure that those who do maximize profits survive while others perish.

A prominent study of this kind was conducted by an economist named Joel Bain in the early 1960's. Bain asked a large number of firms what their main goal was. Only a few (about 5%) indicated that they sought to maximize profits. More common responses were that firms sought to maximize sales or market shares or that they sought to act in a "socially responsible" fashion. First of all, one suspects that the conscious goal of profit maximization was more important than this survey would indicate but that the low public esteem of profit maximizing business enterprises tempered the responses given. More fundamentally, these results must be discounted because any sample of firms at some point in time includes
two categories of firms: those who eventually fail because they do not behave like profit maximizers and other firms who survive when they are forced to adopt such behavior. The interview technique used by Bain could not properly distinguish between the survivors and the losers or anticipate behavior competition would force upon firms that would survive.

Students, too, often find this theory of survival difficult to accept at first. Indeed, the theory abstracts from certain familiar complications which might alter some details of the analysis of a particular case. By and large, however, economists have developed and come to accept the theory of firm survival because it has provided the most reliable explanations and predictions of the behavior of firms in competitive markets.

Any theory is only as good as the predictions about events it yields. Like scientists in other disciplines, economists formulate theories, test their predictions against actual events, refine the theory, test again, and so on. In common with other social scientists, economists seldom have the opportunity to develop their theory under the carefully controlled laboratory conditions of the physical sciences. This makes the task of theory development more complicated and often leads to predictions less precise, although no less scientific, than in the physical sciences.
Now, we shall consider another aspect of the theory of the firm. While we argued previously that firms need not consciously behave like profit maximizers for the theory of survival to hold, we suspect that firms are motivated, to varying degrees, by the desire to maximize profits. This aspect of our theory will demonstrate how such firms behave and will enable us to predict how they respond to changes in market conditions (prices, costs, etc.).

Meet Joe The Pizza Man: Would-Be Profit Maximizer

Since we already know that surviving firms are forced by competition to emulate the successful profit maximizers, let us consider how such an industry leader behaves. Enter Joe the genial pizza man, proprietor of a typical firm of the pizza market we examined previously. Joe wants to find out how to maximize profits. We will go through this learning process as if we, too, were pizza proprietors, looking over Joe's shoulder, as it were.

From the standpoint of our theory's "operations research orientation," Joe and every other firm must successfully answer a number of questions to survive. Joe might never have heard of operations research, or economic theory for that matter, but he still must decide the following:

1) what shall I produce?
2) how shall I make it?
3) for what price shall I sell my product?
4) how much output should I produce?
Because of competitive forces, Joe's choices are fairly circumscribed, provided he wants to survive, let alone maximize his profits. Let us assume that he has already decided to sell pizza only, has discovered the best method of combining the various inputs, and has found that he must charge the same prices as his competitors. Joe probably discovered this all through a process of trial and error as well as employing an observant eye (and nose) to survey what the competition was up to. Many firms fail to find the right answers to these questions and pay the penalty of extinction. The correct answers change over time so competitive forces keep Joe and other pizza men on their respective toes. Those pizza men who most consistently come up with the right answers earn the greatest rewards. Their firms are not more profitable in the economic sense (although they appear to be so in the sense an accountant measures profit) but the "extra" rewards represent a return on their owners' unusual skills in providing pizza buyers with what they want.

We have already seen that market forces have determined what Joe produces, how he produces it, and what he sells it for. The only decision left seems to be choosing the best output. In the long run, of course, even this decision is determined by competitive pressures. But Joe wants to figure out how to be a profit maximizer initially rather than only follow the lead of others in order to survive.
The Profit-Maximizing Output

Like other firms, in and out of the pizza business, Joe's is seeking to maximize profit. Profit is equal to the difference between the firm's sales receipts (or total revenue) and its total costs of operation. Symbolically: Profit = TR - TC. Both revenue and cost vary with output and, in general, different outputs yield different levels of profit. How do we find the best, i.e. the most profitable, output?

Some aspects of the revenue and cost conditions Joe's firm must live with are depicted in figure 29. Joe himself might not go through such a formal analysis himself but, then again, he might not have the complete information we will consider. Businessmen must operate in a world of uncertainty so some of this information Joe will have to discover himself, as he also found some of the other answers previously, in order to survive and prosper.

Figure 29 shows the total revenue and total cost curves as related to output. The total revenue rises at a constant slope, implying that the revenue per unit is the same at every level of output. The firm is so small relative to the entire market that its output decision does not affect the market price. The total cost curve rises with output, too, but the rate of increase differs at different outputs. The most straightforward route to the profit-maximizing output might seem to be to join these two curves together and find the output for which total revenue most exceeds total costs. We shall do this in
Figure 29
The Total Revenue and Total Cost Curves
due course but a more general method of seeking the profit-
maximizing output requires the consideration of two additional
curves shown in figure 30.

The first curve in figure 30 is the marginal revenue
curve, defined as follows: **marginal revenue** the addition to
total revenue received in the sale of one more unit. Since
the firm can sell as many units at the market price as it de­sires (a consequence of its insignificant size relative to
the market), the marginal revenue is equal to the price.
**MR = P** is a general condition for firms in competitive markets.
Thus the MR is horizontal at the market price.

The second curve is the marginal cost of output where
marginal cost is defined as follows: **marginal cost** the addi­tion to total cost incurred in the production of one more unit.
As explained in an earlier chapter, the universal character­istic of marginal cost curves is their upward slope. This
portion of the MC curve indicates that additions to output
require successively more resources. Note, too, that the
falling and rising portions of the MC curve correspond to
changes in the steepness of the total cost curve. In figure
31 are gathered together all four of these curves to derive
a rule for choosing the output than maximizes profit.

The profit maximizing output is \( X^* \). This is the output
for which the vertical distance between TR and TC is max­i­mized. You will note that **MC = P** for this output. This is
Figure 30
The Marginal Revenue and Marginal Cost Curves
Figure 31
MR, MC, TR, TC Curves Shown Together
no coincidence. As a matter of fact, the profit maximizing rule we have been searching for is the following:

**The profit maximizing output is that output where MC = P and the MC curve intersects P from below.** Our task is now to demonstrate why such a rule is the appropriate one, i.e. that it leads to the highest attainable profit.

Consider some output such as $X_2$. If this was the profit maximizing output, a small increase or decrease in output must result in a reduction in profit since, by definition, the profit there would be the highest attainable. If you measure the vertical distance between TR and TC you will see that for output just to the right of $X_2$ profits are greater than at $X_2$. Therefore, an output such as $X_2$ could not be the profit maximizing one. Note that at $X_2$ the $P$ exceeds $MC$. If $P > MC$, the firm would be better off to increase output because the addition to TR exceed, the addition to TC and therefore profit would increase when output was expanded. Such conditions hold for all quantities between $X_4$ and $X_1$. By the same reasoning, outputs such as $X_3$ are not profit maximizing because the $MC > P$. Each unit such as $X_3$ (all units beyond $X_1$) adds more to TC than to TR and therefore must reduce profit.

Finally, we can dismiss points such as $X_4$ from consideration as possible points of profit maximizing output. Units just to the right (greater) than $X_4$ are produced at a $MC$ less than $P$. Thus, $X_4$ could not be the profit maximizing output since others are more profitable. We are left with the point
$X_1$ as the profit maximizing output, the only point that satisfies our rule. The shape of the MC and P curves insures that there will be only one such point for a given set of TC and TR curves.

Can Joe Find The Profit-Maximizing Output?

A rule such as the one we developed for finding the profit maximizing output is certainly not beyond the grasp of our profit seeking friend in pizza. Joe's analysis may be less rigorous but nonetheless successful in discovering such a rule for his own use. Consider figure 32.

The upper curve relates profit to output. This is the curve Joe is likely to consider first in his thinking, although he might not formalize it by graphing it as we have. This curve could be derived from our previous TR - TC curve since each point on the profit curve is simply the vertical distance between those 2 curves. The profit maximizing output $X_1$ is the same one for which $P = MC$ and MC intersects P from below as the lower curve indicates.

Joe's information is probably less complete than ours but he may come to realize the gist of the profit curve. Through experience and thought he may discover that his profit peaks at an output such as $X_1$. Expansions in output beyond $X_1$ add to revenue but by an amount less than MC. Similarly, Joe may realize, outputs less than $X_1$ are less profitable because the units "lost" would have added more to revenue
Figure 32
Two Ways of Finding the Profit-Maximizing Output
than to cost and thus increased profit. The likelihood of Joe's discovering this rule and using it to adjust to changes in market conditions depends mainly upon his "business sense." Survival will require he adopt the crucial output whether or not he understands why. Appreciation of the rule, whether by formal or informal means, will give Joe an edge over competitors. For the competitive firm the survival formula is to produce where $P = MC$. 
Chapter 11. Competition and Economic Efficiency

INTRODUCTION

Capitalists have been accused of many things but promoting economic efficiency is not a charge often made against (or a claim made on behalf of) businessmen in a competitive economic system. We shall see, however, that the operation of competitive markets leads to economic efficiency. Efficiency, like profit maximization, need not be the explicit goal of business firms for it to be the outcome. Of course, firms do have an incentive to be technically efficient (as they would view efficiency) as part of their quest for profits. Overall economic efficiency is a somewhat more complex notion, as we shall see.

Economic Efficiency

Consistent with its concern with the problem of the scarcity of resources relative to wants, economic theory concentrates attention upon the process of the allocation of available resources. The logic of this emphasis is that "if resources are too limited to supply all the commodities desired, it behooves an economic system to employ its resources in the best ways possible, to satisfy their material needs to the greatest extent possible." The notion of economic efficiency is closely related to this most fundamental tenet of economic reasoning. Economic efficiency can be illustrated in the following fashion:
Economic efficiency is said to exist if there is no rearrangement of resources which can make anyone better off without making someone else worse off. To understand the meaning of this definition and to see that it is almost a restatement of those principles near and dear to the hearts of economists, consider the example in the table below.

<table>
<thead>
<tr>
<th>PERSON A</th>
<th>PERSON B</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q. Good X</td>
<td>Q. Good Y</td>
<td>Good X</td>
</tr>
<tr>
<td>1) 10</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2) 8</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>3) 6</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

Figures are given for three alternative distributions of a given total amount of resources. We shall assume that our individuals, A and B, can tell us when a change in the amounts of goods X and Y they have affects their well-being, i.e. whether a change makes them better off, worse off, or leaves their level of "satisfaction" the same. The initial situation (distribution of X and Y between A and B) is shown in the first line. Note that the total amounts of X and Y available to A and B together remain the same throughout the three cases. This is analogous to the state of limited resources for an economic system.

Suppose that after appraising their respective levels of satisfaction in the first arrangement of goods X and Y we
ask A and B to consider a re-arrangement such as is depicted in the second line of the table. In effect we have asked how they would be affected by a trade whereby A gives up 2X for 1Y and B receives the 2X in exchange for the 1Y. It is entirely possible that either or both A and B would consider this situation preferrable to the initial distribution of X and Y without the other feeling himself to be worse off. If this is the case, we declare that the initial distribution is inefficient, i.e. a re-arrangement of existing resources could make someone better off without making someone else worse off.

This line of reasoning may strike you as "much ado about nothing." After all, if such a trade has the possibility of making someone better off, would it not be completed without any complicated economic analysis? Indeed, economic theory predicts such "trades" will be consumated, providing certain facilitating conditions are met, particularly competition among buyers and sellers.

To continue with the example, suppose that the re-arrangement of X and Y between A and B depicted in line 3 is considered and that either or both of our parties declares that he would be worse off than under the previous situation in line 2. We would then state that this arrangement is less efficient than the previous one and that normally such a "trade" would not be made voluntarily. Note that one of the parties might consider such a trade as favorable to his interests but that our principle precludes us from comparing
one's gain to the other's loss. Economic efficiency statements are generally of this kind, comparing arrangements that are demonstrably better or worse without interpersonal comparisons. This conservative stance does not, of course, preclude individual economists or non-economists from making such comparisons when, for instance, they consider the merits of redistributing incomes between individuals. There are many such decisions we make without the "aid" of economic theory and since they involve changes in individuals' welfare, they can become important and controversial issues. Having explained the notion of economic efficiency, let us see its relation to competition in markets.

**Competition and Economic Efficiency**

You will recall that the equilibrium condition in competitive markets was production at an output where $P = MC$ for all firms. The price is a measure of what the last unit produced is "worth" to buyers. The last unit of the good must be worth its price for buyers to choose to make its purchase in the quantities they do. The marginal cost, $MC$, is the cost to producers of the resources necessary for production of the last unit of output. It is our next task to demonstrate that this condition where $P = MC$ for all firms is equivalent to the condition of economic efficiency (as just described) in production.

The economically efficient arrangement of output among firms is that situation which maximizes the output for a given
amount of resources. You should see that this condition is analogous to the condition of economic efficiency in the previous "exchange" examples. We focus upon two representative, identical competitive firms who are to produce a given amount of output, at a given price, between them. Consider figure 33 in which their contributions to the given total output \( X \) are different.

The initial total cost of producing some particular amount \( \bar{X} \) is the sum of the costs of producing \( X_{1A} \) by firm 1 and \( X_{2A} \) by firm 2. \( X_{1A} \) plus \( X_{2A} \) equals \( \bar{X} \), the particular amount to be produced for this entire example. These costs are equal to the areas under the two marginal cost curves at the points \( X_{1A} \) and \( X_{2A} \) respectively. Consider a reallocation of output which while keeping total production equal to \( \bar{X} \), reduces the output of firm 1 by one unit to \( X_{1B} \) but "recovers" the unit by expanding output to \( X_{2B} \). Does this reallocation change total cost? Indeed it does since the marginal cost of the unit in question was higher when it was produced by firm 1 than after firm 2 contributed it to the given total \( \bar{X} \). The first arrangement of output was less efficient than the second since a reduction in the cost of the given output \( \bar{X} \) is equivalent to an increase in output from a given amount of resources. Equivalently, if the same amount of total cost was incurred as in the original arrangement, total output would be somewhat greater. This is a consequence of the fact
Figure 33
The Efficient Allocation of Output Between Two Firms
that MC of output differed between the two firms and the re-
arrangement reduced the total cost of producing the given
output. The differences in MC appear as differences in heights
of the respective curves.

What is true of the initial re-arrangement is true of
any situation where the MC of output differed between 2 or
more firms. Efficiency would be attained only when the MC was
the same for all such firms. We have already seen that the
competitive equilibrium is one where $P = MC$ for all firms,
i.e. the MC for all firms is the same. In this way, competitive
markets ensure economic efficiency. Such a result is very
significant since any economic system, whether it is based
upon free markets, "planning," or whatever must cope with the
fundamental problem of scarcity. The fact that competitive
markets solve the efficiency problem so well has been a major
source of support for the free market arrangement as an eco-
nomic system.

A number of planned economies have paid the competitive
market system a great compliment in recent years if we believe
that imitation is the most sincere form of flattery. The U.S.S.
and other eastern bloc economies have experimented with numerous
aspects of the competitive market mechanism in an attempt,
by their own admission, to increase economic efficiency. These
experiments which include decentralization and greater autonomy
of decision-making and increased importance of "profitability"
have been interpreted as refinements in the economic planning
process by observers both in and outside these economies. It is not unreasonable to interpret these reforms as groping toward an output equilibrium similar to that of competitive markets, i.e. the $P = MC$ output. Of course, tentative and piecemeal adoption of elements of the market system do not "prove" the overall superiority of competitive markets. This development is suggestive, however, of the importance of the market mechanism in the attainment of economic efficiency.
Chapter 12. Competition and Discrimination I

INTRODUCTION

Some of the most "relevant" questions in economics concern discrimination. Most of the forms of discrimination which have been of increasing controversy in recent years are economic in nature—the effects of discrimination upon wages and incomes, upon prices paid for goods and services, upon opportunities for housing and education. We shall investigate some aspects of these kinds of economic discrimination. In particular, we shall be concerned with the interrelationship between the economic system and discrimination. We shall focus primarily on kinds of discrimination existing in competitive markets.

Racial Discrimination in Housing

Casual observation of housing patterns in most American communities reveals that blacks and whites are usually not randomly distributed within an area but, rather, that blacks and whites tend to be concentrated by place of residence within their own separate enclaves. While absolute racial segregation is not the rule, most neighborhoods are predominantly "black" or "white" and relatively few areas sustain significant degrees of racial integration for long periods. Such housing patterns are probably explained by differences in tastes and incomes to a large degree. Still, it seems undeniable that racial discrimination plays a major role in establishing housing patterns.

Suppose that whites prefer living among other whites to living among blacks. We shall focus upon this "taste for dis-
crimination" by whites not because blacks might not have similar tastes but because whites account for a very large proportion of the population and an even larger proportion of income. The sheer size of the white community in its economic impact is large enough to determine the "overall amount" of racial discrimination almost independently of black attitudes. Exactly why whites have such tastes or prejudices is seldom explicitly explored by economists but rather taken as datum. If whites would be willing to pay more for a given good or service when consumption of the commodity can be undertaken without contact with blacks than if such contact was possible, a taste of discrimination is said to exist. For individual consumers with such tastes, the demand curve for "segregated" goods lies farther to the right than for "integrated" goods. Whether the demand for "whites only" goods exceeds the demand for "integrated" goods depends upon the strength of these tastes relative to the demand of blacks for integrated and segregated goods. This analysis and that which follows does not imply that such tastes are good or bad but only traces of the consequences predicted by economic theory.

In figure 34 we consider the market for housing in a white neighborhood. The demand curves are that of the builder of a five house development in the situations where the development is restricted to whites only \( (D_1) \) and unrestricted by race \( (D_2) \). Note immediately that \( D_1 \) exceeds \( D_2 \) implying that the reduction in demand due to whites' tastes for discrimination exceeds any
Figure 34
The Demand for Segregated and Integrated Housing
increase in demand for integrated housing due to blacks desiring to buy in this market. If this were not the case, residential segregation in this market would be unlikely because racial integration would be more profitable.

The builder of the houses faces the same construction costs regardless of who buys them. Let us also assume that his future costs of operation will be unaffected by the race of the purchaser. On the basis of costs alone, the firm does not care who buys the house as long as his money is green. But because buyers care, i.e. the demand is different for the two "different" goods, sellers will be forced to take heed of the race of potential buyers and behave as if they (the builders) themselves have the tastes which their behavior only "reflects."

This builder must compete with other builders capable of producing very similar houses at similar costs. Suppose he decided to sell his houses "integrated" and let others "pander to buyers' prejudices." This noble builder's ideals will be costly to him, given the market's tastes for segregated housing. Other builders will be able to provide "more" homes for the same price (given buyers' tastes) or offer the "as good" for less money. This builder's good deeds can last only as long as he can continue to absorb losses on every home he completes. Eventually, he will abandon building similar projects, voluntarily or involuntarily. Given buyers' tastes, selling integrated housing is simply a losing proposition. An individual
builder, no matter how well intentioned, can have little effect on the pattern of residential segregation.

Proposed Solutions to Residential Segregation

Many people, perhaps most, would find the pattern of segregated housing described above as undesirable and propose ways to rectify the situation. In considering each of these "solutions," you should bear in mind that the costs of discrimination result from the prevailing tastes for housing and that these proposed solutions differ mainly in deciding who should bear the costs. As long as the tastes remain what they are, the costs cannot be "legislated away" anymore than the risk arising from uncertainty is eliminated by the purchase of insurance. In both cases, the cost is shifted, not eliminated.

One solution would be for builders to charge blacks a price high enough to compensate the builders for the expected decline in the market value of their houses which resulted from integration. For example, if a builder anticipated that integrating his 5-house development would reduce the market value of each house by $5,000 to $45,000, he would be willing to sell one of the houses to a black if the black would pay $70,000 for the fifth house. The total revenue of the builder would then be the same ($250,000) whether the project was integrated or not. This solution is unlikely to appeal to one's sense of fair play, partly because the element of discrimination is so blatant, and because the cost is borne en-
tirely by the one against whom the discrimination is directed.

A second solution would be to impose all the costs of discrimination upon the builders of homes. We have seen that individual builders cannot afford to "go it alone" as promoters of integrated housing. Why not pass a law requiring that no builder can discriminate, i.e. to favor one race of buyer over another? This would seem to eliminate the competitive drive to discriminate since builders no longer have to compete with each other by offering segregated housing. By definition, the law would now prohibit such competition.

The effects of such a law upon all builders can be demonstrated. The initial situation in which buyers preferred segregated housing and builders supplied this housing is shown as \( S_1 \) and \( D_1 \) in figure 35. Since home building is a competitive market, at the equilibrium price \( P_1 \), firms earn no more than normal returns. Any unusual profit would be competed away by the entry of new firms in the market.

Assuming that the law against discrimination in housing was enforced, the demand curve for housing would shift to the left since buyers preferred the initial kind of housing (segregated). In the short run, some firms would incur losses since they now must sell housing in a form which is less valuable to buyers even though the costs of construction remain unchanged. The exit of unprofitable firms would cause the supply curve to shift to the left. When a sufficient number
Figure 35
The Effects of Adequately Enforced Fair Housing Laws
of firms have left the industry, the remaining firms will again earn a normal rate of return at $P_2$. Note that the equilibrium quantity $Q_2$ is lower than the initial quantity $Q_1$.

Builders bear the costs of discrimination during the transition period from market output $Q_1$ to $Q_2$. While some firms are forced out of business, it appears that if all builders were prevented from discriminating, that the costs of discrimination, along with discrimination itself, can be quickly legislated away. Of course, owners of firms that are forced out of this field suffer a loss in their wealth. However, the economic cost of this adjustment to the new law is borne by everyone in the form of a less desirable allocation of resources as well as those who own resources especially suited for the housing industry. The initial output $Q_1$ represented the desired quantity of housing in the market, i.e. the best use of the resources required. Changing the overall composition of output by reducing the output of housing and freeing resources for other uses clearly reduces the overall material well-being attainable from a given stock of resources.

It is possible that an overall social well-being might be increased by such a law, if it were successfully enforced. In other words, the gains from integration per se might exceed the losses from resource misallocation for the society as a whole and the losses suffered by individual firms. As always, there are "winners" and "losers" to such a rearrangement—those that gain from integration of residential housing and
those that feel themselves better off when housing remains segregated. The distribution of the costs of discrimination under this plan probably would strike most as more fair than the previous arrangement where the costs were borne entirely by those subject to discrimination. The costs under this second proposal weigh heavily upon the producers of housing and those with tastes for discrimination. An optimist might suggest that the latter costs would tend to disappear once integration was established and tastes for discrimination diminished as "people got to know each other." Such speculation is clearly beyond the realm of economic theory.

Unfortunately, the legally imposed solution above embodied in countless "fair housing" laws has not been conspicuously successful in promoting residential integration. Patterns of residence have changed only very slightly and it is far from clear that such statutes changed patterns from what they would have been otherwise.

Why have fair housing laws been ineffective? This failure might be due to (1) the unanticipated difficulty and costs of enforcement (legal obstacles, public resistance to integration), (2) laxity on the part of elected and appointed officials, (3) the lack of a sincere commitment to integrate. Whatever the causes of the failure of fair housing, the actual economic effects have been different than would have resulted if these statutes had succeeded in their stated objective.
The costs and benefits of integration, per se, have been largely unrealized since the racial pattern of residence has been so little affected by these laws. The existence of these statutes and their occasional effectiveness has added a cost to the business of home construction and other forms of supplying residential housing. Firms now have to bear the risk that such statutes might force them to bear some of the costs of forced integration. This increase in cost shifts the supply of these commodities to the left, resulting in the same kind of misallocation we previously discussed, without the benefits of successful integration. The magnitude of the shift would be less than the previous case, but the gains associated with this increased cost are very small.

Since the legislative approach to fair housing has been unsuccessful, other approaches might be considered. We shall briefly consider a number of similar proposals which differ only in detail. These proposals are worth considering if a community really wanted residential integration (i.e. society is willing to bear the costs) but had failed to achieve it through the legislative route only because of some problems unique to the means chosen (the fair housing laws). If integration is not desired, consideration of the following proposals amounts to an academic exercise and nothing more. One important non-economic consideration in the following proposals is that they all involve a much greater participation of government in the process of the determination of housing patterns.
To implement our original solution that blacks pay a price high enough to compensate builders for the expected reduction in market value of their houses, tax funds could be raised to cover the extra cost due to discrimination. (It appears that private groups sometimes make such arrangements to promote integration). The burden would be shifted to taxpayers as a whole and the dollar cost of discrimination is made very explicit. The "discrimination premium" on houses could be expected to be about equal to the expected cost (reduced market value) builders incurred in selling their homes integrated. If the premium were higher, competition for black buyers by builders would reduce the premium to such a level. The cost of this program is borne by taxpayers together in proportion to whatever criterion is used for the tax funding. Virtually, everyone would share in the costs of discrimination and the allocation of resources to housing would be correct. Objections to this proposal as a "special favor" to blacks who prefer integrated housing could be expected. Partly these objections reflect a desire to 1) shift the costs of discrimination to others or 2) avoid the costs by maintaining residential segregation or 3) an admission that the costs of integration are not worth the benefits. Undoubtedly, some would consider such a plan unfair discrimination in favor of blacks (who benefit) and against themselves (who they see as bearing the costs).
Similarly, direct government subsidy of builders who sell integrated housing or government production or integrated housing could effectively change residential racial patterns. Again, the cost of accomplishing these proposals would be quite clear (the amount of the subsidy or loss on sales that would be expected) and the cost would be borne by society as a whole. Resource allocation might be unaffected by the first form (subsidy) although judging from experience with most government programs, an over-allocation of resources to housing seems likely with this version.

Finally, a government tax policy might be adopted to change the income distribution in favor of discriminated groups to compensate them for the costs of discrimination they currently bear (special, lower rates for blacks and women, for example). This last proposal, as well as the former ones, seem politically impossible now or in the foreseeable future. As previously noted, such programs would be seen by many as favoritism rather than a sharing of the costs of discrimination. Should such costs be shared? How? The reader must decide for himself. We can only conclude that as long as a taste for discrimination persists, the costs of discrimination must fall upon some people. Rearrangements through laws may shift these costs and influence resource allocation but the problem, i.e. the costs of discrimination, remain. That the reader who is sympathetic to integration should feel discouraged after this chapter is both un-
derstandable and appropriate. None of the "solutions" pro-
posed seem both practical and equitable. We have only dis-
cussed some reasons why the problem seems to intractable.
Chapter 13. Competition and Discrimination II

INTRODUCTION

In the previous chapter we considered some aspects of and proposed solutions to the problem of discrimination in housing. Now we shall consider another aspect of economic discrimination, discrimination in labor markets. We shall focus especially on the problems faced by women and blacks. Again, we shall see that the role of the firm in competitive markets is basically a passive one in regard to discrimination.

Popular treatment of labor market discrimination against blacks and women tend to exaggerate the extent of the phenomenon but economists are in agreement that a substantial amount of such discrimination does occur. In measuring discrimination per se, economists attempt to "hold constant" other variables such as education and experience that affect wages, incomes, and other aspects of labor market experience. After accounting for the influence of these variables on wages, etc. any unexplained difference between men and women (or between whites and blacks) is considered an estimate of the amount of discrimination. Many, perhaps most, popular treatments of discrimination fail to distinguish between the effects of these variables and discrimination, per se. Such estimates of discrimination as are obtained by comparing the average salary of women and men (or that of whites and blacks) therefore overestimate the direct effects of discrimination.
Economists' best estimates of the effect of discrimination upon the wages of blacks range from about 15 to 25%. This range should be considered as a crude estimate of the extent of labor market discrimination. Furthermore, discrimination probably contributes indirectly (as opposed to the direct 15-25% effect above) to lower wages for blacks and women by discouraging their accumulation of education and experience.

You will note a fundamental similarity of our treatment of the labor market discrimination manifest against blacks and women and the previous chapter's treatment of residential segregation. All these cases involve tastes for discrimination and immediately imply costs to be borne. As always, we shall focus upon how rearrangements to "improve" these situations distribute costs and benefits in competitive markets.

**Labor Market Discrimination Against Women**

Some of the labor market discrimination against women reflects the tastes of the firm's customers and employees. Such preferences reflect traditions, "prejudices," and other forces beyond the economist's field of inquiry. Economists accept such tastes as given when buyers behave as if the price of a good is greater if contact with women is involved or if employees behave as if their wage is "worth less" if work requires contact with female employees. Such tastes are probably not as important in explaining the overall labor market discrimination against women as they are in explaining racial
discrimination. A different factor explains observed discrimination against women in labor markets.

**Why Firms Usually Prefer Men**

Suppose we consider a job for which men and women are equally qualified and for which customers and employees are totally indifferent as to whether the job is filled by a man or a woman. If these were the only factors which affected the costs and benefits a firm experienced in establishing and maintaining its labor force, we would have a difficult time explaining why firms preferred men or women. We would have to answer that the firm had its own individual tastes for discrimination but we have asserted all along that the firms' own tastes have little effect.

Consider, however, some of the costs firms incur in conjunction with hiring employees. Firms make expenditures to recruit, screen, and train employees. These expenditures take many forms and are more important to some firms than others. But no firm can avoid them entirely and the more significant these costs are to the firm, the greater its concern with "recovering" these expenses, i.e. finding the corresponding benefits that make hiring employees worthwhile. Firms can recover these expenses by paying employees a wage less than their productivity in those periods in the future after these "investments" in employees are incurred. The firm would only undertake these expenses if it could regain them somehow.
The idea that a firm seeks to pay its employees less than their productivity may strike you as a little unethical. Remember, however, that during their training period, employees are paid salaries which generally exceed the value of their output. We could imagine an alternative situation where employees (instead of the firm) paid for their own training and other expenses by borrowing money. Competition for such trained employees would bid their wage to a level where the employees would recover the costs of their own investment. In fact, we do observe that firms offer higher salaries for employees with experience and training. Still, firms generally have to give even experienced recruits some training and incur other expenses if only to familiarize them with the firm itself and to find qualified employees in the first place. So, in fact, all firms have some such costs to recover. And again, we can rely on competition between employers for employees to prevent the recovery of these expenses from rising to levels that "exploit" employees unfairly. Like other kinds of investments, investments in employees (recruitments, training, etc.) will tend to pay only "normal" returns.

Clearly, a firm must consider the period over which it can expect to recover these investments in its labor force. All other things being equal, the longer the firm expects to retain an employee, the smaller will be the differential between the wage and value of output per employee required to adequately cover the costs of training. Firms usually seek to determine
factors which might affect the time period any prospective employee would remain with the firm and may look favorably upon certain personal characteristics it believes might be related to long service (age, marital status, etc.) Of course, firms cannot always predict the tenure of individual employees but they will find worthwhile to attempt to forecast the average duration of employment. Since firms usually have a number of employees for whom costs must be recovered, averages are generally sufficient information.

Figure 36 summarizes the entire problem we have been discussing. The $W$ function describes the wage the firm anticipates paying the average employee. For simplicity, we assume that the average cost of training and other expenses related to adding employees to the firm's labor force is equal to $I$ per week for the time period up to $t_1$. The entire cost is thus the sum of $I$ (direct investment cost) and the wages paid in excess of productivity. These costs are shown by the darkened, jagged-shaped area in figure 36. We assume for additional simplicity that the expected value of employee output is zero until $t_1$ and positive thereafter as described by the function $V_0$ (value of output). Like the wage function, $V_0$ is expected to rise over the time of employment.

Over the time that the cost per employee as indicated by the $W + I$ function exceeds the $V_0$ function (up to $t_2$) the firm is investing an amount indicated by the area $(A)$. For the firm
Figure 36
Profitable Investment in Employees by Firms
to willingly pay the indicated wages and incur the indicated
direct training expenses it must recover an amount equal in
value to this sum. The area (B) where VO exceeds W + I rep­
sents the time period for which these costs can be recovered.
(In comparing present and future sums, we have not explicitly
taken into account the fact that future dollars are worth
less than present ones. Assume that such an adjustment has been
made. The firm requires that the area (B) be equal to (A)
to justify its expenditures. Clearly, the size of (A) depends
upon the time period t₂. This is the time period in which the
firm expects its employees, on the average, to leave the firm.

If firms believe that male employees will, on the average,
remain with the firm for a longer period in which to recoup
these costs, they will prefer men to women as employees. Firms
reveal this preference by "discriminating" between male and
female employees by

(1) paying men higher salaries for women when their quali­
fications appear to be identical, or

(2) maintaining the same wage structure for men and women,
but requiring higher qualification standards for women
than men.

We put the word "discriminating" in quotes above because
many might not consider such behavior by firms to be discrimi­
natory as the term is usually applied. For instance, if the
firm believes that the average tenure of female employees
would be \( t^*_4 \), it would not want to hire them at wage \( W_0 \) and incur the expenses shown unless the applicants were "unusual" on some grounds which might lead the firm to expect to recover its costs. Unless such grounds exist, the firm would only be willing to hire women (or any other employees with an expected tenure of \( t^*_4 \)) at a \( W_1 \) less than \( W_0 \) as shown in figure 36. If the firm did not make such a distinction between its male and female employees, it would find itself incurring losses over time and would, hence, not be able to survive in a competitive market.

Since the time period is shorter, the firm must expect to recover its costs faster so that the new area (A) equals the new (B). The firm is not "prejudiced" as we usually use the term but is acting on the basis of expected relative costs and benefits. The difference between \( W_0 \) and \( W_1 \), would not exceed the amount justified by differences in the expectedly shorter time period for long. If such a difference arose, i.e. women were "exploited," competition among firms for female employees would bid the wages up to level where returns were once again "normal."

**Discrimination Against Women and the Law**

In recent years, agitation has been mounting to reduce and eliminate labor market discrimination against women. In most present and proposed legislation, the behavior of firms just described, viz. a preference for male employees based upon ex-
pected average tenure would be considered illegal discrimination. Courts have recently held that testing devices used by firms to better predict employee tenure are not legitimate criteria for establishing differences in pay because they did not directly measure employee productivity.

If the legislation preventing the kind of labor market discrimination we have discussed could be effectively enforced, certain costs and benefits would be distributed among firms and male and female employees.

Suppose such "equal pay" legislation was effectively enforced. Consider the situation as depicted in figure 37. If the firm was allowed to "discriminate," its demand curve for male applicants of a given quality (DM) would exceed that for women similarly qualified (DW). But the law would force firms to act as if their demand for labor was D regardless of the sex of the applicant. Whereas before the firm would be willing to pay male workers more than women for any conditions of supply, the firm is now forced to pay all the same. Since D lies between DM and DW, wages for men and women would fall and rise, respectively. A period of adjustment would ensue in which some firms would probably incur losses and some unusual profits because of differences in the manner in which the law increased or decreased their costs relative to benefits. Firms with many female employees or ones where the difference in expected tenure were most significant would be adversely affected. After a period of wage adjustment and
Figure 37
The Effect of Effectively Enforced Fair Employment Laws
possible reallocation of resources among firms (towards those whose costs have fallen due to the change), firms in competitive markets would return to normal profits on investment in employees. The costs of this rearrangement fall mainly upon men, the benefits mainly upon women.

Of course, such laws are not the easiest statutes to enforce. As always, when firms have a profit incentive to avoid the spirit of the law, compliance is made more difficult. A law requiring "equal pay for equal work," i.e. one which prevents the firms from "discriminating" in terms of lower wages is likely to induce firms to discriminate in hiring itself. For example, firms may reduce the number of recruiting effort for female employees. Women would find it more difficult to find employment under such a law. Those women who did find jobs might have higher wages than otherwise. Whether women as a whole would benefit under this rearrangement is unclear. Male employees' change in welfare is also ambiguous. To the extent that some labor resources would spend more time unemployed or that firms' costs would rise, the value of total output would diminish.

It is widely alleged that women's behavior in labor markets has changed, i.e. that tenure and labor force attachment have risen because of changes in tastes, "traditions," family size and rearing practices, etc. If this is indeed the case, one would expect the wages of women relative to men to rise in competitive markets since at the "old" wage differential, women
have become profitable "bargains" for employers. A reduced cost of day-care centers would probably have a similar effect by allowing women's labor force behavior to change. Competitive forces rather than the force of law would accomplish such a change in relative wages.

**Racial Discrimination in Labor Markets**

Much of the economics of racial discrimination in labor markets should be apparent to you from previous treatments of residential discrimination and discrimination against women in labor markets. Economists estimate that a significant amount of the wage-income differential between blacks and whites can be accounted on "residual" grounds. Measurable differences in the levels of education, experience, etc. explain only part of the differences in wages, the rest is attributable to discrimination. Rough estimates of the overall "discrimination premium" in labor markets range from about 15-25% although the effect of racial discrimination upon wages and income varies among different black groups. For example, discrimination seems to be greater the higher the skill-education level and greater against black men than black women. Furthermore, the effects of discrimination are cumulative and probably reduce human capital accumulations and make for an inadequate environment for labor market success.

Discrimination against blacks in labor markets, as in housing markets, generally reflects preferences or prejudices
of customers and/or employees rather than the tastes of firms themselves. In competitive markets, employers have a strong incentive to "get around" such tastes as long as race wage differentials persist which tend to reduce such differentials. Where such tastes are absent, an employer with prejudices of his own will find it more costly to exercise them than to be "color blind." But in markets where such tastes are widespread, an employer is forced to reflect these prejudices in his hiring policies. It is easy to conclude that the employer is the one responsible, after all, he establishes the terms of employment. However, it is competitive forces which make him "pander" to prejudices. Like an idealistic, unprejudiced home builder, an employer who ignores the tastes of his customers or employees increases the likelihood of non-survival in competitive markets.

An obvious "solution" to the problem of racial discrimination is "fair employment" legislation. If firms can be prevented from "competing" through discrimination, such behavior would disappear, proponents argue.

If fair employment laws are successfully enforced, the results are virtually the same as when "equal pay" laws for women are carried out. Wage rates would converge for blacks and whites with comparable capacities. Similarly, if the laws simply prevent employers from discriminating in terms of difference in pay, they are likely to do so in qualifications for employment. Black employees are less valuable than "comparable" whites because of prejudices just as women employees are less
valuable than men due to differences in expected tenure. In both cases, firms have incentives to "get around" the law because it prevents them from doing what comes naturally in a competitive market, i.e. employing resources such that marginal costs just equal marginal benefits.

Exactly how well such laws are enforced is a crucial and, as yet, not fully answered question. One piece of evidence strongly suggests that such laws have not had a great deal of effect. The ratio of black to white incomes was lower from 1952 through 1966 then it had been before 1952. It was during this period that most fair employment legislation was passed. What this suggests is that any effect of these laws was outweighed by business conditions. Unemployment was lower in 1952 than in the 1953-1966 period and the ratio of black to white incomes generally rises and falls in the opposite direction of the overall unemployment rate. After 1966, the ratio of black to white incomes exceeded the 1952 level, until the unemployment rate rose in the late 1960's. Thus, if we make some heroic assumptions about unchanging relative labor market characteristics of blacks and whites, the fair employment legislation has had little apparent effect, at least relative to the effect of overall business conditions. Actually, since the unemployment rate in 1966 was slightly higher than 1952 (about 3.6% vs. 2.8%) it might be argued that some success was achieved.
To the extent that such laws are effective, i.e. that the labor market discrimination which would be manifest because of customers and/or employees prejudices are suppressed, the direct beneficiaries will be blacks and the cost will be borne by whites. Since the costs are spread over a much larger group than the benefits, "per capita welfare changes" are much greater for blacks, on the average. However, since the effect of these laws seem minimal, it is not clear that, on the whole, whites or blacks have experienced net benefits or costs from such legislation.
INTRODUCTION

Awareness of pollution and other forms of environmental depreciation has grown in recent years to the point where it is one of our most discussed "crises." In this chapter we shall consider the insights economic theory can lend to understanding and, hopefully, improving this "mess" of a problem. Since most pollution and other environmental problems seem to arise in the processes of production and consumption of goods and services, it would seem that the theory could provide some answers. In particular, we shall be concerned with three broad questions about the environment:

(1) to what extent are these problems caused by competitive capitalism as a system?

(2) how does pollution, etc. affect resource allocation: who bears the benefits and costs of these acts?

(3) what are some solutions to these problems and how do these solutions distribute costs and benefits?

Pollution in Pittsburgh

To illustrate these aspects of the pollution problem, we will concentrate upon the city of Pittsburgh and its largest polluter, the steel industry. The production of steel inevitably produces a variety of undesirable by-products—dirt, noxious fumes, noise, congestion, etc.—although these economic
"bads"* are obtained in varying proportions along with steel itself, depending upon the methods of production used. Of course, we dan generally depend upon steel firms to choose whatever methods seem cheapest to them in their pursuit of profits. Let us see the consequences.

The unwanted by-products of steel-making can be regarded as part of the costs of production just as the raw materials, labor, and other costs associated with the product. There is a fundamental difference between the environmental costs and the "conventional" ones we usually associate with steel production, however. While the latter are borne by steel firms, the former are imposed upon "by-standers" forced to breath dirtier air, spend more to clean the many things "dirtied" including the water they drink, etc. Costs the firm has to pay are called INTERNAL costs, i.e. internal to the firm. Normally, the environment costs of production are EXTERNAL costs to the firm, i.e. the costs are imposed upon others.

We have seen that firms attempt to find those outputs which maximize the difference between their revenues and costs: The costs that matter to firms are the ones it must pay. Costs borne by others can be, indeed must be, ignored by the competitive firm bent on survival. Suppose one firm "got religion" about the costs of pollution it had been imposing

*An economic good is a commodity which increases one's utility or satisfaction so that more of it is preferred to less; an economic bad is exactly the opposite.
on others and decided to adopt a more expensive production method to reduce those costs. We can be sure the new "cleaner" method is more expensive since it would have been adopted previously if it had been cheaper in the first place. If this one firm persists in its "enlightened" behavior, its days are numbered unless it can persuade all other steel firms (in Pittsburgh and elsewhere) to also adopt the new higher cost production methods. If a competitive firm persists in using more costly methods than its rivals, regardless of how admirable its motives, its survival is put in jeopardy by less altruistic but lower-cost, lower-price firms.

How do those individuals and firms affected by the steel industry react to bearing some of the costs of steelmaking? As long as these costs do not seem too burdensome or too obvious, most Pittsburghers will probably be "philosophical" and not too disturbed by the depreciation of their environment. After all, the steel companies employ many citizens (who spend most of their salaries in Pittsburgh), so what is good for these firms is also good for many of these by-standers, or at least, so it will seem. At this stage, those concerned or indignant about the environmental costs the steel firms are imposing in conjunction with steel production are likely to be viewed with a mixture of skepticism and hostility. Sympathies will lie mainly with the steel producers whose interests seem more consistent with their own to most Pittsburghers.
In recent years, the attitude of many people toward the pollution originating in the processes of production and consumption has changed from one of indifference to alarm. The cost and awareness of pollution have mounted to the point where proposals to "do something about it" have become widespread. Firms and individuals are called upon to consider more fully the effects, i.e. the costs, of their actions upon others. The extent to which their behavior has changed in response to public outcries, occasional new laws, and even vigilante harrassment is unclear. Competitive pressures are unlikely to allow much real change to originate within the market. As long as the firms find it less costly to pollute (including these new pressures which must be considered additional costs to the firm) there is a presumption that competition will force firms to continue to impose environmental costs on bystanders.

**Efforts to Stop Pollution**

Individuals or groups interested in reducing or eliminating pollution, i.e. ending the costs imposed upon them, are likely to try several approaches to changing undesired behavior. We have already mentioned appeals to firms' better nature, civic spirit, altruism, etc. When firms respond at all at this stage, it is likely to be in good measure a "public relations" approach without much real change in behavior. Pressure on firms can get results, especially if political
clout is mastered, but by and large, asking firms to change their behavior, however forcefully done, is unlikely to succeed.

Another approach is to make it worth the firm's while to stop its pollution or whatever it is doing that imposes costs on others. If many suffer, they might be willing to "bribe" the firm into behaving differently, i.e. making it more profitable for the firm not to pollute. Two problems arise here, viz. 1) organizing all the sufferers, and 2) getting everyone to pay who would benefit (since any individual would benefit if the action were successful even if he did not contribute). While this solution is a theoretical possibility, in practice the identification and apportioning of the costs of the bribe covering all those affected by pollution in Pittsburgh would be very expensive. With such high costs in coordinating the bribe, the likelihood of ever undertaking the bribe is very small. And, of course, there is the additional question of whether it is fair for those who impose the costs (the polluters) to desist only when they are rewarded for doing so.

Another approach is for the offended parties to seek redress in courts of law. After all, if someone dumped a load of garbage on another's lawn, the offended party would be able to sue for damages. Unfortunately for the victims of pollution, dumping garbage in the air is not so easily punished. Unlike the lawn, the air one breathes belongs to no one, it is not private property. As laws are generally written, "damage" done to natural resources, like air, water,
that are not private property, are extremely hard to recover. Consequently, the polluter is usually relieved of his responsibility for these costs. Here, both the layman and economist concur, is a situation that calls for government action, in the form of a change in law. Before we look at what governments might do at this juncture, let us analyze the entire situation as it now stands, given the present legal framework.

Figure 38 depicts a steel producer's revenue and cost curves. The demand curve indicates the value of output on the margin—what an additional unit of output is worth to buyers. There are two cost curves viz. MPC--marginal private cost and MSC--marginal social cost. MPC includes only those costs internal to the firm such as labor, materials, etc. MSC includes all the costs of production, i.e. the firm's internal costs plus the external costs, such as pollution, which it imposes upon others. The vertical distance between the two curves represents the amount of external cost generated by the firm but borne by others, at all outputs.

The profit maximizing firm is forced by competition to ignore external costs and produce output $Q_3$ where $\text{MPC} = P_3$. From the standpoint of economic efficiency, however, output $Q_3$ uses too many resources in relation to the value of steel output. This is revealed by the fact that at $Q_3$ $\text{MSC} > P$. The appropriate output is $Q_2$ where $P_2 = \text{MSC}_2$. At higher rates of production like $Q_3$ the social marginal cost, including external costs, borne by society, exceeds the additional
Figure 38
The Appropriate Allocation of Production With Pollution as a Byproduct
value placed by consumers on steel. The ideal solution would be to change the firm's behavior so that it regards MSC as its own cost curve. If such a situation could be arranged, the appropriate allocation of resources would be obtained. Note that this would not eliminate pollution or other external costs but merely reduce them to a level where these costs (along with the internal costs) were just matched by the benefits of production, on the margin. Economists term this ideal arrangement "internalizing the external costs," i.e. making the firm responsible for the costs imposed upon others. A simpler device to bring this about would be to levy a tax in steel equal to the external cost at each level of output.

The objection that such a solution is unsatisfactory because all pollution is not eliminated is mistaken. Efficiency is attained at output $Q_2$ where $P_2 = MSC_2$. Assuming for simplicity that there exists no other way to produce steel, eliminating all the external cost would require eliminating all production. At a zero output, price would exceed marginal social cost. This means that a zero output of steel is economically inefficient. Society would improve its welfare by permitting steel production to increase until steel output equalled $Q_2$.

In practice, there are a number of problems associated with the economist's solution to pollution we have developed. Accurately measuring the costs of environmental depreciation is often difficult. Moreover, strong resistance to imposing
the costs of pollution upon the polluter can always be expected from polluting firms and others (employees, suppliers, customers, and sometimes neighbors) who view their interests as consistent with those of the polluter. Usually these groups are smaller than the population suffering the costs of pollution. Therefore, these defendants of the status quo are likely to be better organized in their resistance. Furthermore, political action to reduce pollution is usually crude in comparison with the fine distinctions between marginal costs and benefits, internal and external effects, etc. Rather, the adopted solutions, when the status quo can be overturned are not necessarily improvements in terms of their effects upon resource allocation.
INTRODUCTION

Thus far we have examined only competitive markets. The author would contend that most markets in the American economy are essentially competitive and that the competitive economic theory already developed is generally the most reliable source of predictions about economic events. In this chapter we shall consider several different kinds of objections and criticisms of the emphasis so far presented. First, however, we begin by reviewing some key concepts of the competitive theory.

A Very Brief Review

In figure 39 is depicted the simple supply and demand analysis of competitive economic theory. Pictured are the market conditions for a particular industry where the curves summarize the aggregate behavior of a large number of competing firms. The unfettered operation of market forces tends to establish the equilibrium values for price ($P_1$) and quantity ($Q_1$), given these particular conditions of supply and demand.

We have held that the free operation of competitive markets generally leads to economic efficiency, i.e. the optimal allocation of resources. This condition is satisfied when $P = MC$ in the various markets, a condition usually fulfilled by the competitive market equilibrium. Economic
Figure 39
Supply and Demand Revisited

Figure 40
Two Kinds of Demand Curves
efficiency is a means to an end, viz. the maximization of the value of output from a given stock of resources and technology.

One clear exception to the general rule that competitive markets lead to economic efficiency was provided in the previous chapter on pollution. Because of external costs imposed by polluters upon bystanders, the free market solution was not efficient. A second possible exception, the existence of discrimination in markets, is a somewhat less clear indictment of the competitive market's efficiency. We will consider this issue when we question the desirability of allowing the full expression of people's tastes in competitive markets. By and large, however, equilibrium in competitive markets corresponds to an economically and socially efficient allocation of resources.

General Objections to This Analysis

This text lays greater emphasis upon the importance of competitive markets than many introductory economics texts. The traditional approach is to treat markets as if they were divided into a number of types, not all of which are deemed competitive. One extreme form is often called "pure competition" which more or less corresponds to the model developed here. Furthermore, this competitive model is often alleged to be an ideal type only rarely observed in the American economy. Most markets are held to be less than competitive, i.e. located somewhere along a continuum which leads, ulti-
mately, to the opposite of competition, monopoly. How important are these non-competitive elements? Is the economy essentially competitive (as we have held thus far) or not?

Strictly speaking, very few sellers are "pure" competitors as depicted by the demand curve $D_1$ in figure 40. In the short run, firms usually do not have horizontal demand curves. This means that their sales would not go to zero at some price slightly above the one they presently charge. Similarly, a small reduction in price would not increase the quantity demanded "by an infinite amount." Rather, most firms face demand conditions more like $D_2$ at any given point in time. In this sense, most firms have some degree of "monopoly power" at any given point in time where monopoly power is taken to mean some degree of "insulation" from pure competition.

Does it follow, therefore, that the theory of competitive markets is of limited applicability? Even if noncompetitive elements are pervasive in the short run, most markets' operation might still be better explained and predicted by the competitive theory than any other. Whether or not markets are essentially competitive or not, i.e. whether the competitive theory is the best theory, depends crucially upon whether or not firms and individuals are free to enter and exit markets as they please. The entry and exit of firms responding to "unusual" profits makes any firm's demand curve more elastic, given time to get in and out of the industry. Entry and exit also shifts the industry supply curve, making it more elastic.
when time for such adjustments is allowed. Together, these effects bring a market closer to the competitive equilibrium, the longer the period of adjustment for entry and exit, providing that such an adjustment is forthcoming. Economists differ in how strong they believe these tendencies toward very elastic supply and demand curves (which approach the competitive or "purely competitive" ideal) for the economy as a whole. This is the basic issue of how competitive the economy really is, i.e. what model is most "realistic" and provides the best predictions. We believe that this condition holds for most markets, and that most markets' conditions, given time for adjustments, are therefore best explained by the competitive model.

**Specific Objections to This Analysis**

We shall consider two specific kinds of objections that are often raised to the competitive theoretical framework and show how they do not necessarily vitiate its structure and implications. To be sure, these objections to the competitive theory do not exhaust the lists of critics of the framework. Ultimately, the matter of the best theory must be settled upon the basis of empirical results, i.e. the reliability of predictions. Therefore, to resolve this issue to one's own satisfaction requires a more extensive investigation than can be undertaken at the introductory course level.
Price and Quality Variation in Markets

A first set of objections to the competitive theory relate to "the" market price. The simple theory postulates that there is only one price for a commodity in a given market. A casual glance reveals, however, that the "same" commodity is often sold in "nearby" locations at different prices. Another similar occurrence which is often not treated in the competitive model is the existence of several levels of quality for a given commodity sold, as opposed to the "unique" commodity of the simple theory.

The observation that prices vary among sellers of the same good in the same city, e.g. in "ghettoes" and college neighborhoods, is often asserted to be evidence of the "monopoly power" of the sellers with higher prices. Critics of the competitive theory (and sometimes residents of the high price neighborhood) charge that the sellers' powers derive from the lack of mobility of buyers.

Perhaps college students and ghetto residents do have less access to automobiles and other forms of transportation. This would not, in itself, allow sellers a degree of monopoly power. Potential sellers would quickly enter these markets if prices were high enough to allow a rate of profit above the competitive level. No one has alleged that sellers outside these areas suffer from a lack of mobility. To the extent that prices are higher in these areas seems adequately explained by differences in costs (rent, insurance, shoplifting, risk of burglary, robbery, etc.) relative to other areas.
Often what appears to be differences in price for the same good is actually explained by differences in supplementary services provided by different sellers. "Discount" operations usually provide less sales personnel, credit and other facilities than other sellers. If for these and other reasons their costs and therefore prices are lower than those of other sellers, this alone would hardly constitute good evidence that the high price sellers have some monopoly power. Most people would probably assume that the large volume discount operation is more profitable than these "monopolists." Probably the first discount operations were unusually profitable, while their competition reduced the profits of many other sellers below the competitive level. With time for the entry of new discount stores and the exit (or conversion) of other sellers, the profit rate for different types of retail stores probably approximates the competitive level. In competitive markets, innovations usually give way to imitation and unusual profits are transitory. Even if price differences among different kinds of sellers remain, there is a strong presumption that with free entry and exit, the competitive theory still provides the most reliable predictions and that remaining differences can be reconciled with the theory.

The "differences in quality" argument is sometimes alleged to be evidence of non-competitive markets. We do observe that for many goods a range of qualities is offered by sellers.
Some might allege that this is proof that sellers of different quality goods do not compete with each other or that sellers can vary the quality of commodities as a technique to limit competition from their rivals.

A little more thought about quality differences should convince you that they are not inconsistent with competitive markets. Differences in quality presumably imply differences in costs and each seller attempts to find that combination of quality and costs which maximizes his profit. This combination depends upon numerous variables such as buyers’ tastes and incomes, substitute goods, etc. Innovations and changes in these variables may give rise to temporary high or low profit levels but competition and time are dependable levelers. If some particular quality of good or service proves to be unusually profitable, firms producing other degrees of quality will have an incentive to switch to that quality which apparently suits buyers’ tastes best. "Entry" in this sense is probably even less restricted than entry into a market, per se. Therefore, it would be unusual for a seller to be able to maintain supra-normal profits solely because he produced commodities of some particular quality. Long run price differences due to differences in quality are not evidence of monopoly power if each seller is free to choose those price-quality combinations is expected to be most profitable. Such choices, combined with flows of firms into profitable areas of production, may
be expected to reveal the temporary nature of abnormal profits attributable to differences in quality among similar products. Therefore, it is not at all inconsistent with competitive theory to observe a range of different prices and qualities in the market for a commodity.

Another exception to the competitive market model is the allegation that goods of very high quality (e.g., very safe or very durable) are sometimes not produced, even though feasible, because producers have colluded to keep them off the market. For example, many people believe that sellers have not offered to buyers "safe enough" automobiles. The reaction has been a call for government regulation to force manufacturers to include some safety features as part of standard equipment which might not have been available at all previously. Apart from the effectiveness or propriety of such laws, the question arises as to exactly why these features were not provided by the free market, if indeed they were not even though buyers really wanted them.

All automobile manufacturers offer many of the same features (sometimes in different qualities) because customers want them. Cars usually have more than one door, a spare tire, headlights, and many other features not because they are absolutely essential but because they are "extras" people are willing to pay for. They are extras in the sense that they cannot be produced at a zero cost. (Perusal of ads for early automobiles gives ample evidence that such items as spare
tires have not always been standard equipment, whereas tool sets were standard on some models in early days.) A manufacturer who did not provide these features, even if he reduced the cost correspondingly, would probably not be able to earn a normal return. Are safety features somehow different? Cars already have a number of safety features—brakes, padding, locks, etc.—that have been provided in different qualities. Why have not additional features (e.g. safety belts, air bags, etc.) been provided as "standard equipment?"

Rather than a conspiracy between manufacturers, competitive theory suggests that such features are not profitable to provide because buyers simply do not want to pay the required cost. A car built along the lines of a tank would probably be very safe, and perhaps fewer or less serious accidents would occur if cars' top speed were limited to 50 mph, but such cars are not produced because of buyer preferences for other kinds of cars. One may believe such tastes to be ill-informed, regret the fact that no manufacturer finds the market for tank-cars attractive enough to provide them to those who would buy it, and support laws to force people to buy safety equipment whether they want it or not. These attitudes reflect value judgments about which economics analysis as such has nothing to say. But the absence of products or product quality which one believes should be provided is hardly proof that its markets are not competitive. One's initial
feeling based on casual analysis may well be that producers are colluding to increase their profits at the expense of the general public; we are inclined, however, to echo Walt Kelly's "Pogo" in our search for the public's enemy in this case—"We have met the enemy, and he is us!"

Another version of this argument asserts that fantastic products are withheld from the market (a lifetime lightbulb or tire, etc.) because of a conspiracy among sellers to protect and preserve their markets. While existing sellers might have such an interest, it is difficult to explain why some outsider or maverick firm might not find the innovation profitable. We must concede that we cannot be categorically certain that the suppression of such innovations does not occur; but we suspect that often such alleged commodities really are not feasible or that they would not be commercially successful.

Criticism of the Competitive Model on the Grounds of the Existence of Advertising.

Many individuals and economists alike believe that advertising is antithetical to competitive markets. This belief probably arises among the general public, because successful advertising is associated with successful firms who have bested (in some sense "eliminated") their competitors. Economists who view advertising as a force reducing competition in markets probably think in terms of 1) "pure" competition and undifferentiated products, versus 2) non-competitive markets and differentiated products.
Whether or not advertising is consistent or inconsistent with competitive markets is an empirical question, despite plausible theorizing on the issue. Naturally, advertisers hope that their expenditures will help them prevail over their rivals, but does it follow that advertising therefore reduces the amount of competition rather than increasing it?

Advertising does not appear to reduce the degree of competition in markets and create monopoly-type profits, even though specific instances abound where successful advertising differentiates sellers in the minds of buyers. This should not be surprising for if advertising were the easy road to profits, what would prevent the competition from following suit and thereby compete profits away? A number of specific findings about advertising and various measures of competition would be usefully presented at this point.

1) Advertising intensity (% of sales spent in advertising) appears to be unrelated to consumer product market concentration (a measure degree of monopoly power.)

2) For a sample of heavily advertised consumer goods, the greater the intensity of advertising, the less stable the brand shares.

3) For successive periods studies in the 1950's and 1960's, changes in the degree of advertising intensity were unrelated to changes in consumer product market concentration.

Taken together, these findings, based upon the research of Lester Telser shed great doubt on the proposition that ad-
Economists who believe advertising to be incompatible with competition implicitly compare the actual state of affairs with a mythical world of "perfect" competition, sans advertising. They ignore the very important problem that knowledge is not a free good and that advertising, far from inhibiting competition, may well be a necessary condition for competition rather than antithetical to it. But, again, we emphasize that the matter is an empirical rather than theoretical issue that must be resolved with investigations like that which revealed the results noted above.

Another criticism of the competitive theory which arises from the existence of advertising questions emphasizes not so much the effect of advertising upon the degree of competition among sellers but its effect upon the optimal allocation of resources. This argument asserts that advertising changes consumers' tastes, and, therefore, the composition of output (presumably for the worse.) J. K. Galbraith argues, for instance, that advertising biases the choice between private and public goods in favor of the former. However, since the production of public goods has grown faster in recent

*See "Advertising as a Form of Competition" by Telser in The Journal of Business, vol. 1, no. 2, pp. 166-173, University of Chicago Press, 1968. This article is contained in several "readings" books designed for introductory economics courses.

decades than that of private goods this particular alleged bias effect hardly seems to be cause for great alarm, at least in overall terms.

Even if advertising changes the composition of output, it is not clear that such changes are for the worse. Critics of advertising really seem to object to the human motives and drives to which much advertising is directed. It is often charged, for example, that advertising makes people more materialistic or that it creates the needs heavily advertised goods are alleged to satisfy. However, the logical content of such arguments appears to boil down to the allegation that such needs are less worthy than some other, more "basic" needs, which presumably antecede advertising. To paraphrase Galbraith—"if certain needs would not be perceived in the absence of advertising, they certainly must be less important than other 'spontaneous' needs." But what of the "needs" many kinds of which people learn through "education," or government-provided "information," rather than "advertising," which they come to desire—are such "acquired" tastes less valid, legitimate or important than "natural" ones? To us, it seems more appealing to think of many specific forms in which we satisfy "natural" needs that are the result of many kinds of learning, including the influence of advertising.

As you can sense, much of this issue remains in the realm of ethical evaluations, perhaps inevitably. We therefore
believe it is crucial to recognize that the issue of whether advertising really lowers welfare depends upon whether one approves of allowing consumers' desires, no matter how formed, to determine outputs. While in some cases, almost everyone would agree that advertising should be prevented from influencing allocation (deceptive advertising or advertising for "bad" products like heroin), we believe that the number of such is very small relative to the total amount of advertising.

An example which may illustrate our point of view is the advertising for men's and women's cologne. Typically, such ads make subtle and not-so-subtle appeals to consumers' sex drive, feelings of insecurity, etc. The ads seem to make vaguely extravagant claims on behalf on their products. Since such products are very intensively advertised, their prices reflect heavy advertising expenditures. However, the cosmetics market would appear to be quite competitive, and there is no presumption that cosmetics firms are unusually profitable. Telser found that the brand loyalty for toiletries and cosmetics was less than for less heavily advertised goods.

To say that such advertising distorts the allocation of resources unfavorably is to assert that tastes, or the basic human motives to which such advertising appeals, should not govern resource allocation. Actually, this issue is similar to the question we posed earlier, viz. whether or not the existence of discrimination in markets undermined the asser-
tion that competition allocates resources most efficiently. Both questions are resolved on the issue of whether or not primacy should be granted to consumers' tastes as a determinant of output.

**A Closing Note**

One argument that proponents of advertising often advance is that advertising subsidizes the consumption of various goods and services such as "free" T.V. and printed media. It should be clear from the discussion presented in this chapter, however, that advertisers must recoup these "subsidies" in the form of higher prices for advertised commodities. Advertisers are not indulging in charity but are seeking to maximize their profits. Of course, some viewers of media which carry advertising may experience a net gain from the existence of advertising. But, by and large, advertising is no exception to the economists' adage that "there is no free lunch," or the familiar refrain that "you get what you pay for."
Chapter 16. Monopoly—Markets Insulated from Competition

INTRODUCTION

In the previous chapter we noted a number of criticisms of the competitive economic model. While economists hold varying opinions about the degree to which forces discussed vitiate the applicability of the competitive model, there is general agreement that some markets cannot be described as competitive by almost any criteria. These markets, and of course there is debate over how many such markets there are and even which ones they are, are described as "monopolistic." Monopoly, or monopoly power, should be thought of as a matter of degree rather than an absolute characteristic of markets.

A Brief Review of Monopoly

When monopoly exists, competitive equilibrium conditions are precluded. The price and output are changed and it can no longer be asserted that the market tends to allocate resources efficiently. Consider figure 41.

The competitive equilibrium for such an industry would be at the price $P_c$ and the output $Q_c$. It can be asserted that allocation is optimal since each firm produces where $P = MC$. All this has been explained previously.

The effect of "monopolizing" this market is to raise the price charged to $P_m$ and reduce the output to $Q_m$. In this extreme case of a single seller (remember monopoly power
Figure 41
The Contrast Between the Competitive and Non-Competitive Equilibrium Positions
is a matter of degree) the industry supply and demand curves become identical to the firm's marginal cost and demand curves, respectively. If competition was only lessened, rather than eliminated entirely by a reduction in the number of sellers, each firm's demand curve would change from the horizontal demand of the competitive firm toward the extreme case of the single seller. In any event, each "monopolistic" firm has a downward sloping demand curve and its marginal revenue curve lies below the demand curve as shown in figure 42.

Each monopolistic firm would choose an output such as $Q_m$ and price $P_m$ at which to maximize its profits. Contrast this kind of equilibrium with the competitive one also shown. The reason the monopolist firm produces less and charges a higher price than the competitive firm is the difference in their respective marginal revenue curves. The competitive firm's MR curve is horizontal at the price $P_o$ so he maximizes profits by producing output $Q_o$. The monopolistic firm is too large to have no effect upon the market price. Therefore, the monopolist's MR lies below the demand curve and profit-maximizing output is therefore $Q_m$, for which the price charged is $P_m$. Units beyond $Q_m$ would add more to his costs than to revenues ($MC > MR$), and so would reduce profits. Both kinds of firms attempt to maximize profits, but differences in their MR curves lead to 2 distinct differences in their equilibrium price and output decisions.
**COMPETITOR**

1) produces where $P = MC$
2) earns only normal return

**MONOPOLIST**

1) produces where $P > MC$
2) may earn higher than normal profits

---

**Why Monopoly Is Bad**

The first difference between the competitive and the monopolistic outcomes implies that the monopolist causes resources to be misallocated since it would be optimal to produce where $P = MC$. Units left "unproduced" are worth more than the additional resources required to make them are worth in producing alternative goods, but because their $MC > MR$, the profit seeking monopolist would not knowingly carry production into this range. This misallocation of resources (the unused resources go elsewhere, presumably into more competitive markets) is the chief reason economists favor more competitive over less competitive markets, as a general rule. Of course, this analysis is too technical and "theoretical" to concern most people. But most people are opposed to most forms of monopoly.

Monopoly is opposed in most instances by most people not for its role in misallocating resources but because it is sense that monopolists tend to charge higher prices and earn high incomes. Since most people must compete in open markets, they usually resent sellers who, insulated from competition, can successfully raise their prices and profits. Of course,
not all monopolists are rich, even though the changes of becoming wealthy are enhanced by some kind of "protection" from competitors. Actually, people's judgment about the propriety of monopoly seems to be based upon how "deserving" they believe the monopolist is rather than a reflection of bitterness toward all monopolies. Labor unions are probably tolerated or accepted by most people, at least in principle, despite the fact that a union attempts to monopolize the supply of labor in a market. Furthermore, people seem to distinguish among different unions, generally being more "sympathetic" to the unions least successful in raising wages and incomes—contrast attitudes toward the Teamsters Union versus The United Farm Workers.

A Crucial Assumption

In comparing the competitive and monopolistic equilibrium positions we have made an implicit assumption which we shall now consider directly. We have assumed that the cost conditions of any firm remain the same, regardless of the market structure (degree of competition or monopoly). There are two important cases where this condition does not hold and we must reserve judgment as to whether a competitive or monopolistic market is more desirable under these circumstances.

Natural Monopolies

The first case refers to those markets where a firm's average cost curve declines continuously over a large range
of output as shown in figure 42. It is possible that this
decline in AC could continue beyond the output the entire
market is likely to absorb. If market forces are allowed
to operate, only one firm is likely to survive, but this mo-
nopoly outcome may be desirable.

The reason that this market tends toward monopoly is
that any seller who can expand his output, can underprice
his rivals, which eliminates them from the market, allows
him to further expand and reduce price, and so on. If the
entire market is incapable of consuming an output well beyond
Q₁, one firm is likely to eliminate its rivals and emerge as
the monopoly seller. This case, called a "natural monopoly,"
for obvious reasons, is an extreme form of "economies of
scale." This latter concept will be examined in the next
section in detail. At this point, suffice it to say that
the concept refers to the reduction in cost firms achieve
by increasing the quantity they produce. "Spreading the
overhead" as it is colloquially described, reduces average
cost, for some range of output, for most firms. Virtually
all firms experience rising costs eventually; if they did
not we would expect natural monopolies everywhere. As it
is, natural monopolies arise in a few markets where fixed
cost is very high relative to variable cost and AC declines
over ranges sufficient to absorb the entire market output.
Most of these cases occur in markets involving transporta-
Figure 42
Continuously Declining Marginal Cost
tion, communication, and power production (though, of course, even these markets are not invariably natural monopolies).

Governments have traditionally recognized that competition is not feasible and would probably be undesirable in markets served by natural monopolies. The only way competition could be maintained would be by limiting the size which any firm could reach. In the previous figure, suppose the entire market output would be $Q_1$ without government intervention, an amount that one firm could supply. Competition could be maintained by preventing firms from expanding beyond $Q_2$. This would "leave room" for several firms, each operating near the output limit. But note that the average cost of production and therefore price would be considerably higher because the cost reduction gained by expansion of output (the economies of scale) are foregone.

In recognition of these conditions, governments have made natural monopolies "regulated monopolies" or "public utilities" as they are often termed. The intention of this action is to secure the benefits of large-scale output for consumers while simultaneously avoiding the undesirable consequences of monopoly, per se. Typically, governments have established commissions which set the prices public utilities may charge so that profits may not rise above some "fair" rate of return.

Several problems have arisen in the implementation of this policy, sensible as it sounds. The most basic problem
is securing the benefits of large scale output. By restricting the amount of profit the legal monopolist can earn, the incentive the firm has for minimizing cost is sharply reduced. Of course, these firms do not admit that they are intentionally wasteful and inefficient (on the contrary, they spend considerable sums advertising their own virtues) and public officials ostensibly act as cost "watchdogs." Still, there is a presumption as well as some indirect evidence that such firms pursue cost efficiencies with something less than missionary zeal. As long as profits are limited by law, the firm and its employees can benefit from the monopoly position in more subtle ways. Extravagant salaries and padded expense accounts, discrimination on non-economic grounds, plush offices and well-manicured lawns are but a few of the ways such firms could and do share in the benefits of being part of a natural monopoly. The effect of regulation might be to make the enjoyment of the natural monopoly position turn toward more subtle forms. The effects of all these benefits is reflected in higher operating costs and correspondingly higher prices.

One proposed solution to this cost quandry is to sell the right (to license) to be the natural monopolist to the highest bidder, specifying the rates to be charged to consumers beforehand. Competition among would-be monopolists is likely to force the price of such a license up to the expected value of future "unusual" profits. The license fee
would be treated like tax revenue. The licensed monopolist would have every incentive to reduce costs which presumably would allow the stipulated price rates to be lower than currently prevail.

A second serious problem resulting from public regulation of natural monopoly has arisen when the establishment of a legal monopoly did more to create the monopoly position than to simply recognize the inevitable. The question of what markets are natural monopolies is not always obvious, especially over long periods of time when many conditions affecting the industry may change. It is clear that in a number of cases, regulation has served more to protect existing firms from potential competition or from competition among themselves, than to merely confer legal status to a natural monopoly. National regulation in the transportation and communication industries are areas in which this charge is frequently made. Such charges are open to debate (naturally they are denied by the regulators, but not always by the regulated) and should be examined on their individual merits. One technique is to compare the performance of intrastate (unregulated) and interstate (regulated) markets to estimate the effects of regulation upon costs, etc. It seems that there is considerable substance to such allegations.

Incentives for Innovation

We have seen that the potential benefits of economies of scale in monopolized markets is a complex issue. A second
possible cost advantage to monopolized markets is also somewhat controversial. This potential cost advantage of monopoly relates to the firm's incentive to innovate, to bear the risk and other costs associated with achieving "breakthroughs" - really new products, processes, etc.

One line of argument has it that monopoly firms have a greater incentive to innovate because they have a greater assurance of experiencing the benefits of successful attempts at breakthroughs. If a number of competitors stood ready to copy anything new, any firm would have less incentive to devote resources to finding innovations, so this reasoning continues. Others suggest that monopoly firms, insulated from competition, grow lazy and are less likely to innovate. A way in which public policy enters this issue is in the creation of patents and copyrights granted to innovators for limited periods. Hopefully, this policy is a happy compromise between granting innovators some monopoly incentive to "breakthrough" while ensuring that the benefits are eventually passed along to the general public.

Actually, whether or not firms in monopolistic markets are more innovative than their counterparts in competitive markets is another empirical question, albeit a difficult one to resolve. Some suggestive evidence is provided by observing the vast resources devoted to research and numerous innovations achieved by large firms in highly concentrated
markets such as DuPont (Chemicals), Xerox (Copying), and IBM (Computers).

**Sources of Monopoly Power**

While we have discussed monopolies at some length, it should be re-emphasized that monopoly power is a matter of degree. Competition connotes a number of sellers, each too small to affect the market price, while "pure monopoly" exists when the industry and the firm are coterminous. There is obviously a lot of ground in between, markets where sellers enjoy some monopoly power. In this section, we shall examine a number of conditions which would contribute to monopoly power.

1) **Economies of scale.** This phenomenon already noted under natural monopoly, is actually more likely to limit the number of sellers to a figure less than the competitive "many" but greater than the monopoly "one." The question is how many cost-efficient firms is the market (sales potential) large enough to accommodate. This number depends upon the size of output for efficient operation relative to the industry output. A number of industries, which tend to be stereotypes of American industry "Big Business" are dominated by a relatively small number of producers. There once were hundreds of automobile manufacturers but only a handful have survived. It may surprise you to find out that the degree of concentration of industrial output among a few firms has remained
about the same for American industries taken together for the 20th century. Evidently, the automobile and similar examples are balanced by other less dramatic reversals of concentration in other industries.

2) Private Collusion. Monopoly power may arise because of agreements not to compete among sellers and other methods sellers have used to lessen competitive behavior. The contrivances that might be mentioned seem limited only by the ingenuity of greedy capitalists or the suspicions of the critics of big business. Several of the most prominent historically include price fixing, division of markets, tying contracts,* and predatory actions to "destroy" competitors. Such actions have been deemed anti-social in the U.S., which had developed a legal framework of antitrust law regulating such behavior. Needless to say, antitrust legislation is a vast topic in itself and almost any generalization we might offer about it is subject to debate. It is interesting to note, however, that many other countries, notably on the European continent, allow many of those practices which are illegal in the U.S.

3) "Unintended" Effects of Government Regulation. We have already noted that government regulation often has the

*A tying contract obliges a buyer to purchase more than one item regardless of his preference among the items in the package. This technique is held to be illegal when it is a device used to lessen competition. Note that there are many instances where similar arrangements (products sold jointly") are accepted as commonplace and innocuous.
(presumably) unintended effect of contributing to the lessening of competition. On a national level, suspect areas include the interstate regulatory agencies and interferences with international trade. Regulation may have similar effects at the state and local level, especially with regard to licensed professions. Licensing boards, which decide who is legally allowed to practice a given profession, are typically dominated by the practitioners of the profession themselves. Clearly, such boards might have an interest in limiting the access of potential competitors to their markets.

4) **Ownership of Unique Resources.** A few monopolies arise on these grounds, especially when some essential resource is very limited or concentrated in its distribution. Such cases are rarer than you might imagine because most resources, outside of a few dramatic examples, have some substitutes. Monopolized resources give a strong incentive for the discovery and/or development of substitutes as history has amply demonstrated. The more valuable the resource, the stronger the incentive to develop substitutes—consider the industrial diamond, synthetic rubber, plastics, etc.

In summary, we have noted a number of conditions which contribute to monopoly power in markets. Each implies some degree of insulation for existing sellers or limited access to the market for "outsiders." This limitation upon entry, be it natural or manmade, is the essence of the non-competitive market. If entry is free enough to allow many sellers
to survive, competitive markets are likely to prevail. Conversely, to the degree that entry is somehow restricted, monopoly power may arise.
Chapter 17. The Market For Labor as a Factor of Production

INTRODUCTION

In this and the next two chapters we shall shift our attention from the markets for commodities to the markets for one factor of production, labor. Factors of production are the various resources firms combine to produce commodities. Traditionally, these many factors are classified into 3 categories, viz. land (natural resources), labor (human resources), and capital (man-made resources). Families receive the incomes which enable them to purchase commodities largely from the sale of the services of the factors of production they own. Thus, the operation and outcomes of factor markets are crucial determinants of family welfare. As we shall see, the supply and demand framework we have already developed for examining commodity markets will be the basic engine of analysis for factor markets, too.

Labor-The Human Factor of Production

Buyers and sellers of labor services interact in markets determining wage and employment levels for particular individuals, groups, and the market as a whole. The examination of the operation of labor market is, of course, a vast study in itself. We shall first consider two interesting and important changes that have taken place in the 20th century in U.S. labor markets.

The demand for labor and other factors of production is derived from the demand for the commodities which they help
to produce. Intuitively, you should realize that the prices buyers of labor are willing to pay for different quantities of particular factors depend upon how much they contribute to production. Firms purchase factors not for the direct consumption of their services but because the factors contribute to the production of commodities for which there is a "final demand." "Final demand" refers to the demand for commodities by buyers. The demands for factor services are derived from the final demands as distinguished from the final demands themselves, which are based upon the actual consumption of commodities.

The theory of the supply of labor is also very similar to ones we have already used. We shall consider the household as the decision-making unit which (as in the theory of demand for goods) attempts to maximize its level of satisfaction. The problem for the family is finding that combination of labor market participation (which enables it to buy goods) and leisure (time "consumed") which maximizes the family's "utility" or satisfaction. Whether as buyers or sellers of commodities or of labor, households seek to find their most preferred position, subject to given prices and some kind of constraint which determines their feasible alternatives. In other words, the analysis of the family decision to sell labor (or other resources) bears a fundamental similarity to the analysis of the family decision to buy commodities. In the next few sections, we consider household labor supply
decisions in the context of some interesting changes in labor markets that have occurred over the years.

The Decline in the Average Workweek

One of the most striking changes in the operation of labor markets in this century has been the dramatic decline in the length of the average work week. In the table below, the average work week in manufacturing is shown for selected dates. You will note that the average fell by about \(\frac{1}{3}\) between 1900 and 1950 but has remained about the same since the latter date. While the decline has proceeded more rapidly in some periods than others, and the average for any one year in influenced by current business conditions, the downward trend is unmistakeable in the figures.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>AVERAGE HOURS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
<td>62.1</td>
</tr>
<tr>
<td>1910</td>
<td>59.8</td>
</tr>
<tr>
<td>1920</td>
<td>53.5</td>
</tr>
<tr>
<td>1930</td>
<td>42.1</td>
</tr>
<tr>
<td>1940</td>
<td>38.1</td>
</tr>
<tr>
<td>1950</td>
<td>40.5</td>
</tr>
<tr>
<td>1960</td>
<td>39.7</td>
</tr>
</tbody>
</table>

*For a more detailed treatment of this material see Ch. 2-6 of Labor Economics: Theory and Evidence by Belton Fleisher (1970--Prentice-Hall).*
The figures for 1930 and 1940 are deceptively low because average hours worked is sensitive to the level of unemployment, which was unusually high in both of those years.

There are a number of popular explanations of this decline in the work week. These include the following: the effect of trade union bargaining, the effect of wages and hours legislation, a shift in attitudes which has "allowed" the change to take place. While these arguments may have some explanatory power, most economists do not rely heavily upon them. The first two factors could have had a significant impact only after the late 1930's because of the limited scope of unions and legislation until that time. But the data suggests that most of the reduction in hours took place before these effects were operative. Finally, economists are wont to reject changes in attitudes or tastes as a major explanation of any economic phenomenon. It is not that economists believe that changes in attitudes are never important economic factors but simply that such changes are so difficult to measure that, as a result, almost anything can be "explained" by assuming that attitudes did change in a certain way. Economists would prefer to base their conclusions, whenever possible, on firmer ground.

An alternative theory which most economists find a persuasive explanation of the decline in average hours is based upon families' responses to rising levels of real income. At the same time that average hours have been falling, the
real value (in terms of the purchasing power of commodities) of wage rates earned has risen. Economists believe that the quantity of labor supplied per worker to the economy as a whole has fallen as the real wage has risen. Such a supply curve is shown in figure 43. You should be struck by its unusual slope—negative. Since this explanation posits a relationship completely different than ones we have encountered before, some additional explaining is in order.

We should note some unusual features of this supply curve. It represents the amounts of labor offered to the entire market, rather than to an individual seller. Each firm would still find that as its wage changed relative to the wage of other firms, that the quantity of labor offered to it would respond in the normal way. Second, this supply relationship is dependent upon permanent wage levels, ones that are expected to last more than a short time. Finally, these observations were made at different points in time rather than our usual supply curve which depicts the supply function at some given point in time. With these qualifications in mind, we will develop an explanation of this unusual negatively sloped supply curve relationship.

The Labor Supply Decision

A family must decide how to allocate its resources in its attempt to maximize its material satisfaction. An important aspect of this decision is choosing the amount of time that family members will spend in different activities, such as
Figure 43
The Backward-Bending Supply Curve of Labor

Figure 44
The Choice Between Income and Leisure
supplying labor services to the market. The total amount of time available to the family or to any one of its members is limited. Therefore, time spent working reduces the amount of time available for other pursuits. Time spent working and time "consumed" both yield utility but one can be increased only at the expense of the other.

Even though people have different "tastes" for work and value non-work hours differently, it seems clear that no one is so indifferent to how he spends his time that the number of hours "left over" after work is entirely unimportant. We are assuming that most people value their time and that "leisure" (time spent not working) is a desirable "commodity" to consume just as goods and services are desired.

Consider figure 44 which shows the labor market alternatives for one family member, the household "head." For simplicity, we shall initially assume that the family decision is solely a matter of choosing the number of hours that the head should sell in the market. Of course other family members might decide to supply their services to labor markets, too. We shall soon consider some aspects of their decisions. But for now, let us concentrate upon the family breadwinner.

The family head has 24 hours a day to allocate to various activities, just like all of us. Some of his time, he may sell in labor markets. Each hour sold adds \( W_1 \) (his initial wage rate) to the family income. So if the head decides to
"consume" only 23 hours a day and sell one hour a day in the market, his choice is indicated by point A in the diagram. If \( W_1 = $3.00 \), then the family can consume that much of commodities, ignoring the possibility of other sources of purchasing power (savings, loans, gifts, etc.). Similarly, each point on this wage path is a possible combination of hours consumed (not worked) and income earned. For example, Point B indicates a combination of 8 hours worked and $24 income, per day, for the head. Exactly which point the head would choose depends upon several kinds of tastes (for income, work, leisure, etc.) that differ among individuals. And indeed, people do work different numbers of hours although many work about the customary 8 per day, 40 per week. Some work more, some less and, of course, what is customary has changed drastically over the years.

**Do Individual Decisions Really Matter?**

It is easy to believe that the individual decisions just described about the desired number of hours to be supplied are of little moment in determining actual hours worked. After all, most heads work the normal week (whatever it happens to be) and any individual would probably find some difficulty in negotiating a special schedule for himself. As appealing as this reasoning may seem, do not accept it on face value. Individual decisions about hours worked do matter in the same way that individual decisions about quantities offered for
sale at different prices matter. In competitive markets, the market demand and supply functions are the result of many individual decisions. Although the preferences of any one individual may seem insignificant, just as those of a single worker may seem too insignificant to secure a special work schedule for himself, the aggregate of such decisions is crucial to the market solution.

By and large, employers are interested in hiring a certain total number of hours of labor, given the prices and characteristics of various factors of production and the overall demand for their product. At any point in time, it is no doubt most efficient to have more or less common work schedules for all workers. These common schedules probably represent some kind of average of most workers' preferences since employers are more interested in total hours of labor hired rather than hours per employee. As workers' preferences change, employers have an incentive to adjust the customary schedule toward the workers' choices. Insofar as it is the total number of hours of labor the employer is really interested in, he has an incentive to offer workers the schedules they like. Employers who do not adjust to changed desires on the supply side, will have to pay higher wages per hour to overcome employee preferences for different schedules.

Today, employers seeking workers for 12 hour days do have to pay higher wages per hour than they would for 8 hour days, though this was not always the case. Having stressed the
importance of workers' desires, let us see how these desires may have affected the average hours worked as real wages have risen.

Workers' Response to Higher Wages

We are at last at the point where we can present our theory of workers' adjustment to increasing real wage levels. This theory presents the explanation of the decline in average hours that most economists find most valid. Consider figure 45 which is the beginning of this analysis.

Observe first the two wage paths AB and AC. For any number of hours worked (not consumed in non-work activities) income is higher along AC. Thus, AC represents a path with a higher real wage rate, and the shift from AB to AC is the kind workers have faced over the 20th century as real wages have risen. Contrast these wage paths with CD which runs parallel to AB, implying that the wage per hour is the same along the 2 curves. Since CD lies above AB at all points, despite the identity of their wage rates, along CD the family has some non-labor income measured by the vertical distance between A and D.

Suppose first that the family's income possibility curve shifted from AB to CD as shown below (implying that the family can now consume more at any level of work). If initially it had chosen to supply 8 hours of labor per day to the market, a whole new set of choices as indicated by the line XY would be open to the family, such that less work and more consump-
Figure 45
Result of Gain in Non-Wage Income

Figure 46
Result of Increase in Wage Rate
tion are possible simultaneously. Any choice of income and leisure within this set of points is clearly better than the initial position since it includes more of either of both "goods" without a diminution of the other commodity. The family could now have X—more income (and as much leisure as before)—or Y—more leisure (and as much income)—or some combination with more of both. It seems most likely that some choice between X and Y would be most preferred since leisure is the kind of good that one probably would like to consume more of as income rises.

We cannot be so sure of how one would react to a shift in the income possibility curve AC which results from an increase in the wage rate. Since the higher wage rate can be exploited only by working, one might expect hours worked to rise and hours consumed to fall. Indeed, this is the usual result of a temporary increase in the wage rate, as in overtime provisions. Consider now figure 46.

The individual who has chosen Z initially when his wage was \( W_1 \) now has a new set of attainable combinations of income AC from which to choose at the higher wage rate. He might choose some combination between X and Y, implying an increased consumption of leisure and a reduction in hours worked, just as when his choices increased before. A new possibility, however, is a point such as W which requires more hours to be worked and less leisure to be consumed.
This new kind of point $W$ is now a more likely possibility than when the curve shifted upward in parallel fashion, because of the effect of a wage increase upon the cost of leisure. The AB to CD shift had no effect upon the cost of leisure. An hour of leisure still meant that the same amount of income, $W_1$, was foregone by not working both before and after the shift. The shift from AB to CD implies that leisure is now more costly, because an hour's income foregone has risen from $W_1$ to $W_2$. Thus, two opposing effects are initiated by the wage increase. The outward shift increases the range of consumption possibilities for both kinds of commodities, goods and time. This effect alone would induce one to consume more leisure, i.e. work less. But the wage increase has also made leisure consumption more expensive relative to the consumption of other goods, thereby inducing one to work more.

On strictly theoretical grounds, one could not be sure which of these opposing effects would dominate. We can only conclude that, on the basis of the evidence observed, the "more leisure" effect appears to have prevailed.

We might note in closing that economists remain at something of a loss in explaining the slowdown in the decline in hours worked in recent years. Since 1950, the standard workweek seems to have remained about the same, with virtually no downward trend. One possible explanation is that the preference for working less than 40 scheduled hours per week is weak, at least at present. Further increases in leisure
might be more enjoyable if consumed continuously as in longer vacations, more holidays, etc. rather than in reductions in scheduled hours per day. Indeed, scheduled hours per year have continued to decline even as hours per week have remained about the same. At present, economists' thinking on the subject of the apparently stalled decline in customary hours worked per week has yet to move beyond the stage of largely untested speculations. An interesting development related to this matter is a series of "experimental" work schedules (e.g. 4-10 hour days) that have received some notice. No definite conclusions about such innovations have yet been established.

The Increased Labor Force Participation of Women

Moving beyond the simple case where only the male household head supplies labor to the market, let us consider the decisions of married women about labor market participation. There has been no greater change in labor market behavior for any major group in the labor force than the change for married women. As the table below shows, the percentage of married women in the labor market has increased greatly in the 20th century. The increase in market work does not necessarily imply a reduction in leisure time or an increase in the total amount of work done my married women. Undoubtedly, the major change has been a change in where work is done (home vs. the market) for the increased proportion of women in the labor force, rather than an increase in work per se.
A number of explanations of this dramatic shift should occur to you. These probably include 1) changes in attitudes (both toward women working and toward market work itself by women), 2) favorable legislation (protective, anti-discrimination), and 3) the development of labor saving devices for home production (household appliances, prepared foods, etc.). Economists generally regard these factors as being less significant than most observers would allow. The change in attitudes is not stressed for the same reason we mentioned earlier in connection with the decline in the standard work week. Whether protective legislation has had favorable effects upon female employment is quite debatable (even if it increased supply it may have reduced demand) and anti-discrimination legislation is too recent to explain most of the trend. Household innovations may be as much, or even more, the result as the cause of increased participation.

While these previous explanations, as well as others we might have mentioned, explain the increased participation of women by positing rightward shifts in the supply function of
female labor, most economists regard demand shifts as being at least equally important. A strong piece of evidence supporting the economists' case is the behavior of the wages of women, relative to men, as shown in figure 47.

If the increased participation of women resulted from supply shifts alone, their wages relative to men would have fallen from $W_1$ to $W_2$. If demand shifts had been equally important, the relative wage would not be changed by increased female participation. The latter seems much closer to what has occurred, so the importance of demand shifts cannot be ignored.

A number of factors may have contributed in this increase in the demand for female labor, relative to other groups. These include 1) changes in work requirements, 2) a reduction in immigration levels, and 3) an increase in the educational attainment of women relative to men. Another important development increasing the labor force participation of women has been the apparent substitution of married women's time in the market for that of other family members such as children and the elderly. We shall consider these developments more fully in the next two chapters.
Figure 4-7
Female Wages Relative to Male Wages
Chapter 18. Work Disincentives in Labor Markets

INTRODUCTION

Governments have often undertaken policies in an attempt to improve the performance of markets. A good many of these efforts have been undertaken in labor markets. Various policies have aimed at reducing unemployment, raising or lowering wage rates, and a number of other objectives. These policies sometimes have the effect of altering the incentives families and individuals have to participate in labor markets. In this chapter we shall consider two programs which have the effect of discouraging participation in markets by reducing the economic incentive to participate. Whether these effects are intentional or desirable is a matter of debate, as we shall see. You should always keep in mind that the nominal motives and ostensible objectives of policy actions are not necessarily consistent with the actual results of policy actions.

The Labor Force Participation of Older Men

The percentage of men aged 65 or greater who participate in labor markets has declined markedly over the 20th century as the figures below indicate. A number of explanations consistent with economic theory immediately suggest themselves. The most important reason seems to be a result of legislation rather than the free operation of labor markets.
Before asking why the labor force participation of older men has fallen, we would do well to consider why participation rates for this group have always been lower than for prime age men. In other words, why do people retire when they are older instead of at some other point in their work lives? This question is not a trivial one, although it may seem to be at first, because an understanding of why people do typically behave in this way will aid our understanding of why retirement is so much more common than previous years, i.e. why the labor force participation rate of older men has fallen considerably.

We have seen that decisions made about the amount of time allocated to labor market participation and leisure compare the benefits and costs of various alternatives. One could argue that the benefits of retirement (continuous leisure) are greater when people are relatively young and that a given amount of leisure consumed earlier, rather than later, in life.
would yield more satisfaction. Observing some young "drop-outs" who may well eventually go "straight," i.e. get a job, suggests that there is something to this line of thought.

You might think that the reason that retirement comes at the end rather than the beginning of one's work life is that savings must be accumulated before retirement is possible. But one could conceivably retire on borrowed funds to consume leisure when it was most enjoyable, then begin working for as long as one was able. Actually, families often go substantially into debt when the family is young, to purchase a home or other goods for current consumption rather than saving for it. There is no reason, in principle, why borrowing amounts sufficient to support early retirement, later to be paid off by subsequent periods of work, could not also be arranged. Even though there is uncertainty about the length of work life, not to mention life itself, this problem is not unsurmountable. Average worklife and income can be predicted with sufficient accuracy to permit "early retirement" loans to be insured against unexpectedly short periods in the labor force.

We have already seen that this period when the labor force participation of elderly men has fallen was also one when wage and income levels were rising. One might reason that these increases have made retirement more feasible by increasing the capacity of people to accumulate savings for retirement. There, no doubt, has been such an effect. Of
course, this still does not explain why retirement comes later rather than earlier in one's life since rising income levels could just as easily facilitate either "kind" of retirement, "early" or "late."

**Why Older Men Retire**

To get a more definite explanation of the decreased labor force participation rate (greater frequency of retirement) of men 65 and older, we shall return to the analytical framework developed in the previous chapter. Figure 48 shows the effects of aging upon the choice between work and leisure.

A family of labor market budget lines are shown. Initially, the wage rate attainable in the market rises as shown by the increasing steepness of the lines going from the bottom of the second and third ones. This corresponds to the increase in wage rate normally received by workers as they gain experience, seniority, and other forms of human capital which make them more valuable in labor markets. This period of rising wages has a different time pattern for different workers. (Take for example professional athletes versus lawyers; their respective earnings typically peak at much different age-tenure levels.) Most workers do experience such a period of rising wages at some time in their working career. Similarly, most workers eventually find that their productivity and wage level diminishes with age after a certain point as is shown in the upper two curves which have become less steep
than the previous two. Note that the level of income attainable from non-labor sources rises continuously, i.e. we are assuming that savings and the income earned from saving are continuously rising throughout the lifetime.

We are finally at any explanation of why retirement is postponed until the later years of one's life. Given the typical pattern of lifetime wage rates, it becomes rational for a worker to plan (albeit in an uncertain environment) to concentrate his labor market activity during the period in which his productivity (wage) is highest, postponing leisure until it is less costly. This adjustment is similar to the adjustment workers make to temporary high rates of overtime pay. Here, of course, the "premium rate" is earned in certain years rather than extra hours in a given week but the effect of trading costly current leisure time for cheaper future leisure is the same. If the only adjustment older men made in labor force participation was in terms of hours supplied per period, we would expect to see a curved lifetime hours path such as the one shown in figure 48. Such an effect does exhibit itself in the date, i.e. hours worked rises and then falls with age. However, movements in and out of the labor force entirely is the more important kind of adjustment for older men.

The Effect of Social Security Laws

We have noted that rising wage-income levels have made voluntary retirement more feasible for workers and that the
lifetime pattern of wage rates makes it rational to expect the period of retirement to come later, rather than earlier, in one's life. The most important reason for the sharp rise in retirement is an "institutional" change rather than these economic factors, however. It arises from a strong work disincentive built into Social Security legislation. This legislation at once makes retirement at older ages more feasible by providing a source of income and discourages working (beyond a relatively few hours per week) by reducing retirement benefits for older workers who remain in the labor force. This latter effect results from the provision which limits the labor income a retirement income recipient can earn while still leaving his retirement allotment unaffected. Once this limit is exceeded, the recipient's social security benefit is reduced, 50 cents for each dollar earned. His "net" income (labor earnings plus retirement allotment) remains the same for some amount of labor force participation, depending upon his wage rate. Consider the table below.

<table>
<thead>
<tr>
<th>Wage Rate per Hour</th>
<th>Number of Hrs. Worked per Mo. Before Limitation Effective.</th>
<th>Necessary Number of Hrs. Worked to Raise Income Beyond Maximum at Hrs. Limitation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ 2.00</td>
<td>87 1/2 hrs.</td>
<td>187 1/2 hrs.</td>
</tr>
<tr>
<td>$ 4.00</td>
<td>43 3/4 hrs.</td>
<td>93 3/4 hrs.</td>
</tr>
<tr>
<td>$10.00</td>
<td>17 1/2 hrs.</td>
<td>37 1/2 hrs.</td>
</tr>
</tbody>
</table>
This table shown the effect of these social security provisions at the current maximum monthly labor income allowed before reducing the retirement allotment, $175, and that the amount received as a retirement income is $200 (benefits received currently range from about $85-$300 monthly, depending upon eligibility requirements, age, length of work life, etc.) Given these values for these variables, you can see that the higher one's wage, the sooner the limitation becomes effective. Also, the higher one's wage, the fewer the number of hours one would have to work for his net income to exceed the maximum amount attainable through a combination of work income and social security payments, in this case $375 ($175 + $200).

You should be able to see what effect changing the maximum income limitation (say to $50, or $350, instead of $175) or changing the size of the retirement allotment (say from $200 to $300 or $1,000) would have on these decision points.

Exactly why this provision was written is not entirely clear. At the time of the first FICA Act it may have seemed a good idea to encourage older workers to leave the labor force to "make room" for others. Whether others "deserve" jobs more than the elderly is, of course, a matter of debate and, if it is decided to promote retirement, the resources to support these new non-workers in retirement must come from the production of those remaining at jobs or from investment of social security taxes paid in the past by currently retired workers and their employers. Very few economists subscribe to the notion that there are "only so many jobs to go
around" in the sense that employment possibilities are unaffected by the number of workers seeking jobs as plausible as this may seem from the viewpoint of an observer on the supply side of the labor market. Older workers could add to the potential supply of output and could add to realized output if aggregate demand was sufficient. We will turn to an explanation of the aggregate demand for output in chapter 20. We could at least wonder whether public policy might be somewhat more "neutral" in allowing older workers a choice of continuing at work.

Disincentive Effects of Welfare Laws

Another interesting and probably more important work disincentive effect is built into various provisions of income assistance programs, popularly called "welfare." These laws are administered at the various state and local levels, so specific details also vary. In general, however, the provisions we shall describe do operate in these programs to discourage participation in labor markets. Given frequent complaints about the "welfare mess," viz. the increasing cost and allegedly widespread cheating by recipients, it seems unlikely that these disincentive effects were entirely intentional.

The largest single component of the welfare program is Aid to the Families of Dependent Children (AFDC). Benefits depend partly upon the number of dependent children to be
supported. These benefits are sharply reduced, however, if the household head has labor market earnings. This provision has the same effect as an extra income tax rate of about 67% would have upon the family's net income (earnings plus allowance) in the absence of any welfare program. In other words, earning an extra dollar in income increases the family purchasing power by about 33%.

Actually, the family incentive to earn income is further lessened by other taxes upon income. Additional provisions of the welfare programs add to these effects by restricting the eligibility for certain programs (food stamps, medicaid, etc.) on the basis of income. Adding in costs of holding a job (the journey to work, extra clothing, etc.) means that working might add very little to a welfare family's purchasing power (it could even reduce it) while it would reduce the time the eligible parent spent at home and/or at leisure. While it is frequently alleged that welfare recipients are just lazy and that they should work instead of receive public aid, one should realize how these provisions have the effect of discouraging working.

We have just described the immediate work disincentive effect of welfare laws. The system has further effects which seem to operate to perpetuate the problem it was presumably designed to alleviate; welfare parents are in a poor position to help their offspring to succeed in their future employment efforts. On the contrary, the experience of the parent leaves
children without work models to emulate and gives them only a knowledge of the welfare "way of life." Thus the system may operate to increase the size of the problem which was initially conceived to be mainly temporarily indigent families. The system may have the effect of making this condition permanent. Reducing work incentives in the short run reduces the incentive to develop the skills and other characteristics which would contribute to labor force success. The possibility of working one's way off of welfare is discouraged and diminished by the disincentive effects already discussed. It seems likely that continued dependence upon welfare destroys the "work ethic" or the "will to work" of many individuals and perhaps their offspring, as well. It is not unlikely that the provisions of AFDC which sometimes require a male spouse to be absent for family eligibility and that payments vary directly with the number of children encourage desertion by fathers and excessive childbearing by mothers. Presumably, these effects are not desirable ends for public policy to promote.

Is There A Better Way?

Virtually all income assistance programs have some disincentive effects. One would hope that more carefully structured provisions might reduce these effects. A widely suggested program to replace the present system is the so-called negative income tax. This proposal seems attractive, in principle, across a wide range of the political spectrum.
Proponents of the negative income tax, in its many versions, point to several advantages. The most relevant for our purposes is the reduced disincentive effect. The "implicit tax rate" upon earned income varies with the particular plan but, generally, is far less than the present system which, as we have just seen, may have the effect of a 100% income tax. Thus families could improve their net income position much more from working than the present system allows.

Since the negative income tax would replace the entire welfare structure, specific decisions on some of the other provisions of current welfare programs we have mentioned would have to be made (eligibility for programs in kind, the question of male spouse presence, etc.). Two probable benefits of a national negative income tax program which replaced the present welfare programs would be greater uniformity and the reduction of an administrative bureaucracy. Another change would be in the degree to which the government agency administering the program would "invade" the lives of recipients. Many people object to the present system because the recipient's family life is intimately affected by the intrusion of social workers supervising, advising, etc. Of course, not everyone would agree that such a system is undesirable even if it is demeaning. That is only one of very many value judgments on the questions pertaining to welfare and income redistribution.
WHAT IS INVESTMENT?

The term "investment" is used quite differently by people who have different concepts in mind. For example, one commonly speaks of "investing in the stock market." This appears to be somewhat different from "investing in a home," or "investing in a college education," or a business' investment in its inventories. To reconcile these different kinds of investment, we must gain an understanding of exactly what investment is. Our aim is to develop a framework for analyzing a particular kind of investment, of which an "investment in a college education" is an obvious example. Economists call this kind of investment "human capital" and it refers to efforts and expenditures made designed to make labor resources intrinsically more productive.

Imagine the simplest kind of economic system--an isolated individual on an oceanic island--a "Robinson Crusoe." This paradigm is a favorite of economists and other academicians because it allows one to focus upon fundamentals without (at least initially) certain complications. Within this framework we seek to acquire an understanding of the nature of investment without the complication of a monetary system, a multitude of individuals, and a system of inter-related markets. These features of the real world can be considered once the essence of investment is understood.
An understanding of investment begins with one feature common to all economic analysis. This problem is "the economic problem," scarcity, the topic with which we began our study in Chapter 1. Crusoe's, like every other economic system, must decide what goods to produce and how to produce them (we shall assume that fortune has not been so kind as to make him so rich in resources that the choice among alternative desires is no longer relevant). Crusoe must choose among alternative commodities and production methods because it is not feasible for him to acquire all the goods and services he would like, given the resources at his command. Of course, Crusoe's economy does not have to make the third kind of decision usually implied by economic scarcity, viz. who shall receive the goods produced since he is this economy's sole consumer, at least until a Friday comes along.

The resources our Crusoe has at his command to satisfy his material wants include the natural resources of his island (land, water, plants, etc.). A second set of resources Crusoe will employ include his own personal attributes (knowledge, skills, experience, energy, "resourcefulness", etc.) which he possessed when he was deposited upon his island and which might be augmented and/or diminished by his experience on the island. These personal characteristics are a kind of human capital, as we shall see. A third set of resources includes any man-made objects Crusoe brought with him to the island that would add
to his production and consumption possibilities. These might include clothing, tools, a compass, etc. To survive and perhaps even to prosper, Crusoe seeks to combine his available resources in ways that maximize his utility. Crusoe is not limited solely to the resources available at a moment in time, however. The possibility of changing and adding to the resources which contribute to his production and consumption brings us immediately to the idea of investment.

Suppose that Crusoe "discovers" the coconut when one falls at his feet, just by chance. After sampling this new good, Crusoe concludes that adding it to his diet would increase his utility, provided that the cost (in terms of foregone opportunities) is not "too large." The cost of the first coconut was small but Crusoe can continue to consume them at such a price only if he is willing to have coconuts "come to him." Since Crusoe likes coconuts, he decides that it might be worth some effort (cost) on his part to increase and regularize his consumption of coconuts. How might this be done? Coconuts grow in trees that take a number of years to develop so, for at least such period, Crusoe's coconuts will be forthcoming from the existing trees on the island. The problem for Crusoe to decide is how best to "harvest" the coconuts.

The immediate problem is that coconuts grow high off the ground. Assuming that he is not content just to wait for coconuts to be "delivered" as the first one was, Crusoe wonders about the various ways he might bring down some coconuts.
There are several methods he might employ which require only his own labor efforts. For instance, he might try to climb the tree to pick the coconuts, or he might try to get some coconuts by shaking the tree with his hands, or even to dislodge some coconuts by butting the tree with his head, after a powerful running start. But, of course, Crusoe could also use other resources at hand besides his own labor. He might throw rocks at coconuts or he might be fortunate enough to possess an ax or some other tool with which he might cut down a tree. Any method he considered would be judged in terms of its cost and effectiveness.

Still another alternative exists for Crusoe which requires that he combine some existing resources into a new form. He might decide to build a ladder or make a boomerang or some other "tool" that will facilitate the harvesting of coconuts (and perhaps have additional uses). Again this alternative must be judged in terms of its cost and effectiveness. This alternative differs from the preceding ones because it requires that Crusoe devote resources to the production of some "intermediate" good before he turns his efforts directly to the harvesting and consumption of coconuts. This act of making "intermediate" goods is the process of investment as economists understand it. Investment consists of the expenditure of resources designed to create "new" resources to employ in the production and consumption of goods and services. The amount of such intermediate goods existing at any time is the stock of capital, as economists use the term.
There are numerous kinds of capital a Crusoe might consider undertaking as investment projects—a canal, a boat, a plow, etc. What they all have in common is that their creation requires some of the scarce resources—Crusoe's time and effort, natural resources, existing capital—which immediately implies opportunities foregone. This is precisely the same problem that every individual making an investment decision must make. If one is to use resources to build a house, to accumulate an inventory or (to come around to the main topic of this chapter) to pursue a college degree, other uses of those resources must be foregone. Any investor weighs the expected returns (what he expects to receive from the investment in monetary and/or non-monetary forms) against the opportunity cost of the investment. This cost includes various other uses of the resources—using the resources for more immediate consumption, alternative investment projects, etc. Investments are undertaken if the investor feels that the returns will be more valuable than the costs incurred.

People invest in themselves and other individuals in basically the same way and for some of the same reasons that firms invest in physical assets. The most obvious form of "human capital," originates in expenditures for education and training. Clearly, a major motive for the acquisition of education and training is the expected effect this investment will have upon the future earnings of the individual receiving it. Some people object to the notion of treating education as a kind of
investment and find the concept of "human capital" repugnant. Economists have found the human capital concept a useful one. It is true that investment in human beings has a few unique features on purely technical grounds which we will have occasion to mention. Still, the differences between investments in human beings and non-human objects are more apparent than real. All investments are undertaken by individuals or firms (which, of course, are owned by individuals) with the purpose of securing future benefits which may be monetary or non-monetary. There are many kinds of investment and the classification of investment into two categories, human and non-human, is simply one way of grouping alternative forms of capital.

The Labor Force Participation of Youth

One of the most striking changes in the composition of the labor force during the 20th century has been the decline in the rate of labor force participation of young people. Concurrently, there has been a rapid rise in the school enrollment of individuals who in previous years might have entered the labor force. Common sense and the data below suggest that labor force participation and school enrollment can be alternative activities, that for many people either one or the other activity is chosen to be pursued.
A variety of forces have acted to increase the likelihood that a young person will choose to remain in school instead of entering the labor force while still in his teens. As usual, this change in labor force behavior can be explained as the result of 1) economic forces, 2) institutional forces such as changes in legislation, and/or 3) changes in attitudes. The last factor, as you may have anticipated by now, is not one economists are likely to stress. While attitudes toward school attendance and labor force participation have undoubtedly changed in ways consistent with the changes in the patterns of behavior we noted in the above table, it seems that these attitude changes have been much more the effect than the cause of the increased school attendance and diminished labor force participation of youth. And, as always, there is the problem of measuring attitudinal changes which discourages economists

*Again, see Labor Economics: Theory and Evidence, Ch. 2-4, by Belton Fleisher, Prentice-Hall, 1970 for further elaboration and data sources.
from relying too heavily upon them as explanations. This should not be taken to imply that changes in attitudes have had no effect, but only that other factors have had a larger and more independent effect.

Changes in several kinds of legislation have had significant impact on the school-work choice for young people. Economists generally include 1) minimum wage legislation, 2) protective legislation for child labor, and 3) compulsory school attendance laws in the legislative changes that have affected the basic choice. Minimum wage laws and protective legislation have a similar effect upon the terms at which employers can hire young labor. If these laws are effective (i.e., actual market conditions are different than they would be in the absence of the legislation) their impact is either to raise the wage that must be paid or to increase the cost associated with employing young workers. Either effect would discourage employers from hiring young workers, as opposed to other factors of production (other workers, physical capital, natural resources) and therefore reduce the employment of youth. Thus, these two kinds of law have operated on the demand side of the labor market to induce the shift from work to remaining in school.

It is somewhat more difficult to judge the effect of compulsory school attendance laws upon youth to labor force participation and school enrollment. While the effect of
such laws might seem to be obvious, as changes in attitudes, this change may have been more a result than an actual cause of increased school enrollment. If families had felt as strongly the desire to send their offspring into the labor market as they had previously, it is difficult to believe that such legislation would have been enacted or enforced. Conversely, if this legislation merely affirmed a pattern or trend already established, its real impact might have been minimal. Most economists would probably consider these laws less cause than effect in their relation to the labor force participation of youth.

There are two major "purely economic" variables which have had a major impact upon the choice between early labor force entry and extending the customary period of education. The first effect is the impact of the population shift from rural to urban areas, mirroring the diminished importance of agriculture in the economy. This shift has reduced the relative demand for young workers, as agriculture used to be a relatively heavy employer of young people. The impact of this shift has been large, perhaps larger than the combined effects of all three kinds of legislation we have just discussed upon the rates of labor force participation and school enrollment of young people. Together with the variable next to be discussed, these "economic" effects have outweighed all the others in their impact.
A second important economic variable which has induced the shift we have been discussing is the continued rise in the level of economic development. The process of economic development and concomittant accumulation of physical capital encourages human capital investment such as additional education. Also, the growth of family wealth associated with economic development facilitates the acquisition (investment) in human capital.

Like other investments, outlays on human capital are expected to produce future returns that justify the initial costs, and the return on human capital may sometimes be greater, sometimes less, than originally anticipated by the investor. One way in which human capital may differ from many other kinds of investments is that much of the return to it, especially from investment in education and good health, comes in the form of "consumption benefits" as well as from greater pecuniary income. Critics of the human capital concept who prefer to think of education as "intrinsically good" rather than an investment probably have these benefits in mind (the joy of learning, greater understanding, etc.). While these benefits cannot be ignored, it would certainly be an even greater mistake to ignore the effects of human capital investment upon earnings and to omit them from consideration in understanding decisions about the acquisition of education and other forms of human capital. The table below which shows the level of earnings, classified by the level of education
attained, should give you some idea of the effect of education on income in 1968 for individuals over age 25.

<table>
<thead>
<tr>
<th>Years of Education</th>
<th>Average Income, 1968.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3 Yrs of High School</td>
<td>$6,769.</td>
</tr>
<tr>
<td>High School Graduate</td>
<td>$8,148.</td>
</tr>
<tr>
<td>4 or More Years Beyond High School</td>
<td>$12,938.</td>
</tr>
</tbody>
</table>

While the entire difference between each group cannot be attributed to education alone (for example, the more educated are generally more able as measured by standards other than education) most of it probably can be. Parents or others, including the student himself, implicitly compare the returns on education versus other investments which might increase the recipient's future income. If the parents' aim is to provide offspring with a kind of yielding future income, they could consider several alternative forms. Human capital (such as education) can be thought of as an alternative to physical capital (such as a business) or claims on capital (cash, stock and bonds). All are investments which, hopefully, will provide future returns.

An investment in education seems to be a good one in the sense that the average rate of return (future benefits on the initial cost) has probably not decreased over time,
even though families have, on the average, increased their per capita investments in education. Some economists view the decision to have fewer children (as reflected in lower birth rates in recent decades) as, in part, a choice to increase the quality of offspring (in terms of greater human capital endowment).

**A Simple Example of an Investment**

While most readers probably have an idea of what we mean by the return on an investment, a simple example should clarify some of the concepts we have used so far. The example is quite simple in two senses, 1) the return of the investment is expected to be received entirely in the year immediately following the investment; 2) the return on the investment is certain, there is no possibility of it falling short of or exceeding the expected amounts. While in the real world these two considerations certainly do complicate the process of actual investment making, the basic principles of the investment making, the basic principles of the investment process itself can be fully derived from a model as simple as the one we present. The complications do not alter the principles of investing although they make practical decisions more difficult.

Suppose that an individual initially has a certain income of $1,000. to be received in each of two successive years. He is presented with an alternative choice of cer-
tain incomes which would provide him $500. in the first year and $1,600. in the second year. This second pair of incomes (which we shall call alternative B as opposed to the first, alternative A) in effect offers to the investor an extra $600. in the second year in return for $500. foregone in the first year. Is this a good investment? The answer depends, as for any choice situation in economics, upon a comparison between the return and the best alternative use of the $500. invested. To get an intuitive feel for these concepts suppose that if the individual chooses B, he can borrow $500 the first year, to be repaid in the second year along with interest of 10% ($50) on the loan. This amounts to the acceptance of investment of B without reducing purchasing power available in the first year. The three alternatives are summarized below.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Income Available in First Year</th>
<th>Income Available in Second Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$1,000</td>
<td>$1,000</td>
</tr>
<tr>
<td>B</td>
<td>$500</td>
<td>$1,600</td>
</tr>
<tr>
<td>B (with loan)</td>
<td>$1,000</td>
<td>$1,050 (loan repaid)</td>
</tr>
</tbody>
</table>

Clearly, B (with loan) is preferred to A and, the investor will accept the investment offered under alternative B as long as he can borrow at 10% interest. Even if he cannot or would not borrow to make this investment he would pre-
fer B to A if he considers the $600. received in the second year worth more to him than the $500. foregone the first.

It is convenient to relate choices A and B with the concept of the rate of return on investment. The rate of return in this example is 20% because the initial outlay ($500. less available in year one under option B) yields a return of an extra 20 cents on the dollar (totalling $100.) when received in the second year. This rate of return is compared with the cost of the investment which is measured in two ways. If B is chosen, along with a loan, the cost in percentage terms of the loan is 5% (amounting to $50.).

This alternative is clearly superior to A, since the only difference is that B, with loan, yields a greater income available for the second year and no less the first year. If it were impossible to borrow at 5%, alternative B might still be preferred to A. Here we compare the 20% rate of return with the opportunity cost of foregoing $500. the first year. This cost is measured by the difference in satisfaction the individual feels in having income available in year one versus year two. Most individuals prefer spending in the present, as opposed to the future, all other things being equal. If an individual regarded $1.00 available this year as equal in value to (say) $1.10 available next year then when we could say that the opportunity cost to him of foregoing $1.00 this year is measured by the 10% (10 cents) more
he would have to receive next year to make him feel equally well off.

For most investments, it is possible to borrow at some rate of interest in order to finance foregone consumption. Therefore, we shall usually think of the cost of making an investment as reckoned in terms of an interest rate that must be paid on funds borrowed for that purpose. Clearly, whenever this interest rate is less than the rate of return to the investment, then the investment is a desirable one in the sense that the benefits exceed the costs. Since the benefits of investments in all forms of capital vary widely, we can conceive of a demand curve for investments which slopes downward with respect to the cost (interest rate) as in figure 49. This is because at high rates of interest, only a few investments are worthwhile, at lower rates of interest, successively larger amounts of investment yield returns greater than their interest costs. As usual the slope of the demand curve is downward, reflecting the fact that more investments will be desirable, the lower the cost of making them. The demand for investment in all kinds of capital is thus conceptually similar to other demand curves.

Incidentally, the average rate of return on the investment required for a college degree is estimated to be about 10-15%. Whether such a return, which measures only the pecuniary benefits, justifies the investment depends upon cost of the investment to the recipient of the potential returns
Figure 49
The Demand for Investment
and whether he believes that his benefits might diverge from the average figures.

Who Should Pay For Education

If education and other forms of human capital reward their recipients with higher incomes, one might wonder why so much of the cost of education is paid by governments (i.e. taxpayers) rather than the recipients themselves.

One reason why governments have traditionally undertaken some of the cost of higher education has been the widespread belief that some, perhaps even a large part, of the benefits of college education are received by individuals other than the student himself. In recent years the legitimacy of such a subsidy has been frequently questioned, and it has been proposed that governments lend students amounts to finance their education rather than grant them what amounts to subsidization. In the state of Ohio, such a plan was proposed by the Governor in 1971 but was so widely opposed it never came to a vote. This issue is complex and, of course, heated since the amounts of money involved are large, and people tend to feel strongly about public education and one's "right" to receive it (at less than full cost). For a complete treatment of these issues one should see Chapter VI in Capitalism and Freedom by Milton Friedman, University of Chicago Press.

Actually, the largest cost of a college education is neither the tuition paid by students nor the subsidy paid by governments. This largest cost is the foregone earnings,
the opportunity cost of not working for the time one is receiving a college education. For a new high school graduate, this four year, "indirect", cost might in 1973 amount to $15,000 or more which is considerably greater than the amount of the public subsidy students might someday be asked to repay. The average rate of return on a college diploma (10-15%) already mentioned includes all three of these costs (tuition, subsidy, and foregone earnings).

Two economists commissioned to study the effects of the public subsidy for higher education in the state of California\(^1\) estimated that the total cost of a year's undergraduate study in 1965 ranged from about $2,900-6,800 depending upon the level of education. They found that foregone earnings in all cases amounted to more than half the total cost and that the proportion of the total cost borne by students and their families was about 70% or more. While the state of California differs from most states in that tuition charges are very low and the public higher education system is very extensive, these results are probably representative for other states.

The same study sheds light on another aspect of the public subsidy of higher education. The subsidy is often defended as an appropriate and effective means of redistributing

\(^1\)See Benefits, Cost and Finance of Higher Education by W. Lee Hansen and Burton A. Weisbrod, Markham Publishing Co. (Chicago), 1969.
income from the more wealthy to the less wealthy. It was found, however, that the public subsidy to higher education in California was received more by higher than lower income families. Since the state taxes were roughly proportional to income, the effect of this subsidy was to redistribute income toward the more wealthy. This effect is probably even more pronounced in other states. This surprising result is explained by a further consideration of the opportunity cost of higher education.

The public subsidy for higher education has its impact mainly in the reduction of tuition charges although part of the subsidy is devoted to aid to "needy" students. Still, the opportunity cost, the main cost of attending college is largely not offset by these subsidies. This cost, foregone earnings, is probably most acutely felt by low income families. Certainly such families would have the least access to loans or to accumulated wealth to offset this opportunity cost. If it is a goal of public policy to encourage poor students to attend college, subsidizing all students, most of whom are not poor, would seem to be a rather inefficient way to go about it. To the extent that one would hope to increase the enrollment of low income students, a more effective approach would be to subsidize only these students and to grant or lend them a greater amount of the total cost than is presently done, even under "full" scholarships. Such aid would be designed to compensate families for some or all of the opportunity cost incurred by college attendance.
CHAPTER 20.

Introduction to Macroeconomics

INTRODUCTION

Thus far we have limited our study to the realm of microeconomics, investigating supply and demand conditions in a single market or a small number of markets. Now we will shift our attention to macroeconomics which investigates supply and demand conditions in the large number of markets which taken together make up an entire economy. As in microeconomics, we will attempt to explain the determination of prices and outputs. However, instead of considering a single price or the output of a single industry, we will investigate the determination of the price level (a kind of average of prices in a number of markets) and the levels of output and employment for the entire economy.

Many of the concepts and much of the analysis in macroeconomics will build upon the material of microeconomics. But macroeconomics involves more than a simple aggregation of markets. One crucial task is the analysis of interactions among markets which requires us to keep track of many things that are happening simultaneously. Secondly, macroeconomic theory must abstract from many microeconomic details in order to restrict its explanation to manageable proportions. Thus, you may find macroeconomics somewhat less "intuitive" than microeconomics because it requires a vantage point different from the ones normally experienced in the everyday roles of
consumer and producer and because the theory does abstract from many familiar details of the economic landscape.

THE MARKET FOR LOANABLE FUNDS

Our simple macroeconomic theory will summarize the millions of markets by observing just three sectors, viz. the markets for commodities (currently produced goods and services), money, and earning assets (long-term debt instruments such as bonds). The myriad of goods traded in an economy all fall into one of these three categories. While this model is a simplified picture of an entire economy, it captures the essence well enough to give an understanding of the processes determining the levels of output, prices, and employment. More refined models have been developed to increase our understanding as well as to enable the prediction of future conditions and even (hopefully) provide the means for influencing those future events. These advanced models are based upon the principles incorporated in our simple model.

Our theory explains the workings of an entire economic system in terms of the interaction among the commodities, money, and earning assets sectors. For any given set of economy-wide conditions, there exists a set of equilibrium levels of output, prices, and employment, which simultaneously make demand and supply equal for each of the sectors. These values are the ones an economy approaches, and thus are equilibrium values in the same sense that there exists an equilibrium price and output for given conditions in a single market.
Similarly, the equilibrium values are subject to change when underlying market conditions change.

The interaction between the commodity, money, and earning assets sectors can be summarized in the market for loanable funds. The aggregate equilibrium levels of output, prices and employment can be determined by focusing upon the supply and demand curves for loanable funds. The purpose of the loanable funds market is to provide firms a means of acquiring funds necessary to finance their investment expenditures on capital equipment. We shall assume that firms can secure these funds only through borrowing in the loanable funds market. Thus we ignore the existence of alternative forms of financing employed by firms in the real world (for example, use by firms of their retained earnings) although these other methods of financing can be incorporated into a more sophisticated macro-economic model. As it is, we assume that firms secure all their investment funds through the sale of an earning asset. We assume these earning assets are similar to a bond. A bond is a long term financial asset which is really just a contract between borrower and lender. The lender gives current dollars to the borrower in return for the borrower's promise to repay those dollars with interest over the course of the future. Having seen why firms wish to secure funds through borrowing (why they demand loanable funds), we must inquire why other people are willing to provide funds, i.e. to act as suppliers in the market for loanable funds. The reason is that a bond
is a financial asset that provides the lender with an income in the form of an interest payment for the time during which the bond is held. A bond is really a kind of loan. There are many different kinds of bonds which because of their different characteristics yield different rates of interest. Again, we shall abstract from these details by speaking of a single bond market and "the" rate of interest, recognizing that these concepts represent a kind of average of many different bonds. In figure 50 we depict hypothetical conditions in the loanable funds market.

Let us now try to relate the supply and demand functions in the loanable funds market to the supply and demand functions in the other markets of the economy. First, we note that the supply and demand functions in the loanable funds market have their customary slopes with respect to the price of loanable funds, i.e. the rate of interest. The demand curve indicates that the total amount firms wish to borrow varies inversely with the interest rate. Since we have identified the demand for loanable funds with the demand for new capital equipment by firms, it should come as no surprise that firms' demand for loanable funds increases at lower interest rates. As the rate of interest falls, buying more capital equipment becomes more profitable. However, to finance these additional purchases of capital goods firms will have to borrow more on the loanable funds market. In figure 50 we see that the quantity of loanable funds supplied varies positively with the
Figure 50
The Market for Loanable Funds
interest rate. To see why this is so, let us examine how a typical household makes spending decisions when it receives income from the sale of its resources.

We can imagine that families first choose how much to consume, i.e. the quantity of commodities they wish to buy for consumption purposes. Economists believe that this decision reflects families' expectations about their "long run" income. On the average, families spend less than their entire current income on the consumption of currently produced commodities. We can express this relation between their current income, and desired level of consumption, as \( c = ay \) where \( a \) represents the proportion of income spent on consumption. Once this decision has been made, households have some income left over. The difference between income and consumption is called saving which we designate by \( s \).

Households now face the pleasant task of deciding in what kinds of assets they wish to hold their savings, i.e., the amount by which their current income exceeds their current consumption. In our simple world there are two kinds of financial assets available for savers, money and bonds. Each has its own desirable characteristics so that a family might regard them as partial substitutes for each other, holding more or less (or none) of one in place of the other, depending upon how the family valued the services provided by money and bonds. We have already seen that the most important service of money (whether held in the form of currency, coin, or checking
accounts) is that of a medium of exchange. Holding money facilitates making transactions and makes trading of goods and services much less costly. Holding bonds yields an interest income, however, which money does not. The interest rate then measures the foregone opportunities of holding money rather than bonds. Therefore, we would expect families to want to use their saving in order to acquire more bonds, the higher the rate of interest since money would become a relatively less attractive asset. Herein lies the rationale for our assumption that the quantity of loanable funds supplied varies positively with the rate of interest. Again, consider the supply of loanable funds, \( \Sigma \), in the figure 51. On the same diagram we also graph the vertical line, \( s \), representing the amount of saving or the unconsumed portion of income.

In general, there is some rate of interest \( r_1 \) below which families would not be willing to supply any loanable funds (buy bonds). Because holding money would have such a low price at \( r_1 \) households would prefer to use all their saving for the purpose of adding to them money holdings.

As the rate of interest rises above \( r_1 \) the cost of money holdings would start to rise and households would begin to divert some of their saving into the purchase of bonds. Families will supply increasing amounts of loanable funds as the rate of interest rises. At some rate of interest, \( r_2 \) in figure 51, families would devote all of their current saving to the accumulation of bonds since at \( r_2 \) \( \Sigma \) and \( s \) intersect.
Figure 51
Loanable Funds and Commodity Markets at Equilibrium
one another. Finally, as the rate of interest rises above \( r_2 \), bonds become such an attractive asset that households wish to use all their current saving to buy bonds and, in addition, they wish to draw down their existing stock of money balances to acquire even more bonds. Hence, at these higher interest rates the supply of loanable funds lies to the right of the saving line. The exact values of \( r_1 \) and \( r_2 \) and the degree of responsiveness (elasticity) of the supply of loanable funds to the interest rate are important practical matters to which much advanced research is devoted. In our simple model, however, our task is more modest, viz. to understand in a general way how and why the supply of loanable funds provided by the household sector responds to the interest rate.

THE SUPPLY OF LOANABLE FUNDS FROM COMMERCIAL BANKS

So far we have assumed that the supply of loanable funds came entirely from household saving. We must now take into account the fact that commercial banks also supply loanable funds. Commercial banks are distinguished from other financial institutions such as savings and loan associations in that commercial banks can actually add to the supply of loanable funds by creating money. If a firm wishes to borrow from a commercial bank, the bank can accommodate the firm by simply giving the firm a checking account equal to the amount of the loan. On the other hand, before any other financial institution can make a loan, they must be certain that funds have
previously been deposited with them. We define $\Delta \left( \frac{M}{P} \right)$ as the change in the money supply provided by the commercial banking system. In the figure 52 we show the supply of loanable funds originating from commercial banks (equal to $\Delta \left( \frac{M}{P} \right)$) under two different assumptions: 1) the quantity supplied $\Delta \left( \frac{M}{P} \right)$ is entirely unresponsive to the interest rate, and 2) the quantity supplied $\Delta \left( \frac{M^*}{P} \right)$ responds positively to the interest rate. Exactly how the supply of loanable funds originating from commercial banks does respond to the rate of interest is another question of great practical importance which need not concern us here. Under either assumption for the commercial banking system the supply of loanable funds to the market as a whole remains positively sloped with respect to the rate of interest. The market supply curve for loanable funds is the sum of the supply curves of the household sector (non-bank public) and the commercial banking system. The market's supply curve of loanable funds will have a positive slope with respect to rate of interest as long as the households sector's supply curve is positively sloped.
Figure 52
Supply of Loanable Funds From Commercial Banks
Footnote to Chapter 20.

1Of course, some families spend more than the amount of their current income on commodities by either 1) borrowing, or 2) reducing their holdings of other assets, viz. bonds and/or money. In the aggregate, however, families purchase an amount of commodities which falls short of their current income so that, on the average, families are adding to their holdings of money and/or bonds, as previously noted.
CHAPTER 21.
The Determination of the Aggregate Price Level, Income, and Rate of Interest.

INTRODUCTION

We are now ready to undertake the explanation of macroeconomic equilibrium within the framework of our loanable funds model. We shall explain the determination of the equilibrium values of our variables by examining the interaction among our three markets (commodities, money, and bonds) as these interactions are played out upon the stage of the loanable funds market.

HOW EQUILIBRIUM IS ATTAINED

Equilibrium is the result of millions of individual decisions made simultaneously by various entities in an economy. These decisions determine: first, the supply of output and the demand for output in the form of consumption and investment goods; second, the supply and demand for money balances; and finally the supply and demand for new earning assets. The interaction among these three markets (commodities, bonds, and money) is what determines equilibrium. Equilibrium for the economy as a whole can be attained only when all three of these markets are in equilibrium. Therefore, at the equilibrium values of the price level, output and interest rate and only at those values are the quantities demanded and supplied equal in the markets for commodities, bonds, and money. Since we have already discussed the loanable funds market, let us briefly con-
Consider the demand and supply relationships in the commodity market.

**THE COMMODITY MARKET**

The commodity market encompasses all the individual markets for goods and services of an economy upon which we focused in microeconomics. It would be a virtually impossible task to examine each one of these markets, even for a small economy. Actually, such an effort is unnecessary and our theory abstracts from all the details of individual markets and firms to make one summary statement for commodity market equilibrium. More sophisticated macroeconomic models are richer in their microeconomic foundations but, as you might imagine, remain at a far greater level of abstraction than the theory of individual markets.

We will divide all commodities produced into two types, viz. consumption and investment commodities. Consumption commodities consist of all the goods and services produced for household use (peanut butter, clothing, etc.). There are many ways we might classify these various consumption commodities ("type" of good, durability, etc.) but instead we will merely add the market value of all consumption commodities produced into one grand total. Investment commodities consist of all those goods and services which we will use to produce other commodities in the future and include newly constructed buildings, pieces of capital equipment, and additions to firms' inventories. Ultimately, all economic effort is directed to
the production of consumption commodities so the production
and use of investment commodities ("capital") is sometimes
described as the "roundabout means of production." But for
any given period of time, all the commodities produced can be
classified into either consumption or investment commodities.

Estimates are made of the value of total production of all
commodities. This summary figure is known as the Gross National
Product (GNP) and it can be thought of as either the value of
all current output\(^2\) or, equivalently, the sum of all current
incomes received by factor of productions.\(^3\) Households,
as we have seen, do not on the average spend all of their in­
come on consumption but save a portion. Thus we divided the
income earned by households (\(Y\)) into two parts, consumption
\((C)\) and personal savings (\(s\)) expressed as

\[
Y = C + s
\]

But this income total is equivalent to the value of current
output (which we will also call \(Y\)) expressed as

\[
Y = C + I
\]

Since the definitions of income and output measure exactly
the same thing (\(Y\)), it must always be true that the measured
value of saving equals the measured value of investment. How­
ever, this does not mean that all households and firms are satis­
\(fied with the current value of \(Y\) since some firms may in fact
be selling less output than they desire to sell. If this is
the case, some of the unsold output will be unwillingly added
to firms' inventories and they would have an incentive to change
production in the next time period. Equilibrium \( Y \) is attained only when no firm experiences undesired changes in investing. If we now think of \( s \) and \( I \) as representing desired saving and desired investment, the condition for equilibrium in the commodity market is that desired saving must equal desired investment. Only then will the total supply of output be equal to the total demand for output. In the figure 53 we graph the amount of saving at full employment income, \( s(Y_F) \). Similarly the desired rate of investment, \( I \), is graphed versus the rate of interest. We see that when the rate of interest is \( r_o \) and output equals \( Y_F \), that saving equals investment. Hence at that point (A) the commodity market is in a state of equilibrium.

**MODUS OPERANDI FOR ANALYSIS OF EQUILIBRIUM**

Now that we have examined the loanable funds and the commodity markets in some detail, we can consider how their joint interaction determines the equilibrium levels of our variables. At this grand equilibrium the bond market, the commodity, and the money market will all be in individual equilibria, i.e. the quantities supplied and demanded in each market will be equal.

Actually, our task of analysis will be facilitated by the recognition that if two of the sectors are simultaneously at their equilibria, then the third is "automatically" at equilibrium, too. The explanation is simple. If there are only three markets in which spending can take place, any disequilibrium (excess supply or demand) in one market must be "balanced
Figure 53
Commodity Market Equilibrium
by a compensating (excess demand or supply) disequilibrium in the other two markets. For example, if there is an excess demand for commodities, there must be an excess supply in the markets for bonds and money (taken together). However, if two of the market sectors are individually in equilibrium, so must be the third sector since there is no possibility of excess supply or demand in the first two sectors to offset any disequilibrium in the remaining sector. As previously noted, this conditional guarantee of third sector equilibrium will facilitate our analysis since it becomes necessary only to examine any two of the three markets to determine whether overall equilibrium is attained. Which two markets we consider and which is omitted from explicit examination is merely a matter of expositional convenience. We shall choose to examine the interaction between the commodity and loanable funds markets. An equilibrium output \((YF)\) and interest rate \(r_E\) are shown in the figure 54. This is an output and interest rate combination where equilibrium is attained for the commodity and loanable funds markets (and thus the money market as well).

Let us review the characteristics of this equilibrium. First, the quantities supplied and demanded of loanable funds are equal only at interest rate \(r_E\) since at that interest rate the \(S\) and \(D\) curves intersect one another. The commodity market is also in equilibrium implying that the quantity of investment goods produced is equal to household saving (the difference between income and desired consumption), i.e. the \(S\) line.
Simultaneous Equilibrium of Loanable Funds and Commodities Markets
intersects the i line. Finally, we can infer that the quantities supplied and demanded of money are also equal to each other at $r^E$ and $Y_F$.

While choosing a full employment output, we sidestep the controversy concerning whether or not an economy can actually attain full employment. One group of economists believes that there is a natural tendency for the economy to achieve a commodity market equilibrium which also coincides with the full utilization of the resources (labor, land, capital) available to the economy. They believe that if disequilibrium exists in different markets, prices will rise or fall fast enough so that demands and supplies are again equal. Of course, full employment is taken to mean some high level of utilization of resources corresponding to a low level of unemployment, say 4%. The 4% figure represents the fact that resources cannot be expected to move from one activity to another instantaneously. Some time must be spent in the normal transition from one job to another. These economists do not believe that an economy can experience commodity market equilibrium if the unemployment rate is above this full employment level. They believe that excess supply would prevail in the labor market with an unemployment rate above the full employment level and that a reduction in the wage rate along with an expansion in real output and perhaps a decrease in the price of goods are necessary to reduce the unemployment rate and reach the commodity market equilibrium. This position is an old one in economics which we
could call "The Classical Position" although today it is more commonly associated with the intellectual descendents of that school of thought, the "Neo-Classical" theorists.

An alternative theory associated with the followers of the British economist John Maynard Keynes accepts the Classical or Neo-Classical argument as a long-run tendency of an economy. However, the "Keynesians" feel that this tendency for the economy to produce at a full employment commodity market equilibrium may take many years to materialize. They believe prices respond much more slowly to disequilibrium than do the Neo-Classicals. "Many years" is usually held to be too long to wait for full employment by the Keynesians who propose certain policies to facilitate the more prompt attainment of full employment. In the meantime the Keynesians believe an economy can reach a commodity market equilibrium (a rate of output that remains the same, given underlying conditions) with a relatively high level of unemployment. In this belief and the policy implications that follow from it, the Keynesian position differs from the Classical or Neo-Classical theory.

A casual examination of the course of the unemployment rate in the United States might seem to lend some support to the Keynesian position. The unemployment rate has varied widely, most years ranging from 4-7% of the labor force. Whatever the full employment unemployment rate is thought to be, the actual unemployment rate is usually significantly different, so one might conclude that the economy seldom attains the full
employment equilibrium output. Certainly if 4% is taken to be the full employment rate, full employment has been attained in only about one year in every four in this century in the United States.

The author is persuaded, however, that a closer look at the evidence lends more support to the Neo-Classical position than to the Keynesian. The issue is still in vigorous debate among economists and the arguments are quite technical. Having discussed this important issue, we shall indicate whether we are using Neo-Classical or Keynesian assumptions whenever we apply our simple model to a macroeconomic problem.

THE ATTAINMENT OF EQUILIBRIUM SEEN AS THE ADJUSTMENT TO DISEQUILIBRIUM CONDITIONS

We now describe how an economy reacts to conditions that disturb its initial equilibrium position. What happens to price output, and the rate of interest when some sector of the economy radically changes its plans concerning either the demand or supply of some good? Suppose the economy is initially at full employment equilibrium as described in figure 54 in the section "Modus Operandi" of this chapter and in the figure 55 where the subscripts, \( (S_1, D_1) \) refer to the original state of affairs. Assume that because of adverse expectations about the future that there is a leftward shift in the demand for loanable funds from \( D_1 \) to \( D_2 \). This shift implies a reduced willingness (at every rate of interest) on the part of firms to buy commodities for the purpose of investing in new capital. Initially, down-
A Reduction in Investment Demand
ward pressure would be exerted upon the rate of interest since at the old interest rate $r_1 \Sigma_1 > D_2$. Prices in financial markets respond promptly to discrepancies between supply and demand so that soon we would observe interest rate $r_2$ where $\Sigma_1 = D_2$. The reduction in the interest rate has brought the loanable funds market once again into equilibrium. The decline in the interest rate induces some investors to increase the quantity of loanable funds demanded for the purpose of purchasing more investment commodities. Even though the future appears "less rosy" to all investors (as evidenced by the shift in the entire D schedule), a sufficiently lower interest rate increases the attractiveness of some investment projects enough to bring the supply and demand for loanable funds back into equality at $r_2$. Furthermore, the decline in the interest rate induces households to add to their money balances and to purchase fewer bonds.

But what about the state of affairs in the commodity market? At $r_2$ saving exceeds investment so that there is excess supply in the commodity market. Producers of goods and services find themselves with growing inventories since production exceeds sales. The rate of output $Y_p$ no longer equates supply and demand in the commodity market. At the intersection of $S_1$ and $D_2$ in figure 56 we can predict that the economy will experience a recession (a decline in the level of business activity) which leads to a reduction in output and/or the price level. Under the Keynesian assumption, producers
will start to lower production, leading to a decline in income. For households, this process of lower production comes to a halt when the saving line has shifted far enough to the left so that it again intersects the investment curve. At this point, the aggregate demand and supply of commodities is again equal.

While we cannot indicate precisely the new equilibrium position under the Keynesian assumptions (except to say that it occurs at a lower output) we can specify the new equilibrium under Neo-Classical assumptions. Recall that this theory holds that prices adjust promptly to disequilibria. When firms observe undesired additions to inventories at the intersection of $\Sigma_1$ and $D_2$, they respond not by lowering production but by decreasing their prices. Overall equilibrium is again achieved when the supply of loanable funds shifts to $\Sigma_2$ as in figure 56. This shifting of $\Sigma$ is the result of the following events:

1) the excess supply in the commodity market puts downward pressure on the price level.

2) the effect of lower prices is to shift $\Sigma_1$ towards $\Sigma_2$ since the real quantity of money supplied by banks $\Delta \left( \frac{M}{P} \right)$ increases as the price level falls.

During the period of disequilibrium when prices are falling, the rightward movement of the supply of loanable funds causes the interest rate to continue to decline below $r_2$. Again we observe the power of the interest rate in helping to produce overall equilibrium. In this case the continuing decline in $r$
Figure 56
Re-Attainment of Equilibrium
helps to stimulate the investment component of the demand for goods and services by investors to move down along the $D_2$ curve. The fall in the price level and the interest rate continue until $r_3$ is reached where we see that $E_2 = D_2$ and $s = i$. At this point the demand and supply for commodities, loanable funds, and money balances are again in equilibrium. At the new equilibrium position $i$ is the same as it had been before. investor expectations had taken a turn for the worse. Although the future may be "less rosy" to investors, a lower price can help to offset such adverse expectations.

Our analysis can be used to describe the outcome on prices, output, and the rate of interest under either Keynesian or Neo-Classical assumptions. However, economists who find the Keynesian position on this issue more persuasive are inclined toward policy actions, chiefly by the Federal government to speed the adjustment process. Neo-Classical economists are less enthusiastic about macroeconomic policy actions--so called "discretionary" actions. It is to the question of macroeconomic policy that we now turn.

Footnotes to Chapter 21.

2Actual figures published by the Federal government are quite detailed and divide GNP into four parts, viz. consumption, investment, government, and exports-imports as components of GNP. Actually, the commodities purchased by governments and foreigners can be included in our two-commodity framework depending upon whether the commodity is of the consumption or investment type. In the next chapter we shall consider the effect government purchases of commodities explicitly.

3All production leads to income for someone whether the form of income is wages, dividends, or profits. Ultimately, all this income accrues to households.

4This equivalence is valid as long as we ignore depreciation charges and indirect business taxes.
INTRODUCTION

In this chapter we introduce government as an important entity in the macroeconomic scene. Governments (Federal, State, and Local levels) account for a large fraction of total spending in the economy. A large part of government spending is financed through taxation which can be thought of as a redirection of private spending. These functions of government which have become more important over time in both an absolute and relative sense can easily be incorporated into our model. But macroeconomic policy actions undertaken by government (we speak here almost exclusively of the Federal level) are of even greater significance to macroeconomic theory than the sheer size of government activities.

THE EVOLUTION OF MACROECONOMIC POLICY

The scope of government economic activity has expanded greatly in the United States, chiefly in the last 40 years. This expansion itself has been of great economic significance because it has caused a major reallocation of resources from the private sector to the public sector. Another important effect of this expansion in the levels of government spending and taxation has been its impact on the overall stability of the aggregate economy. But exactly how the enlarged scope of government has affected economic stability (reflected in the
degree of fluctuation in the activity of the private sector) is heatedly debated among economists. The majority of economists would probably hold that the expansion of government has, on balance, done more to promote economic stability than to undermine it.

Governments have always spent and taxed, if on a smaller scale than today. A more radical departure in government economic activity has been the assumption of the function of promoting economic stabilization and growth through actively engaging in macroeconomic policies. This commitment was first expressed in the Full Employment Act of 1946. Most people now accept this role for government as appropriate and commonplace. Yet, a coherent and comprehensive macroeconomic policy, supported and fostered by economic theory, has only recently been developed. In fact, the growth of active macroeconomic policy prescriptions largely parallels the period of the vast expansion in the scope of government spending and taxation. This parallelism is no coincidence. The size of government's economic role in spending and taxation on the one hand, and the adoption of active macroeconomic policies on the other, are, to a large degree, developments that have been mutually reinforcing.

The Great Depression of the 1930's marks the turning point for most of the developments we have been discussing. This period of depressed business conditions marked by high unemployment, low outputs and income, in short, economic stagnation,
was probably the most severe in American history. As is usual for periods of declining prosperity, appeals for government to "do something" were heard. In large part, this was a direct call for help by those affected—the unemployed, those who had lost wealth—for relief of their symptoms. In addition, however, calls for some more fundamental changes in the government's economic policy were made. Socialism, i.e. government ownership and direction of economic activity, was proposed by many. A more moderate course, which is widely believed to be the one adopted, was to have government spending compensate for the apparent "shortage" in private spending. Some believed that government would only have to initiate a business revival ("prime the pump") while others believed that the government would have to more or less continuously supplement private spending to ensure prosperity. At any rate, the government's increased responsibility and role for promoting economic stability had changed in a very fundamental way. Some proposals for new legislative and regulatory changes (e.g. recognition of trade unions, government supervision of stock markets) mark a further extension of government activity presumably designed to promote economic activity, too. Whatever their effects, however, these actions were specific ones with numerous precendents (as the Federal Reserve System) as opposed to the fundamental change in the general role of government reflected in increased levels of spending and taxation.
At the same time that the government was widening its economic responsibilities, a monumental development in economic theory which lent great support to governments' new posture came to light. A British economist John Maynard Keynes, heretofore a supporter of the Classical macroeconomic position, revealed a revolutionary theory in his book *The General Theory of Output, Employment, and Prices* (1936). The classical position held that government spending and taxation had no effect on the level of business activity. Attempts on the part of the Government to "manipulate" the level of demand via changes in spending or taxation were held to be useless at best or destabilizing at worst. Keynes rejected the notion that the best course for government was one of "neutrality" (i.e. an always balanced budget) and asserted that governments could and should promote economic stabilization through adjusting the levels of government spending and taxation.

It's not really very difficult to see why Keynes' theory won rapid acceptance among economists. Classical theory was held to predict an "automatic" full employment equilibrium, a prediction obviously at variance with the depressed business conditions that economies throughout the world had been experiencing since 1929. The return to very high levels of output and employment during World War II coincided, of course, with a huge increase in government spending. Many people saw this experience as support for the effectiveness of governments' new role in promoting economic stabilization. Finally, a rela-
tive mildness of post-war business fluctuations, in contrast with earlier periods, has made for further acceptance of these policies. The proposition that governments can and should stabilize business conditions may today be more strongly accepted by laymen than professional economists. We now turn to Keynesian and neo-classical macroeconomic theory and their respective analyses of the role of the government in the economy.
KEYNESIAN AND NEO-CLASSICAL MACROECONOMIC THEORY

Much of our earlier discussion has touched upon various elements of Keynesian and Neo-classical theory. In this section we shall fully incorporate these theories into the basic equilibrium model we developed in the last chapter. The theories presented are competing explanations of macroeconomic phenomena. Ultimately, these differences in the theories can be resolved only by empirical testing.

We begin by presenting the Keynesian theory in the framework of our 3 market equilibrium model. Government spending enters in both the commodity and bond markets (as well as the money market which we shall not consider explicitly, as usual). Commodity production now consists of three kinds of output: consumption, investment, and government expenditures expressed as

\[ Y = C + I + G \]

If \( T \) represents the amount of taxes paid by the private sector, the equilibrium condition in the commodity market becomes

\[ s + T = I + G \]

since \( Y - C = I + G = \) and \( s + T = Y - C \)

i.e. income not spent on consumption goods can be used for payment of taxes and saving.

Keynesian theory states that the equilibrium level of income can be increased by operating directly upon the level of expenditures in the commodity market. Either government expenditures \( (G) \) can be increased or taxes \( (T) \) reduced to in-
crease the equilibrium level of output. In principle, the increase in \( g \) or the reduction in \( T \) (which indirectly increases \( C \) and/or \( I \)) operate in the same fashion so we shall discuss only the simpler case of increasing \( G \). Changes in the level of \( G \) and \( T \) designed to change the level of demand in the commodity market are the two components of the "discretionary" or "compensatory" fiscal policy prescribed by Keynesians.

A crucial assumption of the Keynesian theory is that the economy can reach an overall equilibrium with an amount of unemployment above the full employment level. The classical and neo-classical contention is that such a situation amounts to excess supply in the labor market and could not, therefore, be an equilibrium position. With this in mind, consider figure 57 where \( Y \) is the "equilibrium" output and unemployment is above the full-employment level (the Keynesian assumption).

For simplicity, assume the government starts out with a balanced budget, \( g = t \). This means that it is financing its expenditures on commodities with sufficient revenue generated by its tax collections. If, however, the government were to have a budget deficit, \( g > t \), it must somehow obtain the funds to finance its budget deficit. We assume that when the government finds itself in such a position, it will become a net demander of loanable funds. It will issue earning assets to the extent of \( (g - t) \) in the loanable funds market and so become a borrower in the loanable funds market. Our demand
Figure 57
The Keynesian View of the Effect of an Increase in Government Spending
curve for loanable funds will then have two parts: i, investor demand, and \((g - t)\), net government demand. Of course, this is not the only way in which a government could finance a deficit. An alternative and commonly used method is the printing of money. Such a financing scheme is used almost continuously by many Latin American governments. We shall, however, restrict ourselves to the case of government borrowing in the loanable funds market.

In figure 57, we see that for a rate of interest of \(r_1\) and income of \(Y_1\) that \(\Sigma_1 = D_1\) (loanable funds market is in equilibrium) and that \(s_1 = i_1 + g_1 - t_1\) or \(s_1 + t_1 = i_1 + g_1\) (demand and supply in the commodity market are equal).

A Keynesian cure for this undesirably low output equilibrium is an increase in government expenditures. Since \(g_1\) initially equalled \(t_1\), the effect of the increase in \(g\) from \(g_1\) to \(g_2\) will be to create a budget deficit. The government must now enter the loanable funds market as a net demander. Hence, as a result the demand curve shifts from \(D_1\) to \(D_2\). In order to equate the new demand and supply curve of loanable funds the interest rate must rise to \(r_2\). But at this higher interest rate the state of affairs has changed in the commodity market substantially. At the new intersection point it is easy to verify that \(i_1 + g_2 > s_1 + t_1\). This means that aggregate demand exceeds aggregate supply in the commodity market. As far as producers of goods and services are concerned, their sales exceed their production. To meet the increased sales,
producers will have an incentive to expand employment and output. At the intersection of $\Sigma_1$ and $D_2$, therefore, we can predict that the economy is in for a period of expansion in aggregate production. The basic motivating force behind this is of course the active fiscal policy adopted by government.

Keynesians believe that the rate of output can be increased because of substantial unemployment, a kind of "slack" in the market—"excess" or unused capacity to produce. Classical and neo-classical economists maintain that such situations are only temporary phenomena (at least in the absence of government policy actions) and that the economy will inherently move close to the full employment level. These economists maintain that the Keynesian policy measures might, at best, speed the economy's natural process of adjustment, and, more likely, will have almost no impact on the aggregate rate of production.

In the figure 58 we illustrate the neo-classical interpretation of Keynesian policy prescriptions. The economy is assumed to be in full employment equilibrium in the commodity market. Starting from a position when $g_1 = t_1$ (balanced budget) let the government increase its spending on goods and services. As before, this will produce a deficit, forcing the government to finance the deficit by entering as a net demander of loanable funds. Initially, the loanable funds market equilibrium shifts to the right where $D_2$ and $\Sigma_1$ intersect. Even though their market is in equilibrium at $r_2$, we can immediately infer that there is excess demand in the commodity market since $(i_1 +$
Figure 58
The Neo-Classical View of an Increase in Government Spending
must exceed the unchanged amount of $s_1 + t_1$. Again firms experience sales greater than production but, unlike the previous case, firms cannot increase production. Their only recourse in the presence of excess demand for commodities is to raise prices. As they do this, $\Delta \left( \frac{M}{P} \right)$, the supply of loanable funds created by commercial banks, shrinks. This causes the $\Sigma$ curve to shift leftwards until overall equilibrium is restored at interest rate $r_3$ when $\Sigma_2 = D_2$. At that point the commodity market is also in equilibrium since $s_1 + t_1 = i_1 + g_2$. As a result of changing prices, firms eventually find that sales have declined sufficiently until they again match current production. At the end of the adjustment process only the price level and the rate of interest have risen with income remaining unchanged. Saving remains the same but it now is apportioned between investment and government expenditure. The increase in government spending has only "crowded out" some private investment without changing total output. In fact, this change in the composition of aggregate demand, away from private investment and towards government expenditure, has been accomplished via an increase in that important economic variable, the rate of interest.

We see that the fundamental difference between the Keynesian and neo-classical positions originates in their acceptance or rejection of the possibility of aggregate equilibrium accompanied by substantial unemployment. Keynesians believe such a position could persist for years and that the continuation
of such conditions is tantamount to equilibrium. Therefore, they believe that manipulating government spending and/or taxation can improve the economy's overall performance in terms of output and employment. The problem, as Keynesians view it, is a matter of determining the exact relationships between the targets (employment, output, etc.) and policy variables (spending, taxation, levels of borrowing and money creation, etc.).

The Neo-classical position maintains that the economy has a strong natural tendency toward a full employment equilibrium. They believe that market forces can adjust to changes in the levels of supply and demand, eliminating excess supplies and demands through the price mechanism in individual markets and, thus, for the economy as a whole. They concede that an all-knowing fiscal policy maker could, in theory, speed the economy's adjustment process by compensating for the effects of changes in the overall levels of supply and demand for the economy arising in the private sector. However, neo-classical economists believe that the lack of sufficient knowledge about fiscal policy timing and impact and the political problems associated with its implementation make unrealizable any theoretical benefits fiscal policy might confer in terms of economic stabilization. The safest course, say these economists, is to design fiscal policy to be neutral with regard to its impact on the overall level of spending and to judge any proposals regarding government spending and taxation on their intrinsic merits (effects upon the allocation of resources
between the public and private sector, effect upon the distribution of income).

Having presented the competing theories, the next logical step is to consider which of them seems to provide a fuller explanation of the past performance of the economy and a more reliable guide to the future. Of course, we cannot really resolve this issue for you, but will attempt only to consider two important periods in recent economic history which offered a dramatic setting to test these theories against each other.

MACROECONOMIC POLICY IN THE 1930's AND 1960's

In 1929 the United States was enjoying prosperous business conditions as reflected in a low unemployment rate and a rapidly growing level of income. Stock market prices, mirroring the belief that the business cycle was a thing of the past, rose almost four-fold from 1925 to 1929. But, of course, starting in the fourth quarter of 1929 our economy experienced the onset of a most severe depression which reached its nadir in 1932. At this time per capita income had fallen by about one-third of the pre-Depression level and unemployment was estimated at one-fourth of the labor force or more. How did government macroeconomic policy respond in the Depression? Did it prevent an even worse experience or did it perhaps contribute to the severity of the Depression?

In terms of our loanable funds diagram, the U.S. economy was jolted by "outside events" in 1929 which caused an initial leftward shift in the demand for loanable funds from \( D_1 \) to \( D_2 \).
as shown in figure 59. According to the classical explanation, our economy should have experienced a brief and fairly mild recession, followed by a quick period of recovery. In fact, no such thing occurred. Instead, the "period of adjustment" turned out to be protracted and severe. On the surface, this observation appears to destroy the classical analysis of macroeconomic equilibrium.

Before we make a final judgment on classical analysis, let us examine another stabilization policy available to governments, monetary policy. Monetary policy operates on the supply curve of loanable funds by inducing the commercial banking system to either expand or contract \( \Delta \left( \frac{M}{P} \right) \).

In retrospect we can see quite clearly what the government should have done to compensate for the unfavorable shift in private demand. The government should have either (A) increased the level of its spending (purposely causing a deficit) thus shifting \( D_2 \) back toward \( D_1 \), or (B) increased the money supply shifting \( \Sigma_1 \) toward \( \Sigma_2 \), or (C) undertaken some combination of these two actions. We have already seen how fiscal policy can help in resting aggregate demand to a full employment level. Monetary policy can also perform the same trick but it does so in an indirect fashion. Returning to the figure 59, we see at the intersection of \( \Sigma_1 \) and \( D_2 \) that demand for goods and services is less than supply. If the monetary authority pro-
Figure 59
Macroeconomic Conditions and Policy Actions, 1929-33
duces conditions which make it profitable for the commercial banking system to expand, then $\Sigma_1$ will shift to $\Sigma_2$.

As a result, the rate of interest will fall to $r_2$ to again make the supply and demand for loanable funds equal. But notice as the rate of interest falls, investors will have an incentive to increase their spending on investment goods. In short, if the monetary authorities are smart enough, they can counteract the leftward shift in the demand for loanable funds by making it attractive for investors to move down along $D_2$ via lower interest rates. Assuming they shift $\Sigma_1$ to $\Sigma_2$ so that $\Sigma_2$ intersects $D_2$ at $r_3$, then $s + t_1 = i_1 + g_1$ and the initial excess supply of commodities will completely disappear. Hence, producers of goods and services will have no incentive to change production or prices since sales will again match full employment production. Policy makers had the capacity to take these actions. Let us examine the policy guides they followed and the policy actions they enacted.

Fiscal policy (expenditures and taxes) was essentially neutral during the 1930's. Expenditures rose but the tax structure more than kept pace so that the overall effect of fiscal policy, taking all three levels of government together, was not stimulating in its effect upon aggregate demand. For fiscal policy to have been neutral, despite the image of pump
priming and a frenzy of government activity during the Depression is understandable once it is recalled that fiscal conservatism was the accepted dogma of policy makers and academic economists. A balanced budget was seen as the appropriate policy and the actual deficits experienced were small and largely undesired ones.

Monetary policy was much more crucial to the course of aggregate products in the 1930's. With the benefit of hindsight, we can see that the appropriate response of policy makers (specifically the Federal Reserve) should have been to cause an increase in the money stock, thereby shifting $\Sigma_1$ toward $\Sigma_2$. Actually, the policy adopted was exactly the opposite. The money supply fell by about one third from 1929 to 1933. In terms of the figure 59, the monetary authorities caused $\Sigma_1$ to shift leftwards to $\Sigma_3$. Instead of trying to stimulate aggregate demand, the monetary authorities made matters worse by facing a decline in demand. For producers in the economy, excess aggregate commodity supply became larger. No wonder that prices and output started to plunge dramatically. To each individual producer it seemed the most plausible thing to do since sales were declining so rapidly in relation to production. In short, the monetary authorities created an
environment that "told" firms to cut output and prices. On whose shoulders can we place the blame for the Depression? A defective private enterprise system or a blundering Federal Reserve? Empirical investigation seems to suggest the latter. The performance of the monetary authority seems explicable only in terms of its desire to maintain the interest rate at a level like $r_1$. During the course of the 1930's the Federal Reserve believed that its actions had effectively reduced the severity of the Depression. In all fairness, we should note that only in recent years has the quantity of money in circulation come to be thought of as a key instrument of monetary policy. Other central banks in the world behaved similarly, compounding the impact of inappropriate policy on an international level. We shall soon see, however, that the Federal Reserve's fascination with the interest rate is not a thing of the past.

In the period between the Great Depression and the Great Society, fiscal policy was in the spotlight. Ironically, the Great Depression which was much exacerbated by the actions of the Federal Reserve was taken as evidence of the relative impotence of monetary policy.\textsuperscript{12} Fiscal policy was given credit for the unparalleled prosperity of the W.W.II period and the relative mildness of post-war business slumps. The regard for fiscal policy and the Keynesian position reached a summit by about 1965 in the wake of the Tax Cut of 1964 and what was taken to be highly successful implementation of the ideas embodied in \textit{The General Theory}. 
We shall not examine this long period in any detail before turning to the post-1965 which has proved a sobering one for most participants. We shall note, however, that a much different interpretation of the 25 years just surveyed seems roughly consistent with the evidence. Fiscal policy and monetary policy were almost always operating in the direction for most of this long period (as measured by adherents of the Keynesian and monetarist approaches, respectively). While fiscal policy received the credit for this period of prosperity and growth, it is not entirely clear that it can be inferred that discretionary fiscal policy can be relied upon to improve economic performance as a general rule. This skepticism has been strongly influenced by the experience of the American economy since 1965.

In 1965 the American economy was essentially at full employment. Unemployment had fallen to about 4% and remained about the same until 1969 when it began to rise. The situation in 1965 and the policy actions of the following years can be seen in the figure 60 where $D_1$ and $\Sigma_1$ represent the initial positions in 1965.

In 1965, a substantial increase in government spending associated with the escalation of the conflict in Indochina caused the demand curve for loanable funds to shift toward $D_2$. At $r_2$ where $\Sigma_1$ intersects $D_2$, we observe excess aggregate commodity demand and can predict a resulting rise in the price level. Clearly, the appropriate response of monetary policy
Figure 60
Macroeconomic Conditions and Policy Actions, 1965-69
to this fiscal policy action, if it desires to return the economy to equilibrium without an accompanying rise in the price level, should have been to restrict the growth of the money supply, shifting the $\Sigma_1$ curve toward $\Sigma_3$. A moment's reflection should indicate to you that this situation is the mirror image of the one we discussed in the context of the 1929 scenario. In the 1965 case, contracting monetary policy successfully dampens the latent tendency for the price level to rise via the higher interest rate produced at the intersection of $\Sigma_3$ and $D_2$. The higher interest rate sufficiently reduces the investment component of aggregate commodity demand so that the sum of the lowered investment demand and increased government demand leave the total amount of aggregate demand unchanged.

In short, contracting monetary policy recomposes aggregate demand away from the private sector and towards the government sector.

In a world of scarcity someone must give up his plans if someone else wants more goods and services. Unfortunately, the response of the monetary authority was again exactly the opposite of what it should have been. The money supply (seen as a shift from $\Sigma_1$ to $\Sigma_2$) was allowed to grow very rapidly from 1965-1969. Already at full employment, the economy's output of commodities had little room to rise and the result was a substantial increase in the rate of inflation. Again, this perverse response of the monetary authority seems best explained by a desire to prevent the interest rate from changing. The monetary authorities apparently decided to keep private demand
constant in the face of increased government demand. They
did not want anyone to "pay" for Viet Nam. This effort, by
the way, failed to keep interest rates from rising to the
highest levels since the Civil War by 1970.

In 1969 the economy faced a choice: either a prospect
of further acceleration in the rate of inflation or a rever­
sal of the inflationary policies conducted by the Federal Re­
serve for the years 1965-69. The latter course was chosen and,
as one would expect, an increase in the unemployment rate re­s­
tulted. By 1969, inflationary expectations were so well es­
established that even a mild bit of monetary restraint would
have been felt as very dampening upon the rate of output and
employment but not immediately on prices. As is often the
case, the price level responded with a significant lag to the
conditions in the commodity market. Through 1970-72 the rate
of inflation did diminish substantially (although, of course,
prices did continue to rise) as a delayed response to the re­
duction in the level of aggregate demand. By the way, the
reduction in the rate of inflation experienced between 1970
and 1972 took place mostly before the adoption of the so-called
New Economic Policy in August, 1971. By late 1972, the rate
of inflation had again begun to rise and very high monthly
rates were experienced in early 1973 (the highest monthly rates
since 1951). Whether another cycle of inflation is underway
is not entirely clear but it seems quite apparent that the
behavior of the price level in this latest period is explained
as usual by monetary policy movements during 1972.

The performance of the economy since 1965 has led many economists toward the neo-classical explanation. Within this period, there were a number of instances when monetary and fiscal policy were pointing in opposing directions. Whatever the unfortunate consequences to the economy, this series of "confrontations" allowed economists to discriminate between the effects of fiscal vs. monetary policy more effectively than the earlier periods we have discussed. In general, monetary policy indicators proved more reliable. Despite this aspect of the triumph of the neo-classical approach, monetarists view the experience of the post-1965 period as further support for the proposition that discretionary macroeconomic policy, whether through fiscal or monetary instruments, proves to be more harmful than stabilizing for overall economic performance.
Footnotes to Chapter 22.

5 We cannot be sure because of the lack of data for earlier periods. Economic historians believe that the Depression of the 1830's was also very severe.

6 FDR advocated a balanced budget (equality between government expenditures and taxes) during the 1932 campaign. Once in office, his administration incurred deficits (expenditures exceeded revenues) consistently. Some economists believe that these Federal Government deficits were largely offset by the actions of state and local governments, however, during the 1930's.

7 The problem was one of insufficient demand for goods. The physical capacity of the economy to produce was not much different in 1933 than in 1928.

8 We could easily add a fourth type of expenditure—exports—but for simplicity will omit it. For the U.S. (but not so for many countries) exports have relatively little effect upon the overall level of output.

9 First of all, economists are far from any consensus as to which of these explanations is the better although a majority probably leans towards the Keynesian view. For about two decades after its initial acceptance, the Keynesian theory triumphed almost completely over the old classical explanation. A re-working of the old theory into the neo-classical explanation gained many "converts" and caused even more to begin to question and modify the Keynesian explanation. The experience of the late 1960's described in the following section was quite significant in lending support and gaining adherents to the neo-classical position.

10 Since the Great Depression began at about the same time as the Stock Market Crash of 1929, many people assume that there was a direct cause and effect relationship between the latter and the former. Actually, the stock market values reflect current and prospective business conditions so any cause and effect relationship runs in the opposite direction. Furthermore, stock prices rallied after reaching a low in November 1929 and by April 1930 had recovered over half of the losses of September and October 1929. Stock prices fell as the Depression deepened, reaching their low in July 1932 when the Dow-Jones average stood at about a tenth of its 1929 peak. The point is that neither the economy nor the stock market "crashed" all at once in late 1929 as has come to be believed.
The importance of monetary policy in affecting the course of events lends strong support to another aspect of the neo-classical position vis-a-vis the Keynesian. The former generally holds that monetary (rather than fiscal) policy is much more potent in its effect upon the overall level of economic activity.

The reinterpretation of the significance of monetary policy we just discussed was one of the major contributions of those economists whose work has come to be identified as the neo-classical position.