

industry may be technologically progressive and have an enviable record of providing products of increasing quality at reasonable prices. Therefore, if the industry has served society well and engaged in no anticompetitive practices, it should not be accused of antitrust violation simply because it is highly concentrated. Why use antitrust to penalize efficient, well-managed firms?

Since the Alcoa decision of 1945, the courts have reverted to the rule of reason. The general sentiment among antitrust economists and those responsible for enforcing the antitrust laws has also swung away from the strict structuralist view. For example, in 1982 the government dropped its 13-year-long monopolization case against IBM on the grounds that IBM had not unreasonably restrained trade.

Defining the Market

Court decisions involving existing market power often turn on the issue of the size of the market share of the dominant firm. If the market is defined broadly, then the firm's market share will appear to be small. Conversely, if the market is defined narrowly, the market share will be large. It is the difficult task of the Court to determine the relevant market for a particular product.

For example, in the *du Pont cellophane case* of 1956 the government contended that du Pont, along with a licensee, had 100 percent of the cellophane market. But the Supreme Court defined the market broadly to include all "flexible packaging materials," that is, waxed paper, aluminum foil, and so forth, in addition to cellophane. Despite du Pont's total dominance of the "cellophane market," it only controlled about 20 percent of the market for "flexible packaging materials." The Court ruled that this did not constitute a monopoly.

Effectiveness

Have the antitrust laws been effective? This is a difficult question, but some insight can be gained by noting how the laws have been applied to existing market structures, mergers, and price fixing.

Existing Market Structures The application of antitrust laws to existing market structures has been lenient. Generally, a firm will be sued if it has more than 60 percent of the relevant market and there is evidence to suggest the firm used abusive conduct to achieve or

maintain its market dominance. The most significant recent "victory" against existing market structure has been the 1982 out-of-court settlement between the government and AT&T. AT&T was charged in 1974 with violating the Sherman Act by engaging in a series of anticompetitive actions designed to maintain its domestic telephone communications monopoly. As part of the settlement, AT&T agreed to divest itself of its 22 regional telephone operating companies. Since 1982, however, the Federal government has filed no significant antitrust suits to break up existing market structure.

Mergers The treatment of mergers varies with the type of merger and its effect on industry concentration.

Merger Types Mergers are of three basic types, as shown in Figure 33-1. This diagram shows two stages of production—the input stage, and the output, or final product stage—for two distinct final-good industries: autos and beer. Each rectangle represents a particular firm.

A **horizontal merger** is a merger between two competitors selling similar products in the same market. In Figure 33-1 this type of merger is shown as a combination of glass producers T and U. Other hypothetical examples of horizontal mergers would be Ford Motor Company merging with General Motors or Anheuser Busch merging with Coors.

A **vertical merger**—the merging of firms at different stages of the production process in the same industry—is shown in Figure 33-1 as a merger between firm Z, a hops producer, and firm F, a brewery. Vertical mergers involve firms having buyer-seller relationships. Actual examples of mergers of this type are PepsiCo's mergers with Pizza Hut, Taco Bell, and Kentucky Fried Chicken. PepsiCo supplies soft drinks to each of these fast-food operations.

A **conglomerate merger** is the merger of a firm in one industry with a firm in another unrelated industry. In Figure 33-1 a merger between firm C, an auto manufacturer, and firm D, a brewery, fits this description. Actual examples: the merger between Philip Morris, a cigarette company, and Miller Brewing; the merger between International Telephone and Telegraph (ITT) and Sheraton Hotel Corporation.

Merger Guidelines: The Herfindahl Index The Federal government has established merger guidelines based on the Herfindahl index (Chapter 26)

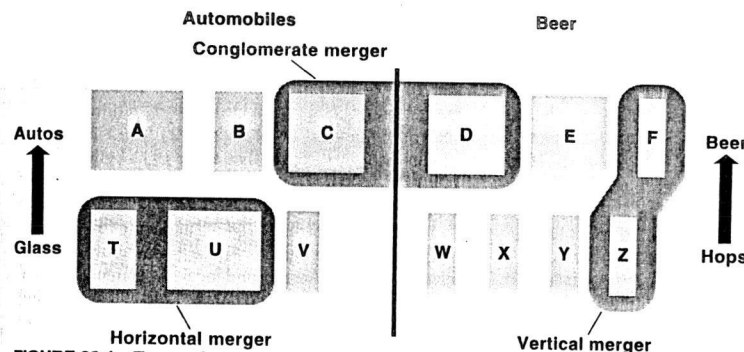


FIGURE 33-1 Types of mergers

Horizontal mergers (T + U) bring together firms selling the same products; vertical mergers (F + Z) connect firms having a buyer-seller relationship; and conglomerate mergers (C + D) join unrelated firms.

which measures the sum of the squared values of market shares within an industry. For example, an industry consisting of four firms, each with a 25 percent market share, has a Herfindahl index of 2500 ($=25^2 + 25^2 + 25^2 + 25^2$). In pure competition, where each firm's market share is minute, the index approaches 0 ($=0^2 + 0^2 + \dots + 0^2$). In pure monopoly, the index is 10,000 ($=100^2$).

Government uses Section 7 of the Clayton Act to block horizontal mergers which will substantially lessen competition. Generally, government is likely to challenge a horizontal merger if the post-merger Herfindahl index for the industry would be high (above 1800) and the merger has substantially increased the index (added 100 or more points). But other factors such as the impact of foreign competition and ease of entry of new firms are also considered. Also, mergers are usually allowed when one of the merging firms is on the verge of bankruptcy.

Most vertical mergers escape antitrust prosecution because they do not substantially lessen competition in either of the two markets. In Figure 33-1 neither the Herfindahl index in the hops industry nor that in the beer industry changes when the vertical merger between firms Z and F occurs. However, a vertical merger between large firms in highly concentrated industries may be challenged. In a 1949 case du Pont had acquired a controlling interest in General Motors'

stock. General Motors subsequently purchased about two-thirds of the paint and almost half the fabrics used in auto manufacturing from du Pont. The impact was to effectively foreclose other paint and fabric manufacturers from selling to GM. The Court ordered du Pont to divest itself of GM stock and sever the tie between the two firms.

Conglomerate mergers are generally permitted. If an auto manufacturer acquires a brewery, no antitrust action is likely to be taken because neither firm has increased the market share of its own market as a result of the merger. That is, the Herfindahl index would remain unchanged in each industry.

Price fixing Price fixing is treated strictly. Evidence of price fixing, even by small firms, will elicit antitrust action as will other collusive activities such as schemes to divide up sales in a market. In the parlance of antitrust law, these activities are known as *per se violations*; they are "in and of themselves" illegal, and therefore *not* subject to the rule of reason. To gain a conviction, the government or other party making the charge need only show that there was a conspiracy to fix prices or divide up sales, not that the conspiracy succeeded or caused serious damage to other parties.

Price-fixing investigations and court actions remain common. Recent examples are:

1 A 1991 court order stopping the Ivy League

schools from engaging in a 35-year practice of not competing with each other in the amount of financial aid offered to individual students

2 A 1991 investigation of several large airlines which allegedly practiced price collusion via their computerized fare network

3 A 1990 guilty plea by Southland Corporation and Borden for rigging bids for school milk contracts in Florida

4 An agreement by Panasonic in 1989 to refund \$16 million to buyers of consumer electronics, as part of a settlement for requiring wholesalers and retailers to charge minimum prices for its products

5 A 1991 investigation by the FTC into price fixing by producers of baby formula being sold to a Federal food program for low-income women and children

6 A 1990 antitrust settlement in Washington state by publishers of high-school yearbooks who rigged bids, fixed prices, and divided market territories

There are two major consequences of the government's vigor in prosecuting price fixing.

1 Price fixing is surrounded by great secrecy; it has been driven underground.

2 Collusive action is now much less formal; price leadership and the use of common cost-plus pricing formulas have often replaced formal price-fixing arrangements.

All the above statements about the application of antitrust laws are broad generalizations. Each potential antitrust case has unique circumstances which may make it an exception. Also, the strictness with which antitrust laws are interpreted has varied greatly among various administrations. The Reagan administration, for example, adopted a "lenient" enforcement posture toward existing market structures and mergers, while taking a "strict" position on price fixing. The Bush administration has continued the heavy emphasis on prosecuting price fixing. At present, there is little sentiment for breaking up existing domestic business monopolies in view of the increasing competition from equally large, or even larger, foreign firms.

Restricting Competition

Our discussion of antitrust laws and their application must not lead us to conclude that government policies are consistently procompetition. It is important to note that there are exemptions from antitrust and that a number of public policies have reduced competition.

Labor unions and agricultural cooperatives have been exempt, subject to limitations, from antitrust laws. We will see in Chapter 34 that Federal legislation and policy have provided some measure of monopolistic power for agriculture and have kept agricultural prices above competitive levels. Similarly, in a subsequent chapter we will discover that, since 1930, Federal legislation on balance has generally promoted growth of strong labor unions. This federally sponsored growth has resulted, according to some authorities, in development of union monopolies whose goal is above competitive wage rates. At state and local levels many occupational groups have been successful in establishing licensing requirements arbitrarily restricting entry to certain occupations, thereby keeping wages and earnings above competitive levels.

American patent laws—the first of which was passed in 1790—provide monetary incentive for innovators by granting them exclusive rights to produce and sell a new product or machine for a period of seven years. Patent grants protect the innovator from competitors who would otherwise quickly imitate this product and share in the profits, though not the cost and effort, of the research.

Few contest the desirability of this particular aspect of our patent laws, particularly when it is recalled that innovation can weaken and undermine existing positions of monopoly power. However, the granting of a patent frequently amounts to the granting of monopoly power in the production of the patented item. Many economists feel that the length of patent protection—seven years—is much too long.

The importance of patent laws in the growth of industrial concentration must not be underestimated. Du Pont, General Electric, American Telephone and Telegraph, Eastman Kodak, Alcoa, and innumerable other industrial giants have attained various measures of monopoly power in part through their ownership of certain patent rights.

Although detailed discussion of tariffs is postponed to a later chapter, at this point we must recognize that tariffs and similar trade barriers shield American producers from foreign competition. Protective tariffs are in effect discriminatory taxes against goods of foreign firms. These taxes make it difficult and often impossible for foreign producers to compete in domestic markets with American firms. The result is a less competitive domestic market and an environment frequently conducive to the growth of domestic industrial concentration.

QUICK REVIEW 33-2

• The Sherman Act of 1890 outlaws restraints of trade and monopolization; the Clayton Act of 1914 as amended by the Celler-Kefauver Act of 1950 outlaws price discrimination, tying contracts, mergers which lessen competition, and interlocking directorates.

• The Federal Trade Commission Act of 1914 and the Wheeler-Lea Act of 1938 give the Federal Trade Commission (FTC) authority to investigate unfair methods of competition and deceptive acts or practices of commerce.

• "Structuralists" contend that industries which are highly concentrated will behave like a monopolist; "behaviorists" hold that the relationship between industry structure and firm behavior is uncertain.

• Presently, government treats existing industrial concentration leniently; blocks most horizontal mergers between large, profitable firms in concentrated industries; and vigorously prosecutes price fixing by firms of all sizes.

• Government policies such as antitrust exemptions, patents, tariffs, and occupational licensure restrict competition.

NATURAL MONOPOLIES AND THEIR REGULATION

Antitrust is based on the assumption that society will benefit by preventing monopoly from evolving or, alternatively, by dissolving monopoly where it already exists. We now consider a special case in which there is an economic rationale for an industry to be organized monopolistically.

Natural Monopoly

A natural monopoly exists when economies of scale are so extensive that a single firm can supply the entire market at lower unit cost than could a number of competing firms. Such conditions exist for the so-called public utilities, such as electricity, water, gas, telephone service, and so on (Chapter 24). In these cases economies of scale in producing and distributing the product are very large and large-scale operations are necessary if low unit costs—and a low price—are to be realized (see Figure 22-9b). In this situation competition is simply uneconomic. If the market were divided among

many producers, economies of scale would not be realized, unit costs would be high, and high prices would be necessary to cover those costs.

There are two possible alternatives for promoting socially acceptable behavior on the part of a natural monopoly. One is public ownership and the other is public regulation.

Public ownership or some approximation thereof has been established in a few instances; the Postal Service, the Tennessee Valley Authority, and Amtrak at the national level, while mass transit, the water system, and garbage collection are public enterprises at the local level.

But public regulation has been the option pursued most extensively in the United States. Table 33-1 lists the major Federal regulatory commissions (industrial regulation) and their areas of jurisdiction. All the states also have such regulatory bodies for intrastate natural monopolies. The regulated sector is quantitatively important; an estimated 10 percent of the nation's output is produced by regulated industries.

The intent of "natural monopoly" legislation is embodied in the public interest theory of regulation. This theory envisions that such industries will be regulated for the benefit of the public, so that consumers may be assured quality service at reasonable rates. The rationale is this: If competition is inappropriate, regulated monopolies should be established to avoid possible abuses of uncontrolled monopoly power. Regulation should guarantee that consumers benefit from the economies of scale—that is, the lower per unit cost—which their natural-monopoly position allows public utilities to achieve. In practice, regulators seek to establish rates which will cover production costs and yield a "fair" or "reasonable" return to the enterprise.

TABLE 33-1 The main Federal regulatory commissions: industrial regulation

Commission (year established)	Jurisdiction
Interstate Commerce Commission (1887)	Railroads, trucking, buses, water, shipping, express companies, etc.
Federal Energy Regulatory Commission (1930)*	Electricity, gas, gas pipelines, oil pipelines, water power sites.
Federal Communications Commission (1934)	Telephones, television, cable television, radio, telegraph, CB radios, ham operators, etc.

*Originally called the Federal Power Commission; renamed in 1977.

More technically, the goal is to set price equal to average total cost. (You should review the "Regulated Monopoly" section of Chapter 24 at this point.)

Problems

There is considerable disagreement as to the effectiveness of regulation in practice. Let's briefly examine three major criticisms of regulation.

1 Costs and Efficiency Regulatory experience suggests that there are a number of interrelated problems associated with cost containment and efficiency in the use of resources.

A major goal of regulation is to establish prices so regulated firms will receive a "normal" or "fair" return above their production costs. But this means, in effect, that firms are operating on the basis of cost-plus pricing and, therefore, have no incentive to contain costs. Higher costs will mean larger total profits, so why develop or accept cost-cutting innovations if your "reward" will be a reduction in price? Stated technically, regulation fosters considerable X-inefficiency (see Chapter 24).

A regulated firm may resort to accounting skulduggery to overstate its costs and obtain a higher, unjustified profit. Furthermore, in many instances prices are set by the commission so that the firm will receive a stipulated rate of return based on the value of its real capital. This poses a special problem. To increase profits the regulated firm might make an uneconomic substitution of capital for labor, contributing to an inefficient allocation of resources within the firm (X-inefficiency).

2 Commission Deficiencies Another criticism is that the regulatory commissions function inadequately because they are frequently "captured" or controlled by the industries they are supposed to regulate. Commission members often were executives in these very industries. Therefore, regulation is *not* in the public interest, but rather, it protects and nurtures the comfortable position of the natural monopolist. It is alleged that regulation typically becomes a means of guaranteeing profits and protecting the regulated industry from new competition which technological change might create.

3 Regulating Competitive Industries Perhaps the most profound criticism of industrial regulation is that

it has sometimes been applied to industries which are *not* natural monopolies and which, in the absence of regulation, would be quite competitive. Specifically, regulation has been used in industries such as trucking and airlines where economies of scale are not great and entry barriers are relatively weak. In such instances it is alleged that regulation itself, by limiting entry, creates the monopoly rather than the conditions portrayed in Figure 24-6. The result is higher prices and less output than would have been the case without regulation. Contrary to the public interest theory of regulation, the beneficiaries of regulation are the regulated firms and their employees. The losers are the public and potential competitors barred from entering the industry.

Example: Regulation of the railroads by the Interstate Commerce Commission (ICC) was justified in the late 1800s and the early decades of this century. But by the 1930s the nation had developed a network of highways and the trucking industry had seriously undermined the monopoly power of the railroads. At this time it would have been desirable to dismantle the ICC and let railroads and truckers, along with barges and airlines, compete with one another. Instead, the regulatory net of the ICC was cast wider in the 1930s to include interstate truckers.

Legal Cartel Theory

The regulation of potentially competitive industries has given rise to the **legal cartel theory of regulation**. In the place of socially minded officials *forcing* regulation on natural monopolies to protect consumers, this view sees practical politicians as supplying the "service" of regulation to firms which *want* to be regulated. Regulation is desired because it constitutes a kind of legal cartel which can be highly profitable to regulated firms. Specifically, the regulatory commission performs such functions as dividing up the market (for example, the Civil Aeronautics Board, prior to deregulation, assigned routes to specific airlines) and restricting potential competition by enlarging the cartel (for example, adding the trucking industry to the ICC's domain). While private cartels are unstable and subject to breakdown (Chapter 26), the special attraction of a government-sponsored cartel under the guise of regulation is that it endures. In short, the legal cartel theory of regulation suggests that regulation results from rent-seeking activities (Chapter 32).

Proponents of the legal cartel theory of regulation point out that the Interstate Commerce Act was sup-

ported by the railroads and that the trucking and airline industries both supported the extension of regulation to their industries on the grounds that unregulated competition was severe and destructive. Proponents also note occupational licensing (Chapter 28) as the labor market manifestation of their theory. Certain occupational groups—barbers, interior designers, or dietitians—demand licensure on the ground that it is necessary to protect the public from charlatans and quacks, but the real reason may be to limit occupational entry so that practitioners may receive monopoly incomes.

DEREGULATION: THE CASE OF THE AIRLINES

The legal cartel theory, increasing evidence of inefficiency in regulated industries, and the contention that government was regulating potentially competitive industries all contributed to the deregulation movement of the 1970s and 1980s. Important legislation was passed which deregulated in varying degrees the airline, trucking, banking, railroad, natural gas, and television broadcasting industries.

Controversy

Deregulation has been controversial and the nature of the controversy is quite predictable. Basing their arguments on the legal cartel theory, proponents of deregulation contend it will result in lower prices and more output, and will eliminate bureaucratic inefficiencies. Some critics of deregulation, embracing the public interest theory, argue that deregulation will result in gradual monopolization of the industry by one or two firms, in turn leading to higher prices and diminished output or service. Other critics contend that deregulation may lead to excessive competition and industry instability and that vital services (for example, transportation) may be withdrawn from smaller communities. Still others stress that, as increased competition reduces each firm's revenues, firms may lower their safety standards as they try to reduce costs and remain profitable.

Perhaps the most publicized case of deregulation involves the airlines. The **Airline Deregulation Act (ADA)** was passed in 1978. Before the ADA the airlines were regulated by the Civil Aeronautics Board (CAB), which controlled airline fares and allocated interstate

routes among airlines. By assigning routes the CAB was also able to control industry entry. In fact, no new carriers were permitted to enter major interstate routes from the CAB's creation in 1938 until deregulation began in the late 1970s. This changed in the 1980s in that the 1978 deregulation act freed airlines to set their own rates and select their own routes. Under the terms of this act, the CAB was abolished in 1985.

Effects of Airline Deregulation

Determining the impact of airline deregulation has been complicated by such factors as fluctuating airplane fuel costs, recessions, the dismissal of striking air traffic controllers in 1981, and the significant expansion of national income between 1982 and 1990. Moreover, deregulation is less than two decades old and adjustments are still incomplete. Nevertheless, some of the effects of airline deregulation have become clear.

Fares Deregulation has exerted substantial downward pressure on fares, which overall have risen less than the general price level. Discount air tickets, in particular, have dramatically increased in availability and declined in price. Specifically, in 1976 only 15 percent of passengers used discount air fares; by 1987, that number had risen to 90 percent. The price of discount tickets for long intercoastal flights (such as from New York to Los Angeles) declined by 35 percent in real terms between 1976 and 1987.¹

Fare reductions have slowed since 1988, but in general fares remain about 20 percent lower in real terms than before deregulation. Of course, the benefits of fare reductions have not been spread uniformly among all airline markets. Passengers flying from some cities have enjoyed greater decreases than those flying from other cities.

There are two reasons why deregulation has produced lower air fares.

1 Competition among air carriers has driven ticket prices down to a closer proximity of the average cost of service than under the legal cartel form of regulation.

2 Competition has greatly pressured firms to reduce their costs. For example, the industry has adopted a "hub and spoke" route system analogous to a bicycle wheel. Passengers are flown from smaller cities along "spokes" into major "hub" airports where they change

¹*Economic Report of the President, 1988* (Washington, D.C.: 1988), p. 203.

planes and then fly to their more distant destinations. This system has reduced unit costs by allowing airlines to use smaller planes on the spoke routes and make use of wide-bodied craft between the major hub airports. Wide-body aircraft cost less to operate per seat-mile than smaller aircraft.

Also, the entry of "nonunion" airlines has forced the major carriers to negotiate wage reductions—often as much as 10 to 20 percent—with their unions. Some airlines have established a two-tier wage system whereby new workers are paid less on a given job than current employees. In many instances union work rules have been made more flexible to increase worker productivity and reduce wage costs. In short, some of the major cost reductions and adjustment problems associated with deregulation have occurred in relevant labor markets.

Service and Safety Although critics of airline deregulation predicted that airline service—particularly to smaller communities—would be curtailed or abandoned, this fear has turned out to be exaggerated. While some major airlines have withdrawn from a few smaller cities, commuter airlines have often filled the resulting void. Statistics indicate that the hub and spoke system has increased flight frequencies at most airports. It also has reduced the amount of airline switching required of passengers.

On the negative side, more frequent stopovers now required in hub cities have increased average travel time between cities. Also, by increasing the volume of air traffic, deregulation has contributed to greater airport congestion resulting in more frequent and longer flight delays.

Has deregulation reduced the "safety margin" of air transportation as some critics charge? There is mixed evidence on this question. On the one hand, the increased volume of air traffic has resulted in higher reported instances of near-collisions in midair. On the other hand, the accident and fatal accident rates of airlines are lower today than they were before deregulation. Furthermore, because lower air fares have caused people to substitute air travel for more-dangerous automobile travel, deregulation has prevented an estimated 800 deaths annually on the nation's highways.

Industry Structure Airline deregulation initially brought with it entry of numerous new carriers. In the past few years, however, the industry has gone through a "shakeout" in which many firms have failed

and others have merged with stronger competitors. In 1991 eight airlines accounted for 90 percent of domestic air traffic.

Growing concentration in the airline industry is of considerable concern. Some think consolidation of the industry may be detrimental to the very goals of deregulation itself. Single carriers control 75 percent or more of departures at six large "fortress" hubs. The General Accounting Office found that in 1990 fares at fifteen airports dominated by one or two airlines were about 20 percent higher than at twenty-two airports where competition was more brisk. Moreover, two factors make entry of new carriers into lucrative hubs difficult.

1 Lack of Airport Capacity The lack of airport capacity—at least in the short term—means that airline markets are far from being perfectly contestable. A firm wishing to enter a market because existing carriers are earning economic profits cannot do so if long-term leases allow existing carriers to control the airline gates at the profitable airports. It is alleged that some gates, in fact, go unused because dominant carriers refuse to release them to competitors.

2 Airline Practices Several tactics make it difficult for new firms to successfully enter the airline industry. Airline reservation systems developed by the major carriers give their own flights priority listings on the computers used by travel agents. Also, frequent-flyer programs—discounts based on accumulated flight mileage—encourage passengers to use dominant existing carriers rather than new entrants. Finally, price matching by existing carriers makes it exceedingly difficult for new entrants to lure customers through lower ticket prices.

Although it is too soon for a definitive assessment of airline deregulation, most economists view the outcome to date as positive. The Federal government has estimated that airline deregulation produced a \$100 billion net benefit to society during the 1980s. The tight oligopoly which seems to be emerging in the industry, however, is a mixed blessing. While it may lead to a handful of financially strong airlines, each having cost-minimizing route structures, it predictably will lead to price leadership and other business practices associated with oligopoly. At a minimum, strict enforcement against anticompetitive mergers and business tactics may be needed to preserve the gains from deregulation. Another procompetitive option would be to repeal

laws which bar foreign airlines from flying domestic routes in the United States.

QUICK REVIEW 33-3

- *Natural monopoly occurs where economies of scale are so extensive that only a single firm can produce the product at minimum cost.*
- *The public interest theory of regulation holds that government must regulate business to prevent allocative inefficiency arising from monopoly power.*
- *The legal cartel theory of regulation suggests that firms seek government regulation to reduce price competition and ensure stable profits.*
- *Although deregulation of the airline industry has reduced fares and produced net social benefits, it has also created growing industry concentration.*

SOCIAL REGULATION

The "old" regulation just discussed has been labeled economic or industrial regulation. Here government is concerned with the overall economic performance of a few specific industries, and concern focuses on pricing and service to the public.

Beginning largely in the early 1960s, government regulation of a new type evolved and experienced rapid growth. This social regulation is concerned with the conditions under which goods and services are pro-

duced, the impact of production on society, and the physical characteristics of goods themselves. For example, the Occupational Safety and Health Administration (OSHA) is concerned with protecting workers against occupational injuries and illnesses and the Consumer Products Safety Commission (CPSC) specifies minimum standards for potentially unsafe products.

The main Federal regulatory commissions dealing with social regulation are listed in Table 33-2.

Distinguishing Features

Social regulation differs from economic regulation in several ways.

1 Social regulation is often applied "across the board" to virtually all industries and directly affects far more people. While the Interstate Commerce Commission (ICC) focuses only on specific portions of the transport industry, OSHA's rules and regulations apply to every employer.

2 The nature of social regulation involves government in the very details of the production process. For example, rather than simply specifying safety standards for products, CPSC mandates—often in detail—certain characteristics which products must have.

3 A final distinguishing feature of social regulation is its rapid expansion. Between 1970 and 1979 government created twenty new Federal regulatory agencies.

The names of the better-known regulatory agencies in Table 33-2 suggest the basic reason for their creation and growth: Much of our society had achieved a reasonably affluent level of living by the 1960s and attention shifted to improvements in the quality of life. This improvement called for safer and better products, less pollution, better working conditions, and greater equality of opportunity.

Costs and Criticisms

It is generally agreed that the overall objectives of social regulation are laudable. But there is great controversy as to whether the benefits of these regulatory efforts justify the costs.

Costs The costs of social regulation are of two types: *administrative costs*, such as salaries paid to employees of the commissions, office expenses, and the like; and *compliance costs*, which are the costs incurred by businesses and state and local governments in meeting the requirements of regulatory commissions. In 1991 total

TABLE 33-2 The main Federal regulatory commissions: social regulation

<i>Commission (year established)</i>	<i>Jurisdiction</i>
Food and Drug Administration (1906)	Safety and effectiveness of food, drugs, and cosmetics
Equal Employment Opportunity Commission (1964)	Hiring, promotion, and discharge of workers
Occupational Safety and Health Administration (1971)	Industrial health and safety
Environmental Protection Agency (1972)	Air, water, and noise pollution
Consumer Product Safety Commission (1972)	Safety of consumer products

LAST WORD

DOES THE UNITED STATES NEED AN INDUSTRIAL POLICY?

Should government be more actively involved in determining the structure of industry?

There has been a growing concern in recent years that the United States' industrial preeminence has been seriously eroded. Our domestic markets have been flooded with foreign steel, automobiles, motorcycles, cameras, watches, and electronics equipment, suggesting that our competitive edge has been lost.

Noting apparent Japanese successes, many political, union, and business leaders—but only a limited number of economists—feel that the United States needs an industrial policy to reverse our alleged industrial decline. It is argued that government should undertake a more active and direct role in determining the structure and composition of American industry. Government, it is held, should use low-interest loans, loan guarantees, special tax treatment, research and development subsidies, antitrust immunity, and even foreign trade protection to accelerate the development of “high-tech” industries and to revitalize certain core manufacturing industries such as steel. Conversely, it should hasten the movement of resources out of declining “sunset” industries. Presumably the net result will be that the American economy will enjoy a higher average level of productivity and be more competitive in world markets.

Opponents of industrial policy make a number of points.

1 Deindustrialization? Has the United States in fact deindustrialized? Has our manufacturing sector



experienced serious decline? Statistics suggest not. While the composition of manufacturing output has changed, manufacturing in the aggregate accounts for virtually the same percentage of the nation's output in 1991 (about 20 percent) as it did in 1950. Similarly, manufacturing's share of the nation's expenditures on new plant and equipment was about the same in 1991

whether we are fully aware of the costs and benefits involved.¹²

¹²Testimony of Juanita M. Kreps in *The Cost of Government Regulation* (Washington, D.C.: 1978), p. 7.

CHAPTER SUMMARY

1 The case against industrial concentration centers on contentions that it a causes a misallocation of resources; b retards the rate of technological advance; c promotes income inequality; and d poses a threat to political democracy.

2 The defense of industrial concentration is built around

the following points: a firms have obtained their large market shares by offering superior products; b inter-industry and foreign competition, along with potential competition from new industry entrants, make American industries more competitive than generally believed; c some degree of monopoly may be essential to realize economies

as in 1950. Employment in manufacturing has declined from 34 to 17 percent of total employment in the 1950–1991 period, but that reflects the growth of labor productivity rather than industrial demise. Output and employment in manufacturing suffered in the early and mid-1980s, but the primary cause was a very strong dollar which made foreign imports relatively cheap and American exports relatively expensive.

2 Foreign experience Advocates of an industrial policy typically cite Japan as a model. In the post-World War II era Japan has achieved rapid economic growth; it has been highly successful in penetrating world markets; and it has had a much-publicized industrial policy. Yet the overall role of industrial policy as a causal factor in Japanese industrial success is not clear. The picture is mixed.

Japanese industrial policy has had both successes and failures. Some targeted industries, including semiconductor and machine tools, are almost certainly stronger than they would have been without government support and can be claimed as successes for Japanese industrial policy. Other industries, such as shipbuilding and steel, probably grow more quickly because of government aid, but undoubtedly would have developed without any government intervention. However, the Japanese government has also picked losers. Aluminum smelting and petrochemicals were favored industries fifteen years ago, but the public and private investments have paid off very poorly and now their capacity is being reduced. There are also several examples of successful industries that did not receive gov-

ernment assistance, including motorcycles and consumer electronics.*

3 Markets and politics While a proposal to create an industrial policy which subsidizes “sunrise” industries and hastens the phasing out of “sunset” industries sounds appealing, critics question the government's ability to identify future industrial “winners” and “losers.” The issue here is whether private investors using capital markets have better foresight than public officials in determining industrial winners and losers. Critics argue that private investors have a greater incentive in investing their own funds to obtain accurate information on the future prospects of various industries than might government bureaucrats in investing the taxpayers' funds.

Furthermore, might not government use its power to allocate investment funds to buy the political support of various industries? Might not the economic goal of enhanced industrial efficiency be subverted to the political goal of getting reelected? It is feared that the creation of a new industrial policy may lead to “lemon socialism,” that is, government support or ownership of declining industries and tying companies.

Those who are skeptical of industrial policy contend that government can best stimulate American industry by (1) using monetary and fiscal policy to create a favorable macroeconomic environment (high employment, low inflation, low interest rates) and (2) adjusting tax and regulatory systems to enhance incentives for investment and technological advance.

**Economic Report of the President, 1984*, p. 98.

Source: Adapted from *Economic Report of the President, 1984*, chap. 3; and Federal Reserve Bank of Kansas City, *Industrial Change and Public Policy* (1983). Updated.

of scale; and d monopolies and oligopolies are technologically progressive.

3 The cornerstone of antitrust policy consists of the Sherman Act of 1890 and the Clayton Act of 1914. The Sherman Act specifies that “Every contract, combination . . . or conspiracy in the restraint of interstate trade . . . is . . . illegal,” and that any person who monopolizes or attempts to monopolize interstate trade is guilty of a misdemeanor.

4 The Clayton Act was designed to bolster and make more explicit the provisions of the Sherman Act. To this end the Clayton Act declared that price discrimination, tying contracts, intercorporate stockholdings, and interlocking directorates are illegal when the effect of their use is the lessening of competition.

5 The Federal Trade Commission Act of 1914 created the Federal Trade Commission to investigate antitrust violations

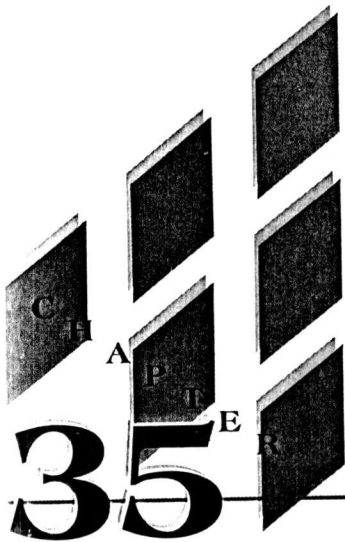
and to prevent the use of “unfair methods of competition.” Empowered to issue cease-and-desist orders, the Commission also serves as a watchdog agency for the false and deceptive representation of products.

6 The Celler-Kefauver Act of 1950 prohibits one firm from acquiring the assets of another firm where the result is a lessening of competition.

7 Major issues in applying antitrust laws include a the problem of determining whether an industry should be judged by its structure or its behavior and b defining the scope and size of the dominant firm's market.

8 Generally, antitrust officials are more likely to challenge price fixing and horizontal mergers among large firms than they are to attempt to break up existing market structures.

9 With respect to agriculture, labor, occupational licens-



Income Inequality and Poverty



It is not difficult to muster casual evidence which suggests substantial economic disparity in the United States.

Boxer Mike Tyson was estimated to have made over \$28 million in 1990; actor Jack Nicholson will reportedly earn \$50 million or more for his role in the *Batman* movie.

A recent study concludes that 5.5 million American children—one in eight—go hungry and that another 6 million are nutritionally “at risk.” In certain rural counties of the Deep South infant mortality rates exceed those of some less developed countries of Asia and Latin America.

The Census Bureau reports that almost 34 million Americans—over 13 percent of the population—live in poverty. Estimates indicate that 500,000 to 600,000 Americans are homeless.

The average salary in major league baseball is about \$900,000 per year. Boston Red Sox pitcher Roger Clemens earns almost \$5.4 million per year, which has been estimated to be \$1536 per pitch.

Government data indicate that income disparity is increasing in the United States; at present the richest fifth of the population receive almost 45 percent of total income while the poorest fifth receive less than five percent.

The question of how income should be distributed has a long and controversial history in both economics and philosophy. Should our national income and wealth be more or less equally distributed than is now the case? Or, in terms of Chapter 5, is society making the proper response to the “For whom” question?

We begin by surveying some basic facts concerning the distribution of income in the United States. Next, the major causes of income inequality are considered. Third, we examine the debate over income inequality and the tradeoff between equality and efficiency implied by this debate. Fourth, we will look at the poverty problem. Finally, we consider public policy; existing income-security programs are outlined and alternative approaches to welfare reform are discussed.

INCOME INEQUALITY: THE FACTS

How equally—or unequally—is income distributed in the United States? How wide is the gulf between rich and poor? Has the degree of income inequality increased or lessened over time?

Personal Income Distribution

Average income in the United States is among the highest in the world. The average income for all families was \$35,353 in 1990. But now we must examine how income is distributed around the average. In Table 35-1 we see that at the low end of the scale 9 percent of all families receive only 1 percent of total personal income. Only 3 percent of the total income went to the 17 percent of families receiving under \$15,000 per year in 1990. At the top of the income pyramid 6 percent of families received incomes of \$100,000 or more per year; this group received about 18 percent of total personal income. These figures suggest *there is considerable income inequality in the United States.*

Trends in Income Inequality

Over a period of years economic growth has raised incomes: *Absolutely*, the entire distribution of income has been moving upward over time. Has this changed the *relative* distribution of income? Incomes can move up absolutely, and the degree of inequality may or may

not be affected. Table 35-2 is instructive on the relative distribution of income. We divide the total number of income receivers into five numerically equal groups, or *quintiles*, and show the percentage of total personal (before-tax) income received by each in selected years. It is useful to examine the data in Table 35-2 over three periods: 1929–1947, 1947–1969, and 1969–1990.

1929–1947 Period Comparison of the income distribution data for 1929 and 1947 suggests that a significant reduction in income inequality occurred between these years. Note in Table 35-2 the declining percentage of personal income going to the top quintile and the increasing percentage received by the other four quintiles. Many of the forces at work during World War II undoubtedly contributed to this decline in inequality. Warborn prosperity eliminated the many low incomes caused by the severe unemployment of the 1930s, brought a reduction of wage and salary differentials, boosted depressed farm incomes through sharp increases in farm prices, temporarily diminished discrimination in employment, and was accompanied by a decline in property incomes as a share of the national income.

1947–1969 Period Many of the forces making for greater equality during World War II became less effective after the war. During the period between 1947 and 1969, the quintile distribution continued its previous trend toward less inequality, but at a far slower pace. The income share of the lowest income group rose by

TABLE 35-1 The distribution of personal income by families, 1990

(1) Personal income class	(2) Percentage of all families in this class	(3) Percentage of total personal income received by families in this class	(4) Percentage of all families in this class and all lower classes	(5) Percentage of income received by this class and all lower classes
Under \$10,000	9	1	9	1
\$10,000–\$14,999	8	2	17	3
\$15,000–\$24,999	16	8	33	11
\$25,000–\$34,999	16	11	49	22
\$35,000–\$49,999	20	19	69	41
\$50,000–\$74,999	18	27	87	68
\$75,000–\$99,999	7	14	94	82
\$100,000 and over	6	18	100	100
	100	100		

Source: Bureau of the Census, *Money Income of Households, Families, and Persons in the United States: 1990*, Current Population Reports, Series P-60, No. 174, 1991.

TABLE 35-2 Percentage of total before-tax income received by each one-fifth, and by the top 5 percent, of families, selected years

Quintile	1929	1935-1936	1947	1955	1969	1979	1990
Lowest 20 percent	12.5	4.1	5.0	4.8	5.6	5.3	4.6
Second 20 percent		9.2	11.8	12.2	12.4	11.6	10.8
Third 20 percent	13.8	14.1	17.0	17.7	17.7	17.5	16.6
Fourth 20 percent	19.3	20.9	23.1	23.7	23.7	24.1	23.8
Highest 20 percent	54.4	51.7	43.0	41.6	40.6	41.6	44.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Top 5 percent	30.0	26.5	17.2	16.8	15.6	15.7	17.9

.6 of a percentage point between 1947 and 1969, while that of the wealthiest quintile fell by 2.4 percentage points.

1969-1990 Period The distribution of income by quintiles has become more unequal since 1969. In 1990 the lowest 20 percent of families received only 4.6 percent of total before-tax income, compared to 5.6 percent in 1969. Meanwhile, the income share received by the highest 20 percent rose from 40.6 percent to 44.3 percent. The reasons for this latest trend in income inequality are in dispute. This chapter's Last Word considers a number of possible causes.

In summary, income inequality fell significantly between 1929 and 1947 but declined more slowly in the 1947-1969 period. More recently, income inequality has increased. But, as a direct comparison of the data for 1947 and 1990 in Table 35-2 reveals, the relative distribution in 1990 was similar to what it was over four decades ago.

The Lorenz Curve

The degree of income inequality can be seen through a **Lorenz curve** as shown in Figure 35-1. Here we *cumulate* the "percentage of families" on the horizontal axis and the "percentage of income" on the vertical axis. The theoretical possibility of a completely equal distribution of income is represented by the diagonal line brown because such a line indicates that any given percentage of families receives that same percentage of income. That is, if 20 percent of all families receive 20 percent of total income, 40 percent receive 40 percent, 60 percent receive 60 percent, and so on, all these points will fall on the diagonal line.

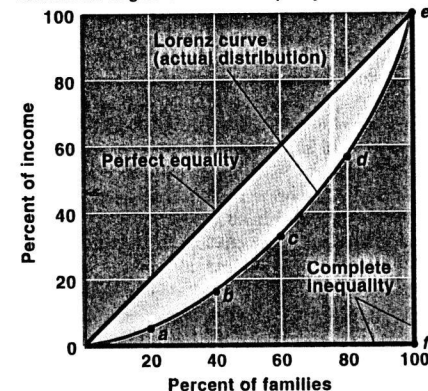
By plotting the 1990 data from Table 35-2 we locate the Lorenz curve to visualize the actual distribution of

income. Observe that the bottom 20 percent of all families received about 4.6 percent of the income as shown by point *a*; the bottom 40 percent received 15.4 percent ($=4.6 + 10.8$) as shown by point *b*; and so forth. The orange area, determined by the extent to which the resulting Lorenz curve sags away from the line of perfect equality, indicates the degree of income inequality. The larger this area or gap, the greater the degree of income inequality. If the actual income distribution were perfectly equal, the Lorenz curve and the diagonal would coincide and the gap would disappear.

At the opposite extreme is the situation of complete inequality where 1 percent of families have 100 percent of the income and the rest have none. In this

FIGURE 35-1 The Lorenz curve

The Lorenz curve is a convenient means of visualizing the degree of income inequality. Specifically, the orange area between the line of perfect equality and the Lorenz curve reflects the degree of income inequality.



case the Lorenz curve would coincide with the horizontal and right vertical axes of the graph, forming a right angle at point *f* as indicated by the heavy black lines. This extreme degree of inequality would be indicated by the entire area southeast of the diagonal.

The Lorenz curve can be used to contrast the distribution of income at different points in time, among different groups (for example, blacks and whites), before and after taxes and transfer payments are taken into account, or among different countries. As previously observed, the data in Table 35-2 tell us that the Lorenz curve shifted slightly toward the diagonal between 1947 and 1969 and then back away from the diagonal between 1969 and 1990. Comparisons with other nations suggest that the distribution of income in the United States is quite similar to those in most other industrially advanced countries.

ALTERNATIVE INTERPRETATIONS

There has been controversy in recent years as to whether the Bureau of Census data of Tables 35-1 and 35-2 provide an accurate portrayal of the degree of income inequality. Some scholars feel that the yearly census figures are inadequate. To understand the nature of these alleged deficiencies, we must first review these data. The census figures of Tables 35-1 and 35-2 show the distribution of *nominal* income and include not only wages, salaries, dividends, and interest, but also all *cash transfer payments* such as social security and unemployment compensation benefits. The data are *before taxes* and therefore do not account for the effects of personal income and payroll (social security) taxes which are levied directly on income receivers.

Two major criticisms of the census data are, first, that the income concept employed is too narrow, and second, that the income accounting period of one year is too short.

Broadened Income Concept

Edgar K. Browning¹ has made several adjustments in the Census Bureau data, resulting in a quite different picture of income distribution. Among other adjustments, Browning estimates the market value and dis-

¹Edgar K. Browning, "The Trend Toward Equality in the Distribution of Net Income," *Southern Economic Journal*, July 1976, pp. 912-923; and Browning, "How Much More Equality Can We Afford?" *The Public Interest*, Spring 1976, pp. 90-110.

tribution of *in-kind transfers*, that is, transfers of goods and services under such programs as Medicare, Medicaid, housing subsidies, and food stamps. Similarly, he takes into account the value and distribution of governmentally provided education. Next, he adds capital gains such as increases in the value of stocks, bonds, and real estate. Finally, he subtracts the amounts families pay as Federal personal income and payroll taxes. The picture which emerges from these adjustments not only is a much more equal distribution of income in each year, but it also indicates a trend toward greater equality over time. The movement toward greater equality is primarily a reflection of the rapid growth of in-kind transfers in the past twenty years or so.

It should be noted that Browning has been criticized for overadjusting the census data and thereby concluding that there is greater income equality than actually exists. Indeed, our point is that income distribution data are subject to many interpretations. In this regard, a recent study by the Census Bureau has confirmed that a broader definition of income translates into reduced inequality in the distribution of income. The Census Bureau has found, however, that its broader income concept tightens the "official" income distribution by only 4 percent.²

Lifetime Income

Another objection to the census data is that they portray the distribution of income in a single year and thereby conceal the possibility that the *lifetime earnings* of families might be more equal. If Ben earns \$1000 in year 1 and \$100,000 in year 2, while Holly earns \$100,000 in year 1 and only \$1000 in year 2, do we have income inequality? The answer depends on the period of measurement. Annual data would reveal great income inequality; but for the two-year period we have complete equality.

This is important because there is evidence to suggest that there is considerable "churning around" in the distribution of income over time. In fact, most income receivers follow an age-earnings profile where their income starts at relatively low levels, reaches a peak during middle age, and then declines. A glance back at Figure 28-9 reveals this general pattern. It follows that, even if people received the same stream of income over their lifetimes, considerable income inequality would still exist in any given year because of

²Bureau of the Census, *Measuring the Effect of Benefits and Taxes on Income and Poverty: 1990*, Current Population Report, Series P-60, No. 176, 1991, p. 11.

age differences. In any year the young and old would receive low incomes while the middle-aged received high incomes. This would occur despite complete equality of lifetime incomes.

Morton Paglin³ has adjusted the quintile data of Table 35-2 for age differences. He found that (1) there is greater income equality when the time factor is taken into account and (2) there was a trend toward greater income equality during the period studied: 1947–1972. The latter conclusion is attributed to the expansion of postsecondary education.

GOVERNMENT AND REDISTRIBUTION

One of the basic functions of government is to redistribute income. As Figure 35-2 and the accompanying table reveal, the distribution of household income *before* taxes and transfers are taken into account is substantially less equal than the distribution *after* taxes and transfers are included.⁴ *Government's tax system and transfer programs do reduce significantly the degree of inequality in the distribution of income.* Most of the reduction in income inequality—roughly 80 percent of it—is attributable to transfer payments. Recall from Chapter 32 that our tax system (Federal, state, and local taxes combined) is not highly progressive and, hence, the before-tax and after-tax distributions of income do not differ greatly. But transfers are vital in contributing to greater income equality. More specifically, government transfer payments account for over 75 percent of the income of the lowest quintile and have clearly been the most important means of alleviating poverty in the United States.

INCOME INEQUALITY: CAUSES

Why does the United States have the degree of income inequality evidenced in Tables 35-1 and 35-2? In general, we note that the market system is an impersonal mechanism. It has no conscience, and does not cater to

ethical standards concerning what is an "equitable," or "just," distribution of income. In fact, the basically individualistic environment of the capitalist economy is very permissive of a high degree of income inequality. Factors contributing to income inequality include:

1 Ability Differences People have different mental, physical, and esthetic talents. Some have inherited the exceptional mental qualities essential to entering the high-paying fields of medicine, dentistry, and law. Others, rated as "dull normals" or "mentally retarded," are assigned to the most menial and low-paying occupations or are incapable of earning income at all. Some are blessed with the physical capacity and coordination to become highly paid professional athletes. A few have the talent to become great artists or musicians. In brief, native talents enable some individuals to make contributions to total output which command very high incomes. Others are in much less fortunate circumstances.

2 Education and Training Individuals differ significantly in the amounts of education and training they have obtained and, hence, in their capacities to earn income. In part, these differences are a matter of voluntary choice. Smith chooses to enter the labor force upon high school graduation, while Jones decides to attend college. On the other hand, such differences may be involuntary: Smith's family may simply be unable to finance a college education.

3 Tastes and Risks Incomes differ because of differences in "job tastes." Those willing to take arduous, unpleasant jobs—for example, underground mining and garbage collecting—and to work long hours with great intensity will tend to earn more. Some people boost their incomes by "moonlighting," that is, by holding two jobs. Individuals also differ in their willingness to assume risk. We refer here not only to the stepladder and prize fighter but to the entrepreneur who assumes risk. Though most fail, the fortunate few who gamble successfully on the introduction of a new product or service may realize very substantial incomes.

4 Discrimination Simple supply and demand analysis suggests how discrimination—in this case labor market discrimination—generates income inequality. Suppose that gender discrimination restricts women to such occupations as secretaries, nurses, and teachers—once considered strictly "female" jobs. This means that the supplies of female workers will be great relative to demand in these few occupations so that

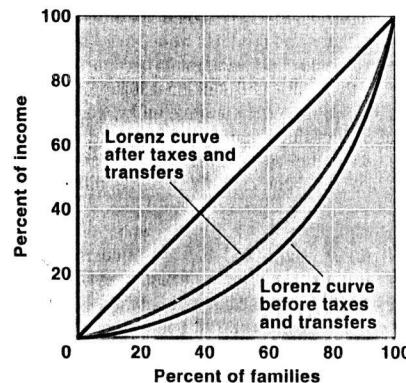


FIGURE 35-2 The impact of government taxes and transfers on income inequality

The distribution of personal income is significantly more equal after taxes and transfer payments are taken into account. Transfers account for most of the lessening of inequality and provide most of the income received by the lowest quintile of families.

Percent of income received, 1990

Quintile	Before taxes and transfers	After taxes and transfers
Lowest 20 percent	1.1	5.1
Second 20 percent	7.9	11.1
Third 20 percent	15.5	16.5
Fourth 20 percent	24.7	23.8
Highest 20 percent	50.7	43.5

Source: Bureau of the Census, *Measuring the Effect of Benefits and Taxes on Income and Poverty: 1990*, Current Population Report, Series P-60, No. 176-RD, 1991, p. 5. The data include all money income from private sources, including realized capital gains and employer-provided health insurance. The "after taxes and transfers" data include the value of noncash transfers as well as cash transfers.

wages and incomes will be low. Conversely, discrimination means males do not have to compete with women in "male" occupations (carpenters, pilots, accountants). This means supply is artificially limited relative to demand in these occupations, with the result that wages and incomes are high.

5 Property Ownership Ownership of property resources and receipt of property incomes are very unequal. The vast majority of households own little or no property resources, while the remaining few supply very great quantities of machinery, real estate, farmland, and so forth. A government study shows that in 1983 the top 10 percent of income receivers in the United States (those with annual incomes of \$50,000 or more) owned 72 percent of all stocks, 86 percent of all tax-free bonds, 70 percent of all taxable bonds, and 50 percent of all real estate. The top 2 percent of American income receivers (with annual incomes of \$100,000 or more) owned 50 percent of all stocks, 71 percent of all tax-free bonds, 39 percent of all taxable bonds, and 20 percent of all real estate.⁵ Similarly, an IRS study for

1986 indicates that nearly 28.5 percent of the nation's personal wealth is held by the richest 1.6 percent of its adults. Asset holdings are much more highly concentrated than are family incomes. Basically, property incomes account for the position of those households at the very pinnacle of the income pyramid. The right of inheritance and the fact that "wealth begets wealth" reinforce the role played by unequal ownership of property resources in determining income inequality.

6 Market Power Ability to "rig the market" on one's own behalf is undoubtedly a major factor in accounting for income inequality. Certain unions and professional groups have adopted policies limiting the supplies of their productive services, thereby boosting the incomes of those "on the inside." Legislation which provides for occupational licensing for barbers, beauticians, taxi drivers, and so forth, can also exert market power favoring the licensed group. The same holds true in the product market; profit receivers in particular stand to benefit when their firm develops some degree of monopoly power.

7 Luck, Connections, and Misfortune There are other important forces which play a part in explaining income inequality. Luck, chance, and "being in the right place at the right time" have all caused individuals

³Morton Paglin, "The Measurement and Trend of Inequality: A Basic Revision," *American Economic Review*, September 1975, pp. 598–609.

⁴The "before" data in this table differ from the data of Table 35-2 because the latter includes cash transfers. Also, the data in Table 35-2 are for families (a group of two persons or more related by birth, marriage, or adoption and residing together), whereas the data in Figure 35-2 are for all households (one or more persons occupying a housing unit). Finally, the data in Figure 35-2 are based on a broader concept of income than the data in Table 35-2.

⁵"Survey of Consumer Finances, 1983," *Federal Reserve Bulletin*, September 1984, pp. 679–692. Also see "Financial Characteristics of High-Income Families," *Federal Reserve Bulletin*, March 1986, pp. 163–177.

to stumble into fortunes. Discovering oil on a run-down farm or meeting the right press agent have accounted for some high incomes. Nor can personal contacts and political influence be discounted as means of attaining the higher income brackets. On the other hand, economic misfortunes such as prolonged illness, serious accident, death of the family breadwinner, and unemployment may plunge a family into poverty. The burden of such misfortunes is borne very unevenly by the population and hence contributes to the degree of income inequality.

QUICK REVIEW 35-1

- Income inequality has increased in the last decade; currently the top fifth of all families receive almost 45 percent of before-tax income and the bottom fifth receive under 5 percent.
- The Lorenz curve portrays income inequality graphically.
- Broadening the income concept and recognition of "churning" within the income distribution over time both lessen perceived income inequality.
- Government taxes and transfer payments significantly reduce income inequality.
- Differences in ability, education, job tastes, property ownership, and market power—along with discrimination and luck—help explain income inequality.

EQUALITY VERSUS EFFICIENCY

The critical policy issue concerning income inequality is: What is the optimal amount? While there is no generally accepted answer to this question, much can be learned by exploring the cases for and against greater equality.

The Case for Equality: Maximizing Utility

The basic argument for an equal distribution of income is that income equality is necessary if consumer satisfaction or utility is to be maximized. The rationale for this argument is shown in Figure 35-3 where it is assumed that the money incomes of two individuals, Anderson and Brooks, are subject to diminishing marginal utility (Chapter 21). In any time period income receivers

spend the first dollars received on those products they value most, that is, on products whose marginal utility is high. As their most pressing wants become satisfied, consumers then will spend additional dollars of income on less important, lower marginal utility, goods. The identical diminishing "marginal utility from income" curves reflect the assumption that Anderson and Brooks have the same capacity to derive utility from income.

Now suppose there is \$10,000 worth of income (output) to be distributed between Anderson and Brooks. The best or optimal distribution would be an equal distribution which causes the marginal utility of the last dollar to be the same for both persons. We can prove this by demonstrating that, for an initially unequal distribution of income, the combined total utility of two individuals can be increased by moving toward equality.

For example, suppose that initially the \$10,000 of income is distributed unequally so that Anderson gets only \$2500 and Brooks receives \$7500. The marginal utility from the last dollar received by Anderson is high (0a) and the marginal utility from Brooks' last dollar of income is low (0b). Clearly, redistribution of a dollar's worth of income from Brooks to Anderson—that is, toward greater equality—would increase (by $0a - 0b$) the combined total utility of the two consumers. Anderson's utility gain (the dark gray area in Figure 35-3a) exceeds Brooks' loss (the light red area in Figure 35-3b). This will continue to be so until income is equally distributed with each person receiving \$5000. At this point the marginal utility of the last dollar is identical for Anderson and Brooks ($0a' = 0b'$) and further redistribution cannot increase total utility.

The Case for Inequality: Incentives and Efficiency

Although the logic of the argument for equality is sound, critics attack its fundamental assumption that there exists some fixed amount of income to be distributed. Critics of income equality argue that the way in which income is distributed is an important determinant of the amount of income produced and available for distribution.

Suppose in Figure 35-3 that Anderson earns \$2500 and Brooks earns \$7500. In moving toward equality, society (government) must tax away some of Brooks' income and transfer it to Anderson. This tax-transfer process will diminish the income rewards of high-income Brooks and raise the income rewards of low-

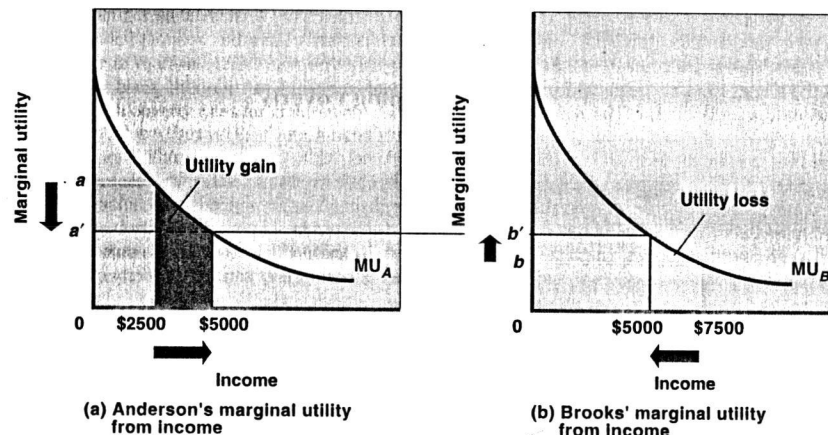


FIGURE 35-3 The utility-maximizing distribution of a given income

Proponents of income equality argue that, given identical "marginal utility from income" curves, Anderson and Brooks will maximize their combined utility when any given income (say, \$10,000) is equally distributed. If income is unequally distributed (\$2500 to Anderson and \$7500 to Brooks), the marginal utility derived from the last dollar will be greater for Anderson (0a) than for Brooks (0b) and, hence, a redistribution toward equality will result in a net increase in total utility. The utility gain shown by the dark gray area in panel (a) exceeds the utility loss indicated by the light red area in panel (b). When equality is achieved, the marginal utility derived from the last dollar of income will be equal for both consumers ($0a' = 0b'$) and, therefore, there is no further redistribution of income which will increase total utility.

income Anderson and in so doing reduce the incentives of both to earn high incomes. Why should Brooks work hard, save and invest, or undertake entrepreneurial risks, when the rewards from such activities will be reduced by taxation? And why should Anderson be motivated to increase his income through market activities when government stands ready to transfer income to him?

In the extreme, imagine a situation in which government levies a 100 percent tax on income and distributes the tax revenue equally to its citizenry. Why work hard? Indeed, why work at all? Why assume business risks? Why save—that is, forgo current consumption—to invest? The economic incentives to "get ahead" will have been removed and the productive efficiency of the economy—and the amount of income to be distributed—will diminish. The way the income pie is distributed affects the size of that pie! The basic argument for income inequality is that it is essential to maintain incentives to produce output and income.

The Equality-Efficiency Tradeoff⁶

The essence of the income (in)equality debate is that there exists a fundamental tradeoff between equality and efficiency.

The contrasts among American families in living standards and in material wealth reflect a system of rewards and penalties that is intended to encourage effort and channel it into socially productive activity. To the extent that the system succeeds, it generates an efficient economy. But that pursuit of efficiency necessarily creates inequalities. And hence society faces a tradeoff between equality and efficiency.⁷

Thus the problem for a society inclined toward egalitarianism is how to achieve a given redistribution

⁶This section is based on Arthur M. Okun, *Equality and Efficiency: The Basic Tradeoff* (Washington, D.C.: The Brookings Institution, 1975).

⁷Ibid., p. 1.

of income so as to minimize the adverse effects on economic efficiency. Consider this *leaky-bucket analogy*. Assume society agrees to shift income from the rich to the poor. But the money must be transferred from affluent to indigent in a leaky bucket. The leak represents an efficiency loss—the loss of output and income—due to the harmful effects of the tax-transfer process on incentives to work, to save and invest, and to accept entrepreneurial risk. It also reflects the fact that resources must be diverted to the bureaucracies which administer the tax-transfer system.

How much leakage will society accept and continue to endorse the redistribution? If cutting the income pie in more equal slices tends to shrink the pie, what amount of shrinkage will society tolerate? Is a loss of one cent on each redistributed dollar acceptable? Five cents? Twenty-five cents? Fifty cents? This is clearly a basic question which will permeate future political debates over extensions and contractions of our income-maintenance programs.

Fueling this debate over the equality-efficiency tradeoff are studies which suggest that the loss from the redistribution bucket may be quite high.

Edgar Browning and William Johnson . . . concluded that the upper-income groups bearing the costs of the taxes would sacrifice \$350 for every \$100 that the poor gained—a net efficiency loss of \$250. In Arthur Okun's terms, the leaks in the redistribution bucket are enormous—starting out with a bucket of \$350 raised from the nonpoor, \$250 is lost on the way to delivering it to the poor. For several reasons, critics of this study have found the estimate to be substantially too high. However, even if cut in half, this loss would be troublesome. Would our society be willing to accept a loss of economic efficiency of \$125—or even \$100—in order to equalize the distribution of income by transferring \$100 to the poor? The answer is by no means clear.⁸

THE DISMAL ECONOMICS OF POVERTY

Many people are less concerned with the larger question of income distribution than they are with the more specific issue of income inadequacy. Therefore, armed with some background information on income inequality, we now turn to the poverty problem. How extensive is poverty in the United States? What are the character-

istics of the poor? And what is the best strategy to take to lessen poverty?

Defining Poverty

Poverty does not lend itself to precise definition. But, in general, we might say that a family lives in poverty when its basic needs exceed its available means of satisfying them. A family's needs have many determinants: its size, its health, the ages of its members, and so forth. Its means include currently earned income, transfer payments, past savings, property owned, and so on.

The definitions of poverty developed by concerned government agencies are based on family size. In 1990 an unattached individual receiving less than \$6,652 per year was living in poverty. For a family of four the poverty line was \$13,359. For a family of six, it was \$17,839. Applying these definitions to income data for the United States, it is found that *about 13.5 percent of the nation—some 33.6 million people—lives in poverty.*

Who Are the Poor?

Unfortunately for purposes of public policy, the poor are heterogeneous; they can be found in all geographic regions, they are whites and nonwhites, they include large numbers of both rural and urban people, they are both old and young. Yet, as Table 35-3 clearly indicates, poverty is far from randomly distributed. While the total poverty rate—the percentage of the population living in poverty—was 13.5 percent for the entire population, blacks and Hispanics bore a disproportionate share compared to whites. On the other hand, thanks to a generous social security system, the incidence of poverty among the elderly is less than that for the population as a whole.

TABLE 35-3 The distribution of poverty, 1990

Population group	Percent in poverty
Total population	13.5
Whites	10.7
Blacks	31.9
Hispanics	28.1
Families headed by women	33.4
Children under 18	20.6
Elderly (65 or older)	10.9

Source: Bureau of the Census, *Money Income and Poverty Status in the United States: 1990*, Current Population Reports, series P-60, no. 175, 1991.

The incidence of poverty is extremely high among female-headed families and a full one-fifth of all children under 18 years of age live in poverty. The poverty rate among black children was 45 percent in 1990.

The high poverty rates for children are especially disturbing because in a very real sense poverty breeds poverty. Poor children are at greater risk for a range of long-term problems, including poor health and inadequate education, crime, drugs, and teenage pregnancy. Many of today's impoverished children will reach adulthood unhealthy, illiterate, and unemployable. The increased concentration of poverty among children bodes poorly for reducing poverty in the near future. It also implies problems for increasing the future productivity of the labor force because poor children receive less and generally inferior education.

Recalling our previous comments on movement or "churning" within the income distribution, we know that there is considerable movement in and out of poverty. Just over half of those who are in poverty one year will remain below the poverty line the next year. On the other hand, poverty is much more persistent for some groups, in particular black families and families headed by women.

Poverty Trends

Not revealed in Table 35-3 is the fact that the percentage of the population living in poverty was higher in 1990 than it was a decade or so ago. This disturbing reality is revealed in Figure 35-4, which traces out the percentage of people in poverty—or the poverty rate—for each year since 1960. Observe that the poverty rate fell significantly between 1960 and 1968, remained relatively unchanged from 1969–1978, and then increased sharply during the early 1980s. This recent increase in the poverty rate resulted from sluggish economic growth, high unemployment rates, and lower real levels of transfer payments. Beginning in 1984, the poverty rate gradually declined as the economy vigorously expanded toward full employment. As has been observed, the poverty rate in 1990 was 13.5 percent.

We need to add a qualification: Although the income levels used to compute the poverty rates shown in Figure 35-4 include cash transfer payments, they do *not* include the monetary value of such noncash transfers as medical care, housing assistance, and food stamps the poor receive. These noncash transfers are similar to income in that they enable the poor to purchase needed goods and services. Recently, the Census Bureau began estimating an alternative poverty

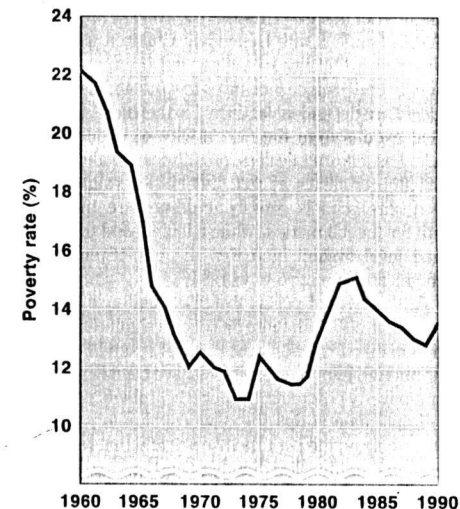


FIGURE 35-4 The U.S. poverty rate, 1960–1990

The percentage of the population living in poverty fell dramatically between 1960 and 1968, remained relatively constant for the next decade, and then climbed between 1978 and 1983. Beginning in 1984, the poverty rate once again began to decline but remained higher in 1990 than it was in the 1970s. Poverty rates for blacks have persistently exceeded those for whites.

rate which includes the value of noncash transfers. The poverty rate for 1990 was 9.8 percent using this expanded definition of income. But, irrespective of definitions of income, the basic point remains: poverty continues to be a persistent and difficult problem.

A "Black Underclass"?

Some observers contend that the city ghettos are spawning a "black underclass" which is trapped in a permanent cycle of poverty, broken homes, welfare, and, frequently, drugs and crime. Relevant statistics are alarming: 1 out of 2 black youths lives in poverty; 1 out of 2 black youths grows up without a father; nearly 40 percent of black teenagers are unemployed; 1 out of 4 births is to a teenager; more than 80 percent of children born to black teenagers are illegitimate; and 1 of every 21 young black men is a homicide victim.

It is argued that in the social and economic isolation of the urban ghetto a new culture—a culture of poverty and dependency—has evolved where attitudes, values, and morality are substantially different

⁸Robert H. Haveman, "New Policy for the New Poverty," *Challenge*, September-October 1988, p. 32.

from those of mainstream America. Welfare programs—Aid to Families with Dependent Children (AFDC), food stamps, housing subsidies, and the rest—have allegedly undermined incentives to work and have created welfare-dependent families. Furthermore, the historical exodus from the central city of middle-class blacks has left drug dealers, prostitutes, hustlers, and small-time criminals as role models for youngsters. Low-quality schools grossly underprepare minority youth for the job market, while a lenient and overburdened legal system increases the attractiveness of crime as an alternative to work.

At the level of policy, the black underclass view asserts that, although there has been a significant diminution in discrimination over the past several decades and although hundreds of billions have been expended on antipoverty programs, the poverty problem persists. The implication is that the responsibility for poverty rests largely on the poor themselves and that self-help is essential to the alleviation of poverty.

Critics of the black underclass view contend that it is a simplistic and callous position which incorrectly implies that the blame for poverty rests with its victims and not with larger social and economic considerations. Critics of the underclass view also contend that central-city poverty is heterogeneous and has a multitude of causes. What is needed is a far-reaching effort to eliminate racial segregation in housing and schooling, compensatory training and education, more accessible job opportunities, and an income maintenance program which does not discourage work.

The "Invisible" Poor

These facts and figures on the extent and character of poverty may be difficult to accept. After all, ours is an affluent society. How does one square the depressing statistics on poverty with everyday observations of abundance? The answer lies mainly in the fact that much American poverty is hidden; it is largely invisible.

There are three major reasons for this invisibility. First, a sizable proportion of the people in the poverty pool change from year to year. Research has shown that as many as one-half of those in poverty are poor for only one or two years before successfully climbing out of poverty.⁹ Hence, many of these people are not visible to us as being permanently downtrodden and needy. Second, the "permanently poor" are increasingly iso-

lated. Poverty persists in the slums and ghettos of large cities and is not readily visible from the freeway or commuter train. Similarly, rural poverty and the chronically depressed areas of Appalachia, the South, and the Southwest are also off the beaten path. Third, and perhaps most important,

The poor are politically invisible. . . . [They] do not, by far and large, belong to unions, to fraternal organizations, or to political parties. They are without lobbies of their own; they put forward no legislative program. As a group they are atomized. They have no face; they have no voice.¹⁰

Indeed, the American poor have been labeled "the world's least revolutionary proletariat."

THE INCOME MAINTENANCE SYSTEM

The existence of a wide variety of income-maintenance programs (Table 35-4) is evidence that alleviation of poverty has been accepted as a legitimate goal of public policy. Despite cutbacks in many programs in recent years, income-maintenance programs involve substantial monetary outlays and large numbers of beneficiaries. Total spending for income maintenance has expanded from about 4 percent of domestic production in 1940 to about 13 percent currently.

Our income-maintenance system consists of two kinds of programs: (1) social insurance programs and (2) public assistance or "welfare" programs.

Social Insurance Programs

Social insurance programs partially replace earnings lost due to retirement and temporary unemployment. "Social security" (technically Old Age, Survivors, and Disability Health Insurance or OASDHI), unemployment compensation, and Medicare are the main social insurance programs. Benefits are viewed as earned rights and do not carry the stigma of public charity. These programs are financed primarily out of Federal payroll taxes.

OASDHI and Medicare OASDHI is a gigantic social insurance program financed by compulsory payroll taxes levied upon both employers and employees. Ge-

¹⁰Michael Harrington, *The Other America: Poverty in the United States*, rev. ed. (New York: The Macmillan Company, 1970), p. 14.

TABLE 35-4 Characteristics of major income-maintenance programs

				Fiscal 1990	
Program	Basis of eligibility	Source of funds	Form of aid	Expenditures* (billions of dollars)	Beneficiaries (millions)
Social Insurance Programs					
Old Age, Survivors, and Disability Health Insurance (OASDHI)	Age, disability, or death of parent or spouse; individual earnings	Federal payroll taxes on employers and employees	Cash	\$248	40
Medicare	Age or disability	Federal payroll tax on employers and employees	Subsidized health insurance	98	34
Unemployment compensation	Unemployment	State and Federal payroll taxes on employers	Cash	7	14
Public Assistance Programs					
Supplemental Security Income (SSI)	Age or disability; income	Federal revenues	Cash	6	5
Aid to Families with Dependent Children (AFDC)	Certain families with children; income	Federal-state-local revenues	Cash and services	18	12
Food stamps	Income	Federal revenues	Vouchers	14	20
Medicaid	Persons eligible for AFDC or SSI and medically indigent	Federal-state-local revenues	Subsidized medical services	55	24

*Expenditures by Federal, state, and local governments; excludes administrative expenses.

Source: *Social Security Bulletin*, September 1991, and *Statistical Abstract of the United States*, 1991.

nerically known as "social security," the program replaces earnings lost because of a worker's retirement, disability, or death. A payroll tax of 7.65 percent is levied on both worker and employer and applies to the first \$55,500 of wage income. Workers may retire at 65 with full benefits or at 62 with reduced benefits. When the worker dies, benefits accrue to the survivors. Special provisions provide benefits for disabled workers. Currently, social insurance covers over 90 percent of all employed persons in the United States. In 1989 some 40 million people received OASDHI checks averaging about \$600 per month.

Medicare was appended to OASDHI in 1965. The hospital insurance it provides for the elderly and disabled is financed out of the payroll tax. Medicare also makes available a low-cost voluntary insurance program which helps pay doctor fees.

Unemployment Compensation All fifty states sponsor unemployment insurance programs. **Unemploy-**

ment compensation is financed by a modest payroll tax which varies by state and according to each firm's employment history. Any insured worker who becomes unemployed can, after a short waiting period (usually a week), become eligible for benefit payments. Almost 90 percent of all civilian workers are covered by the program. Size of payments and the number of weeks they may be received vary considerably from state to state. Generally speaking, benefits approximate one-half of a worker's after-tax wages up to a certain maximum payment. Benefits averaged \$152 weekly in 1990. The number of beneficiaries and the level of total disbursements vary greatly over the business cycle.

Public Assistance Programs

Public assistance, or *welfare*, programs provide benefits for those who are unable to earn income because of permanent handicaps or dependent children. These

⁹Greg J. Duncan, *Years of Poverty, Years of Plenty* (Ann Arbor, Mich.: University of Michigan Press, 1984).

programs are financed out of general tax revenues and are regarded as public charity. Individuals and families must demonstrate low incomes in order to qualify for aid. The Federal government finances about two-thirds of the welfare program expenditures.

Many needy persons who do not qualify for social insurance programs are assisted through other programs. Beginning in 1972 Federal grants to states for public assistance to the aged, the blind, and the disabled were terminated and a new Federally financed and administered **Supplemental Security Income (SSI) program** was created. The purpose of SSI is to establish a uniform, nationwide minimum income for these three categories of people who are unable to work. Over half the states provide additional income supplements to the aged, blind, and disabled.

The **Aid to Families with Dependent Children (AFDC) program** is state-administered, but partly financed with Federal grants. The program provides aid to families in which dependent children do not have the financial support of a parent, usually the father, because of death, disability, divorce, or desertion.

The **food stamp program** is designed to provide all low-income Americans with a "nutritionally adequate diet." Under the program eligible households receive monthly allotments of coupons which are redeemable for food. The amount of food stamps received varies inversely with a family's earned income.

Medicaid helps finance medical expenses of individuals participating in both the SSI and the AFDC programs.

QUICK REVIEW 35-2

• The fundamental argument for income equality is that it maximizes consumer utility; the basic argument for income inequality is that it is necessary to stimulate economic incentives.

• By government standards almost 34 million people or 13.5 percent of the population live in poverty.

• Our income maintenance system comprises both social insurance programs and public assistance ("welfare") programs.

"The Welfare Mess"

There is no doubt that the income maintenance system—not to mention local relief, housing subsidies, minimum-wage legislation, veterans' benefits, private

transfers through charities, pensions, and supplementary unemployment benefits—provides important means of alleviating poverty. On the other hand, the system has been subject to many criticisms in recent years.

1 Administrative Inefficiencies Critics charge that the wily-nilly growth of our welfare programs has created a clumsy and inefficient system, characterized by red tape and dependent on a huge bureaucracy for its administration. Administrative costs account for relatively large portions of the total budget of many programs.

The amount necessary to lift every man, woman, and child in America above the poverty line has been calculated, and it is *one-third* of what is in fact spent on poverty programs. Clearly, much of the transfer ends up in the pockets of highly paid administrators, consultants, and staff as well as higher income recipients of benefits from programs advertised as anti-poverty efforts.¹¹

2 Inequities Serious inequities arise in welfare programs in that people with similar needs may be treated very differently.

Benefit levels vary widely among States and among different demographic and family groups. Geographic differentials arise primarily because benefits under the two major public assistance programs—AFDC and Medicaid—are essentially controlled by the States. As a result, sharp disparities in benefit levels exist between the poorer, rural States and the wealthier, more urban areas. . . .¹²

A family in New York City might receive welfare benefits two times as great as the same family in Mississippi. Furthermore, control of the system is fragmented and some low-income families "fall between the cracks" while other families collect benefits to which they are not entitled.

3 Work Incentives A major criticism is that most of our income-maintenance programs impair incentives to work. This is because all welfare programs are constructed so that a dollar's worth of earned income yields less than a dollar of net income. As earned income increases, program benefits are reduced. An in-

dividual or family participating in several welfare programs may find that, when the loss of program benefits and the effect of payroll taxes on earnings are taken into account, the individual or family is absolutely worse off by working. In effect, the marginal tax rate on earned income exceeds 100 percent!

There are other criticisms. Noncash transfers interfere with freedom of consumer choice. Public assistance programs sap initiative and encourage dependency. AFDC regulations in some states promote family breakup by encouraging unemployed fathers to abandon their families so the spouse and children can qualify for benefits. AFDC benefits subsidize birth outside of marriage; nearly one-half of the mothers in the AFDC program have illegitimate children. Various welfare programs foster social divisiveness between workers and welfare recipients. For example, working mothers with small children may wonder out loud why poor mothers receiving AFDC should not also work for their money.

REFORM PROPOSALS

These criticisms have led to calls to reform the public assistance system. Although reform proposals have taken numerous forms, two broad approaches have dominated: negative income tax schemes and "workfare" plans.

Negative Income Tax

One contention is that the entire patchwork of existing welfare programs should be replaced by a **negative income tax (NIT)**. The term NIT suggests that, just as

the present (positive) income tax calls for families to "subsidize" the government through taxes when their incomes rise *above* a certain level, the government should subsidize households with NIT payments when household incomes fall *below* a certain level.

Comparing Plans Let's examine the two critical elements of any NIT plan. First, a NIT plan specifies a **guaranteed annual income** below which family incomes would not be allowed to fall. Second, the plan embodies a **benefit-loss rate** which indicates the rate at which subsidy benefits are reduced or "lost" as a consequence of earned income. Consider Plan One of the three plans shown in Table 35-5. In Plan One guaranteed annual income is assumed to be \$8000 and the benefit-loss rate is 50 percent. If the family earns no income, it will receive a NIT subsidy of \$8000. If it earns \$4000, it will lose \$2000 (\$4000 of earnings *times* the 50 percent benefit-loss rate) of subsidy benefits and total income will be \$10,000 (= \$4000 of earnings *plus* \$6000 of subsidy). If \$8000 is earned, the subsidy will fall to \$4000, and so on. Note that at \$16,000 the NIT subsidy becomes zero. The level of earned income at which the subsidy disappears and at which normal (positive) income taxes apply to further increases in earned income is called the **break-even income**.

One might criticize Plan One on the grounds that a 50 percent benefit-loss rate is too high and therefore does not provide sufficient incentives to work. Hence, in Plan Two the \$8000 guaranteed income is retained, but the benefit-loss rate is reduced to 25 percent. However, note that the break-even level of income increases to \$32,000 and many more families would now qualify for NIT subsidies. Furthermore, a family with any given earned income will now receive a larger NIT sub-

TABLE 35-5 The negative income tax: three plans (hypothetical data for a family of four)

Plan One (\$8000 guaranteed income and 50% benefit-loss rate)			Plan Two (\$8000 guaranteed income and 25% benefit-loss rate)			Plan Three (\$12,000 guaranteed income and 50% benefit-loss rate)		
Earned income	NIT subsidy	Total income	Earned income	NIT subsidy	Total income	Earned income	NIT subsidy	Total income
\$ 0	\$8,000	\$ 8,000	\$ 0	\$8,000	\$ 8,000	\$ 0	\$12,000	\$12,000
4,000	6,000	10,000	8,000	6,000	14,000	8,000	8,000	16,000
8,000	4,000	12,000	16,000	4,000	20,000	16,000	4,000	20,000
12,000	2,000	14,000	24,000	2,000	26,000	24,000*	0	24,000
16,000*	0	16,000	32,000*	0	32,000			

*Indicates break-even income. Determined by dividing the guaranteed income by the benefit-loss rate.

¹¹Thomas Sowell, *Markets and Minorities* (New York: Basic Books, Inc., Publishers, 1981), p. 122.

¹²*Economic Report of the President, 1978*, pp. 225–226.

sidy. For both of these reasons, a reduction of the benefit-loss rate to enhance work incentives will raise the cost of a NIT plan.

Examining Plans One and Two, still another critic might argue that the guaranteed annual income is too low in that it does not get families out of poverty. Plan Three raises the guaranteed annual income to \$12,000 and retains the 50 percent benefit-loss rate of Plan One. While Plan Three does a better job of raising the incomes of the poor, it too yields a higher break-even income than Plan One and would therefore be more costly. Furthermore, if the \$12,000 income guarantee of Plan Three were coupled with Plan Two's 25 percent benefit-loss rate to strengthen work incentives, the break-even income level would shoot up to \$48,000 and add even more to NIT costs.¹³

Goals and Conflicts By comparing these three plans we find that there are conflicts or tradeoffs among the goals of an "ideal" income-maintenance plan. First, a plan should be effective in getting families out of poverty. Second, it should provide adequate incentives to work. Third, the plan's costs should be reasonable. Table 35-5 tells us that these three objectives conflict with one another and that compromises or tradeoffs are necessary.

Plan One, with a low guaranteed income and a high benefit-loss rate, keeps costs down. But the low-income guarantee means it is not very effective in eliminating poverty and the high benefit-loss rate weakens work incentives. In comparison, Plan Two has a lower benefit-loss rate and therefore stronger work incentives. But it is more costly because it involves a higher break-even income and therefore pays benefits to more families.

Compared to Plan One, Plan Three entails a higher guaranteed income and is clearly more effective in eliminating poverty. While work incentives are the same as with Plan One, the higher guaranteed income makes the plan more costly. The problem is to find the magic numbers which will provide a "decent" guaranteed income, maintain "reasonable" incentives to work, and entail "acceptable" costs. While abolishing most of our current public assistance programs in favor of the NIT might be an improvement, the NIT is fraught with internal tradeoffs and should not be regarded as a panacea.

¹³You may have sensed the generalization that, given the guaranteed income, the break-even level of income varies *inversely* with the benefit-loss rate. Specifically, the break-even income can be found by dividing the guaranteed income by the benefit-loss rate. Hence, for Plan One, $\$8000/.50 = \$16,000$. Can you also demonstrate that, given the benefit-loss rate, the break-even level of income varies *directly* with the guaranteed income?

cea. In fact, reform efforts have moved away from the NIT in recent years.

Workfare Plans

However desirable the establishment of a NIT might be, political realities are such that piecemeal changes to the income maintenance system are more likely. In fact, most critical attention has focused on AFDC, for several reasons. First, as noted earlier, AFDC may encourage family dissolution. Second, it is contended that the program encourages—or at least subsidizes—illegitimate births. Third, some critics contend that AFDC is conducive to a "culture of poverty" where poverty becomes a way of life and is passed from generation to generation. Also, according to government studies, many recipients of AFDC receive benefits fraudulently. Fifth, as more middle-class mothers with children join the labor force, a consensus is emerging that poor mothers receiving AFDC should also work for their incomes.

These criticisms have led to a variety of **workfare proposals**—also called "welfare-to-work" plans—which would alter the AFDC program by providing work, training, and education activities to help, and eventually require, welfare recipients to move from public assistance to employment. People on welfare who undertake training or enter the labor force would also receive child care and transportation subsidies. As an additional aspect of this overall approach, earnings of absentee parents—whether married or unmarried—would be taken directly from workers' paychecks to pay child support.

Several states have had some success in their experiments with the "welfare-to-work" approach to poverty. The success of these state programs helped generate support for an overhaul of the AFDC program nationally. In late 1988 Congress passed and the President signed into law the **Family Support Act of 1988**, more commonly called the *Welfare Reform Act of 1988*. This important act embraces the workfare approach and includes the following provisions:

- 1 Each state must establish a Job Opportunities and Basic Skills program (JOBS) through which AFDC parents will be offered basic and remedial education, literacy classes, job skills training, job readiness activities, and job placement.
- 2 States must provide child care and Medicaid coverage for 12 months to welfare families switching from welfare rolls to employment. The purpose is to reduce the costs of moving from welfare to work and thus to lessen the incentives to stay on welfare.

LAST WORD

THE CAUSES OF GROWING INCOME INEQUALITY

Government data indicate that the gap between rich and poor in the United States has increased in the past twenty years.

A number of interrelated hypotheses have been put forth to explain growing income disparity.

1 Industrial Restructuring One possible explanation is that our industrial mix has changed from goods to services. The service sectors have both lower average wages and greater wage variation than the goods sectors. The result is greater income inequality.

2 Import Competition A related explanation is that more competition from imports in the 1970s and 1980s severely reduced the demand for and employment of less skilled but highly paid workers in such industries as automobiles and steel. The decline in such jobs reduced the average wage for less skilled workers. It also swelled the ranks of workers in already low-paying industries, placing further downward pressure on wages in such industries.

3 Returns to Education The college wage premium—the earnings advantage enjoyed by college graduates in comparison to high-school graduates—rose significantly in the 1980s. This implies a growing wage gap between more skilled and less skilled workers. The widening premium suggests that the demand for college graduates may have increased sharply relative to the demand for less skilled workers, reflecting the emergence of new high-tech industries such as computers and biomedicine. There has also been a dramatic increase in the compensation of chief executive officers.

4 Demographic Changes The entrance of large numbers of less experienced and therefore less skilled "baby boomers" into the labor force in the 1970s and 1980s may have contributed to greater income inequality. As large numbers of younger people entered the labor force, the median age of the average worker fell. Since younger workers typically earn less than



older workers, overall income inequality rose. In addition, the labor force participation of the wives of high-income husbands increased at a faster rate than for low-income husbands, adding to family income disparity. Finally, the number of unmarried or divorced women with children—who are very likely to be low income—has increased greatly.

5 Taxes and Transfers Legislation in the 1980s reduced Federal marginal tax rates such that high-income people are being taxed at rates below the levels of the 1960s and 1970s. Conversely, welfare benefits for the poor have either been cut or have not kept pace with inflation.

Conclusion: The widening of our income distribution is a complex phenomenon which is not readily explained in terms of one or two factors.

3 All states must begin offering welfare benefits to qualified two-parent families when the main wage earner is unemployed. The purpose of this provision is to reduce the incidence of family break-up associated with the AFDC program.

Supporters of the law believe that it can play an important role in helping end a "culture of welfare" in which dropping out of school, having a child, and going on welfare have allegedly become a normal way of life for part of the welfare population.

Labor-Market Issues: Unionism, Discrimination, and Immigration*



In this chapter we examine three important labor market issues: unionism, discrimination, and immigration. Although largely unrelated to each other, each issue is significant in its own right.

- 1 Much of the chapter consists of a detailed look at organized labor, collective bargaining, and the economic effects of unionism. What are the reasons for the historical growth and the recent decline of unionism? What impact do unions have on wages, efficiency and productivity, the distribution of earnings, and inflation?
- 2 We next discuss discrimination, its dimensions, and its costs.
- 3 Finally, we consider the much publicized issue of immigration of foreign labor to the United States. How many people enter the United States legally and illegally each year? What are the economic ramifications of this inflow of people?

BRIEF HISTORY OF AMERICAN UNIONISM

Some 17 million workers—16 percent of the employed labor force—now belong to labor unions. Bare statistics, however, may understate the importance of unions. The wage rates, hours, and working conditions of nonunionized firms and industries are influenced by those determined in organized industries. Unions are clearly important economic institutions of American capitalism.

As we consider how the labor movement evolved in the United States, we must examine government policy toward organized labor because labor legislation and union growth are intimately related. In terms of national labor policy, the American labor movement

has gone through three phases: repression (1790 to 1930), encouragement (1930 to 1947), and intervention (1947 to date). Though the dates are somewhat arbitrary, these three phases serve as an excellent guide for our discussion.

Repression Phase: 1790 to 1930

Labor unions have existed in the United States for 200 years. Shoemakers, carpenters, printers, and other skilled craftsmen formed unions of some permanence in the early 1790s. As Figure 36-1 indicates, despite this early start, union growth was relatively slow and spo-

*Instructors may choose to treat the three topics in this chapter selectively.

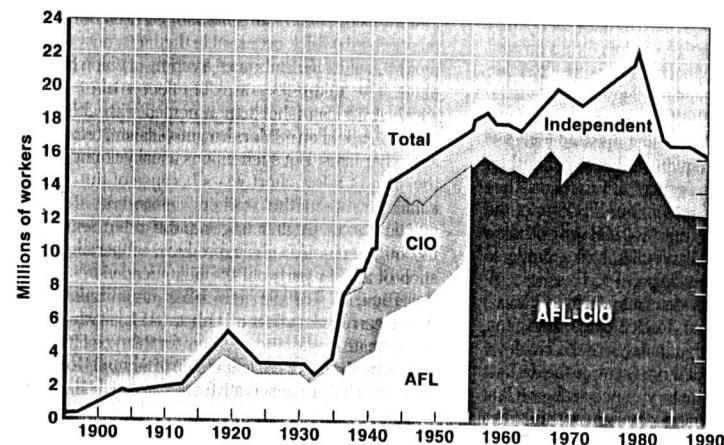


FIGURE 36-1 The growth and decline of union membership. Most of the absolute growth in organized labor has occurred since 1935. However, organized labor has been declining as a percentage of the labor force for some time and, in recent years, the absolute number of union members has also diminished. (U.S. Bureau of the Census and Bureau of Labor Statistics.)

radic until the 1930s. Two factors may account for this meager progress: (1) the hostility of the courts toward labor unions, and (2) the reluctance of American businesses to recognize and bargain with unions.

Unions and the Courts Not until the 1930s did legislation spell out the Federal government's policy toward labor unions. Lacking a national labor policy, it was up to the courts to decide on specific union-management conflicts. And, much to the dismay of organized labor, the courts were generally hostile toward unions. Their hostility had two sources. First, most judges had property-class backgrounds. Second, the courts are inherently conservative institutions charged with the responsibility of protecting *established* property rights. Unions, throughout the 1800s and the early decades of the 1900s, were in the unenviable position of seeking rights for labor at the expense of the *existing* rights of management.

The hostility of the courts was first given vent in the **criminal conspiracy doctrine**. This doctrine, "imported" by American courts from English common law at the turn of the nineteenth century, concluded that organizations of workers to raise wages were criminal conspiracies and hence illegal. Although unions as such were later recognized by the courts as legal organizations, the techniques employed by unions to press their demands—strikes, picketing, and boycotting—were generally held to be illegal. And, in the latter part

of the 1800s, the courts employed both antitrust laws and injunctions to impede the labor movement significantly.

Although Congress passed the Sherman Act of 1890 (Chapter 33) for the expressed purpose of thwarting the growth of business monopolies, the courts interpreted the loose wording of the act to include labor unions as conspiracies in restraint of trade and frequently so applied the act.

A simpler and equally effective antiunion device was the injunction. An injunction, or restraining order, is a court order directing that some act not be carried out, on the ground that irreparable damage will be done to those affected by the action. The attitude of the courts toward unions was such that it was easy for employers to obtain injunctions from the courts, prohibiting unions from enforcing their demands by striking, picketing, and boycotting. Stripped of these weapons, unions were relatively powerless to obtain the status and rights they sought.

Antiunion Techniques of Management The business community, hostile to unions from their inception, developed a group of techniques to undermine unions. A simple antiunion technique was ferreting out and firing prounion workers. Too, many employers felt it their duty to inform fellow employers that the discharged workers were "troublemakers" and "labor agitators" not fit to be hired. This combination of **discriminatory**

discharge and blacklisting made it extremely risky for workers to seek to organize a union. One's present and future employment opportunities were at stake.

Another potent weapon in management's struggle to keep unions down was the **lockout**, management's counterpart of the strike. By closing up shop for a few weeks, employers were frequently able to bring their employees to terms and destroy any notions they might have about organizing a union. Workers of the late 1800s and early 1900s were not blessed with savings accounts or multimillion-dollar strike funds to draw upon in such emergencies.

Where workers were determined to organize, pitched battles often ensued. Rocks, clubs, shotguns, and an occasional stick of dynamite were the shadowy ancestors of collective bargaining. Some of the darkest pages of American labor history concern the violent clashes between workers and company-hired *strikebreakers*. The Homestead strike of 1892, the Pullman strike of 1894, and the Ludlow Massacre of 1914 are cases in point. Less dramatic skirmishes erupt down to the present time.

But management tactics were often more subtle than a cracked skull. The **yellow-dog contract** was one of the more ingenious antiunion devices fostered by management. In such contracts workers agreed to remain nonunion as a condition of employment. They often had little choice but to sign such contracts—no contract, no job. Violation of a yellow-dog contract exposed a worker to a lawsuit by his employer, the result of which might be a court-imposed fine or even imprisonment.

As a last resort, an employer might shower his work force with such amenities as group insurance, pension programs, and stock ownership and profit-sharing schemes to convince them that employers would look after workers' interests as effectively as unions established by "outsiders." The next step beyond company *paternalism* was employee-representation schemes or **company unions**, that is, employer-dominated "dummy" unions which, it was hoped, would discourage the establishment of genuine unions. Paternalism and company unions were decidedly effective in retarding union growth as late as the 1920s.¹

¹During a prolonged strike in the bituminous coal industry in 1902, a spokesman for the mine operators, George F. Baer, issued the classic statement of business paternalism: "The rights and interests of the laboring man will be protected and cared for—not by the labor agitators, but by the Christian men to whom God in His infinite wisdom has given the control of the property interests of this country."

Evolution of Business Unionism The labor movement growth which occurred in the 1800s not only was modest, but it also embraced a variety of union philosophies. The mid-1800s were in effect a laboratory in which American labor experimented with alternative forms of unionism—Marxism, utopianism, reformism, and other isms. But such unions usually floundered in the span of a few short years because of the internal conflict between the workers' interest in short-run practical goals (higher wages and shorter hours) and the long-run utopian goals (producer cooperatives, creation of a labor party) of the union leaders.

Then, in 1886, a new labor organization—the **American Federation of Labor (AFL)**—which was to dominate the labor movement for the next fifty years was formed. Under the leadership of Samuel Gompers, labor charted a conservative course which has been very influential down to the present.² Appropriately honored as "the father of the American labor movement," Gompers preached three fundamental ideas: (1) practical business unionism, (2) political neutrality for labor, and (3) the autonomy of each trade or craft.

1 Business Unionism Gompers was firmly convinced that "safe and sane" **business unionism** was the only course for American labor to follow. Gompers rejected long-run idealistic schemes aimed at overthrow of the capitalistic system. He spurned intellectuals and theorists and emphasized that unions should be concerned with practical short-run economic objectives—higher pay, shorter hours, and improved working conditions. In the words of one scholar, Gompers felt that "you must offer the American working man bread and butter in the here and now instead of pie in the sky in the sweet by and by."³

2 Political Neutrality Gompers was convinced that government should keep its nose out of labor-management relations and collective bargaining. Although he recognized that governmental interference on behalf of labor might be a boon to union growth, Gompers was equally certain that antiunion government policies could stifle the progress of the entire labor movement.

²This is not to say that all unions have followed conservative paths since Gompers first espoused the virtues of business unionism. The Industrial Workers of the World, founded in 1905, advocated a decidedly revolutionary brand of left-wing unionism. In 1949 and 1950, the CIO expelled eleven affiliated unions whose leadership had come to be dominated by Communists.

³Charles C. Killingsworth, "Organized Labor in a Free Enterprise Economy," in Walter Adams (ed.), *The Structure of American Industry*, 3d ed. (New York: The Macmillan Company, 1961), p. 570.

Gompers cautioned organized labor not to align itself with any political party. Preoccupation with long-run political goals, he argued, causes labor to lose sight of the short-run economic objectives it should seek. Gompers admonished organized labor to reward labor's friends and punish its enemies at the polls regardless of political affiliation.

3 Trade Autonomy Finally, Gompers was firmly convinced that "autonomy of the trade," that is, unions organized on the basis of specific crafts, was the only permanent foundation for the labor movement. Unions composed of many different crafts lack the cohesiveness essential to strong, hard-hitting, business unionism. These craft unions should then be affiliated in a national federation. "One union to each trade, affiliated for one labor movement."

This philosophy—conservative business unionism, political "neutrality," and the craft principle of union organization—was destined to dominate the AFL and the entire labor movement for the next half-century. Indeed, the AFL, operating under Gompers' leadership, met with considerable success—at least for a time. AFL membership hit a high-water mark of about 4 million members by the end of World War I. Then a combination of circumstances in the 1920s forced the AFL into an eclipse (see Figure 36-1). One factor was a strong antiunion drive by employers. Also, many firms introduced employee representation plans, company unions, and a host of paternalistic schemes to convince workers that employers were better prepared to look out for their employees' interests than were labor leaders. Finally, the AFL clung tenaciously to the craft principle of union organization, ignoring the ever-increasing number of unskilled workers employed by the rapidly expanding mass-production industries—the automobile and steel industries in particular.

Encouragement Phase: 1930 to 1947

Two significant events occurred in the 1930s which revived the labor movement and inaugurated a period of rapid growth.

1 The attitude of the Federal government toward unions changed from one of indifference, not to say hostility, to one of encouragement.

2 A major structural change in the labor movement accompanied the founding of the Committee (later the Congress) of Industrial Organizations in 1936. Both events, coupled with the wartime prosperity of the 1940s, greatly swelled the ranks of organized labor.

Pro-labor Legislation of the 1930s Against the background of the depressed thirties, the Federal government enacted two decidedly pro-labor acts. In part, the passage of these acts reflected the strong opposition of organized labor to the previously described weapons employed by the courts and by management to suppress unions. In part, they reflected a Democratic administration replacing a Republican administration. In part, they echoed the widely held opinion that strong unions, by achieving higher wages through collective bargaining, would increase aggregate demand—or at least prevent it from falling—and help alleviate the Great Depression.

Norris-La Guardia Act of 1932 The Norris-La Guardia Act of 1932 did much to clear the path for union growth by outlawing two of the more effective antiunion weapons. Specifically, the act

- 1 Made it decidedly more difficult for employers to obtain injunctions against unions
- 2 Declared that yellow-dog contracts were unenforceable

Wagner Act of 1935 Three years later, in 1935, the Federal government took more positive steps to encourage union growth. The **Wagner Act of 1935** (officially the National Labor Relations Act) guaranteed the "twin rights" of labor: the right of self-organization and the right to bargain collectively with employers.

The act specified a number of "unfair labor practices" on the part of management. Specifically it

- 1 Forbade employers from interfering with the right of workers to form unions
- 2 Outlawed company unions
- 3 Prohibited antiunion discrimination by employers in hiring, firing, and promoting
- 4 Outlawed discrimination against any worker who files charges or gives testimony under the act
- 5 Obligated employers to bargain in good faith with a union duly established by their employees

The Wagner Act was clearly "labor's Magna Charta."

A **National Labor Relations Board (NLRB)** was established by the act and charged with the authority to investigate unfair labor practices occurring under the act, to issue cease-and-desist orders in the event of violations, and to conduct worker elections in deciding which specific union, if any, workers might want to represent them.

The Wagner Act was tailored to accelerate union growth and was extremely successful in achieving this goal. The protective umbrella provided to unions by

this act along with the Norris-La Guardia Act played a major role in increasing the ranks of organized labor from about 4 million in 1935 to 15 million in 1947.

Industrial Unionism: the CIO Recall that one of the causes of stagnation in the AFL during the 1920s was its unwillingness to organize the growing masses of unskilled assembly-line workers. Though the majority of AFL leaders chose to ignore unskilled workers, a vocal minority under the leadership of John L. Lewis contended that craft unionism would be ineffective as a means of organizing the hundreds of thousands of workers in the growing mass-production industries. According to Lewis and his followers, the basis for organization should be shifted from craft unionism to industrial unionism, that is, away from unions which only encompass a specific type of skilled workers (carpenters, bricklayers) to unions including all workers—both skilled and unskilled—in a given industry or group of related industries (steelworkers, autoworkers).⁴ This conflict came to a head, and in 1936 Lewis and his sympathizers withdrew their unions (and were simultaneously expelled) from the AFL.

The withdrawing unions established themselves as the Congress of Industrial Organizations (CIO). The CIO met with startling success in organizing the automobile and steel industries. So great was this success that the AFL also moved in the direction of organizing on an industrial basis. By 1940, total union membership approximated 9 million workers.

Intervention Phase: 1947 to Date

The prolabor legislation of the 1930s, the birth of industrial unionism, and the booming prosperity of the war years brought rapid union growth (see Figure 36-1). As unions gathered strength—both numerical and financial—it became increasingly evident that they could no longer be regarded as the weak sister or underdog in negotiations with management. Just as the growing power of business monopolies brought a clamor for public control in the 1870s and 1880s, the upsurge of union power in the 1930s and 1940s brought a similar outcry for regulation. This pressure for union control came to a head in the years immediately following World War II and culminated in the passage of the Taft-Hartley Act of 1947.

⁴Figures 28-6 and 28-7 compare the techniques employed by craft and industrial unions in attempting to raise wages.

Taft-Hartley Act of 1947 Officially called the Labor-Management Relations Act, the provisions of this detailed piece of legislation generally fall under four headings: (1) provisions which designate and outlaw certain “unfair union practices,” (2) provisions which regulate the internal administration of unions, (3) provisions which specify collective bargaining procedures and regulate the actual contents of bargaining agreements, and (4) provisions for handling of strikes imperiling the health and safety of the nation.

1 Unfair Union Practices The Wagner Act outlined a number of “unfair labor practices” on the part of management. A new and crucial feature of the Taft-Hartley Act was that it listed a number of “unfair labor practices” on the part of unions. These unfair practices, which constitute some of the most controversial sections of the act, are as follows: (a) Unions are prohibited from coercing employees to become union members. (b) **Jurisdictional strikes** (disputes between unions over the question of which has the authority to perform a specific job) are forbidden, as are **secondary boycotts** (refusing to buy or handle products produced by another union or group of workers) and certain **sympathy strikes** (strikes designed to assist some other union in gaining employer recognition or some other objective). (c) Unions are prohibited from charging excessive or discriminatory initiation fees or dues. (d) **Featherbedding**, a mild form of extortion where the union or its members receive payment for work not actually performed, is outlawed. (e) Unions cannot refuse to bargain in good faith with management.

2 Union Administration Taft-Hartley also imposed controls on the internal processes of labor unions: (a) Unions must make detailed financial reports to the National Labor Relations Board and make such information available to its members. (b) Welfare and pension funds must be kept separate from other union funds and jointly administered by the union and management. (c) Unions are prohibited from making political contributions in elections, primaries, or conventions which involve Federal offices. (d) Originally, union officials were required to sign non-Communist affidavits.

3 Contract Contents Other Taft-Hartley provisions are designed to control the actual collective bargaining process and the contents of the resulting work agreement: (a) The **closed shop** (which requires that a firm

hire only workers who are already union members) is specifically outlawed for workers engaged in interstate commerce; that is, a closed-shop arrangement cannot be written into a collective bargaining agreement. (b) Bargaining agreements must contain termination or *re-opening clauses* in which both labor and management must give the other party 60 days' notice of intent to modify or terminate the existing work agreement.

4 “Health and Safety” Strikes Finally, the Taft-Hartley Act outlines a procedure for avoiding major strikes which might disrupt the entire economy and imperil the health or safety of the nation, for example, a nationwide strike of port workers. According to this procedure, the President may obtain an injunction to delay such strikes for an 80-day “cooling off” period. Within this period striking workers are polled by the NLRB on the acceptability of the last offer of the employer. If the last offer is rejected, the union can then strike. The government's only recourse—one of questionable legality—is seizure of the industry.

Landrum-Griffin Act of 1959 Government regulation of the internal processes of labor unions was extended by passage of the Landrum-Griffin Act (officially the Labor-Management Reporting and Disclosure Act) in 1959. The act regulates union elections by requiring regularly scheduled elections of officers and the use of secret ballots; restrictions are placed on ex-convicts and Communists in holding union offices. Furthermore, union officials are now held strictly accountable for union funds and property. Officers handling union funds must be bonded; the embezzlement of union funds is made a Federal offense; and close restrictions are placed on a union's loans to its officers and members. The act is also aimed at preventing autocratic union leaders from infringing on the individual worker's rights to attend and participate in union meetings, to vote in union proceedings, and to nominate officers. The act permits a worker to sue his union if it denies him these rights.

UNIONISM'S DECLINE

In 1955 unity was formally reestablished in the American labor movement with the merger of the AFL and CIO. Two factors were especially important in closing the breach which had existed for almost two decades.

1 The political and legislative setbacks which labor

had encountered since the prolabor era of the 1930s convinced labor leaders that unity in the labor movement was a necessary first step toward bolstering the political influence of organized labor.

2 Failure to achieve the desired rate of growth in the ranks of organized labor in the post-World War II years made it evident to organized labor that a concerted, unified effort was needed to organize currently nonunion firms and industries.

In fact, however, the period since the AFL-CIO merger has *not* been characterized by a resurgence of organized labor. The growth of union membership has failed to keep pace with the growth of the labor force. While 25 percent of the labor force was organized in the mid-1950s, currently less than 15 percent are members. Indeed, in recent years the absolute number of union members has declined significantly. Over 22 million workers were unionized in 1980; that figure had fallen to only about 17 million in 1990.

Let's consider two possible explanations as to why this has happened.

1 Structural Changes One view, the **structural-change hypothesis**, is that many structural changes unfavorable to the expansion of union membership have occurred both in our economy and in the labor force.

1 Consumer demand and therefore employment patterns have shifted away from traditional union strongholds. Generally, the industry-mix of domestic output has been shifting away from manufactured goods (where unions have been strong) to services (where unions have been weak). This change in industry-mix may be reinforced by increased competition from imports in highly unionized sectors such as automobiles and steel. Growing import competition in these industries has curtailed domestic employment and therefore union membership.

2 An unusually large proportion of the increase in employment in recent years has been concentrated among women, youths, and part-time workers, groups allegedly difficult to organize because of their less firm attachment to the labor force.

3 Spurred by high energy costs, the long-run trend for industry to shift from the Northeast and Midwest where unionism is “a way of life” to “hard to organize” areas of the South and Southwest may have impeded expansion of union membership.

4 An ironic possibility is that the relative decline of unionism may in part reflect the success unions have

had in gaining a sizable wage advantage over nonunion workers in the United States and abroad. Confronted with high union wages, we would expect union employers to substitute machinery for workers, subcontract more work to nonunion suppliers, open nonunion plants in less industrialized areas, or have components produced in low-wage nations. These actions reduce the growth of employment opportunities in the union sector compared to the nonunion sector. Perhaps more important, we would also expect output and employment in low-cost nonunion firms and industries to increase at the expense of output and employment in higher-cost union firms and industries. In short, union success in raising wages may have changed the composition of industry to the disadvantage of union employment and membership.

2 Managerial-Opposition Hypothesis Another view is that intensified **managerial opposition** to unions has been a major deterrent to union growth. It is argued that in the past decade or so unions have increased the union wage advantage which they enjoy compared to nonunion workers and, as a result, union firms have become less profitable than nonunion firms. As a reaction, managerial opposition to unions has crystallized and become more aggressive. One managerial strategy has been to employ labor-management consultants who specialize in mounting aggressive anti-union drives to dissuade workers from unionizing or, alternatively, to persuade union workers to decertify their union.

It is also alleged that there has been a dramatic increase in the use of illegal antiunion tactics. In particular, it has become increasingly common to identify and dismiss leading prounion workers even though this is prohibited by the Wagner Act. Coupling these antiunion strategies with evidence that unions are devoting fewer resources to organizing the unorganized and that NLRB rulings have become increasingly anti-labor, the labor movement has gone into relative and absolute eclipse.

COLLECTIVE BARGAINING

Despite the decline of unionism, collective bargaining remains an important feature of labor-management relations. Nearly 2000 major collective bargaining agreements—those involving 1000 or more workers—cover 8.5 million workers in the United States. Many million other workers are covered under collective bargaining agreements in smaller firms.

The Bargaining Process

To the outsider, collective bargaining is a dramatic clash every two or three years between labor and management. It is easy to get the impression from the newspapers that labor and management settle their differences only with strikes, picketing, and occasional acts of violence.

These impressions are largely inaccurate. Collective bargaining is a somewhat less colorful process than most people believe. In negotiating important contracts, the union is represented by top local and national officials, duly supplemented with lawyers and research economists. Management representatives include top policy-making executives, plant managers, personnel and labor relations specialists, lawyers, and staff economists.

The union usually assumes the initiative, outlining its demands. These take the form of specific adjustments in the current work agreement. The merits and demerits of these demands are then debated. Typically, a compromise solution is reached and written into a new work agreement. Strikes, picketing, and violence are clearly the exception and not the rule. About 95 percent of all bargaining contracts are negotiated without resort to work stoppages. Generally, in recent years less than one-fifth of 1 percent of all working time has been lost each year from work stoppages resulting from labor-management disputes. *Labor and management display a marked capacity for compromise and agreement.* Strikes and labor-management violence are newsworthy, whereas peaceful renewal of a work agreement hardly rates a page-5 column.

The Work Agreement

Collective bargaining agreements assume many forms. Some agreements are brief, covering two or three typewritten pages; others are highly detailed, involving 200 or 300 pages of fine print. Some agreements involve only a local union and a single plant; others set wages, hours, and working conditions for entire industries. There is no such thing as an “average” or “typical” collective bargaining agreement.

At the risk of oversimplification, collective bargaining agreements usually cover four basic areas: (1) the degree of recognition and status accorded the union and the prerogatives of management, (2) wages and hours, (3) seniority and job opportunities, and (4) a procedure for settling grievances.

Union Status and Managerial Prerogatives Unions enjoy differing degrees of recognition from management. Listed in order of the union's preference are (1) the closed shop, (2) the union shop, and (3) the open shop.

Prior to being outlawed by the Taft-Hartley Act, the closed shop afforded the greatest security to a union. Under a closed shop a worker must be a member of the union before being hired. A **union shop**, on the other hand, permits the employer to hire nonunion workers but provides that these workers must join the union in a specified period—say, thirty days—or relinquish their jobs. Some twenty states now have so-called **right-to-work laws** which make compulsory union membership, and therefore the union shop, illegal.

Under the **open shop**, management may hire union or nonunion workers. Those who are nonunion are not obligated to join the union; they may continue on their jobs indefinitely as nonunion workers. Finally, there is the **nonunion shop**. Here no union exists, and the employer makes a conscious effort to hire those workers who are least inclined to form or join a union.

The other side of the union-status coin is the issue of **managerial prerogatives**. Most work agreements contain clauses outlining certain decisions which are to be made solely by management. These managerial prerogatives usually cover such matters as size and location of plants, products to be manufactured, types of equipment and materials used in production, and production scheduling. Frequently the hiring, transfer, discipline, discharge, and promotion of workers are decisions made solely by management but are subject to the general principle of seniority and to challenge by the union through the grievance procedure.

Wages and Hours The focal point of any bargaining agreement is wages and hours. Both labor and management tend to be highly pragmatic and opportunistic in wage bargaining. The criteria, or “talking points,” most frequently invoked by labor in demanding (and by management in resisting) wage boosts are (1) “what others are getting,” (2) ability to pay, (3) cost of living, and (4) productivity. If a given firm's basic rates are below those of comparable firms, the union is likely to stress that wages should be increased to bring them into line with what workers in other firms are getting. Similarly, if the firm has had a very profitable year, the union is likely to demand high wages on the ground that the company has ample ability to grant such incre-

ments. Unions have often achieved considerable success in tying wages to the cost of living. About 40 percent of all union workers are covered by some kind of *cost-of-living adjustment* (COLA). Finally, unions bargain for their “fair share” of the additional revenues associated with increases in productivity.

The four wage criteria are clearly two-edged propositions. For example, the cost-of-living criterion is invoked by the union only when prices are hurrying upward; unions conveniently ignore this criterion when prices are stable or declining. Similarly, the union considers the ability-to-pay argument to be important only when profits are large. Management is equally opportunistic in the evaluation it places on the various wage-bargaining standards.

Hours of work, overtime pay, holiday and vacations provisions, and **fringe benefits**—health plans and pension benefits—are other important “economic” issues which must be addressed in the bargaining process.

Seniority and the Control of Job Opportunities The uncertainty of employment in a market economy, coupled with the fear of antiunion discrimination on the part of employers, have made workers and their unions decidedly “job-conscious.” The explicit and detailed provisions covering job opportunities which most work agreements contain reflect this concern. Unions stress **seniority** as the basis for worker promotion and for layoff and recall. The worker with the longest continuous service has first chance at relevant promotions, is last to be laid off, and first to be recalled from a layoff.

Grievance Procedure Even the most detailed and comprehensive work agreement cannot anticipate all the issues and problems which might occur during its life. What if workers show up for work on a Monday morning to find that for some reason—say, a mechanical failure—the plant is closed down? Should they be given “show-up” pay amounting to, say, two or four hours' pay? Or management and the union may disagree as to whether the worker with the most seniority has the ability to perform the job to which he or she wants to be promoted. Such events and disagreements cannot be anticipated by even the most detailed collective bargaining contracts and therefore must be ironed out through a *grievance procedure*. Virtually all bargaining agreements contain an explicit grievance procedure to handle disputes which arise during the life of an agreement.

QUICK REVIEW 36-1

- *Union growth was slowed during the repression phase (1790–1930) by a use of the criminal conspiracy doctrine and injunctions by the courts, and b employer hostility.*
- *In the encouragement phase (1930–1947) union growth was stimulated by prounion legislation (the Norris–La Guardia and Wagner acts) and the evolution of industrial unionism.*
- *The Taft-Hartley and Landrum-Griffin acts inaugurated the intervention phase (1947 to the present) by regulating union tactics and their internal operations.*
- *The decline of unionism in recent decades has been attributed to a changes in the structures of the economy and the labor force, and b growing managerial opposition to unions.*
- *Collective bargaining agreements determine a union status and managerial prerogatives, b wages and hours, c control of job opportunities, and d the resolution of grievances.*

Given the historical and legislative background of the labor movement and some understanding of collective bargaining, let's now consider the economic implications of unions.

THE ECONOMIC EFFECTS OF UNIONS

Are the economic effects of labor unions positive or negative? We will respond to this important issue by examining several questions: Do unions raise wages? Do they increase or diminish economic efficiency? Do they make the distribution of earnings more or less equal? Do unions contribute to inflation? The reader should be forewarned that there is considerable uncertainty and debate on the answers to these questions.

The Union Wage Advantage

The three union models of Chapter 28 (see Figures 28-5, 28-6, and 28-7 and the accompanying discussions) all imply that unions have the capacity to raise wages. Has unionization in fact caused wage rates to be higher than otherwise?

Empirical research overwhelmingly does suggest that *unions do raise the wages of their members relative to comparable nonunion workers*, although the size of the union wage advantage varies according to occupation,

industry, race, and sex. There is also evidence to suggest that the union wage advantage increased in the 1970s. Hence, early research suggests that over the 1923–1958 period the average union–nonunion pay difference was on the order of 10 to 15 percent.

More recent studies indicate that the difference widened to 20 to 30 percent in the 1970s. Note that these are average differentials and that there is considerable variation among industries and occupations. Furthermore, the wage freezes and pay cuts (“wage give-backs”) suffered by organized labor in the early and mid-1980s most likely have significantly diminished the 20 to 30 percent union wage advantage. Labor economists have speculated that the union wage advantage may have returned to the 10 to 15 percent range by the early 1990s.

These estimates of the union wage advantage tend to be understated because union workers enjoy substantially larger *fringe benefits* than nonunion workers. Union workers are more likely to have private pensions, medical and dental insurance, and paid vacations and sick leaves than nonunion workers. Where such benefits are available to both union and nonunion workers, their magnitude is greater for union workers. Thus the total compensation (wage rates plus fringe benefits) advantage of union workers is greater than the previously indicated 10 to 15 percent.

Economists also generally agree that *unions have probably had little or no impact on the average level of real wages received by labor—both organized and unorganized—taken as a whole*. At first, these two conclusions—that unions gain a wage advantage but do not affect the average level of real wages—may seem inconsistent. But they need not be if the wage gains of organized workers are at the expense of unorganized workers. As we will see (Figure 36-2), higher wages in unionized labor markets may cause employers to move back up their labor demand curves and hire fewer workers. These unemployed workers may seek employment in nonunion labor markets. The resulting increase in the supply of labor will depress wage rates in these nonunion markets. The net result may well be no change in the average level of wages.

Indeed, the tight relationship between productivity and the average level of real wages shown in Figure 28-1 correctly suggests that unions have little power to raise real wage rates for labor as a whole. But Figure 28-1 is an average relationship and therefore compatible with certain groups of (union) workers getting higher relative wages if other (nonunion) workers are simultaneously getting lower real wages.

Efficiency and Productivity

Are unions a positive or negative force insofar as economic efficiency and productivity are concerned? How do unions affect the allocation of resources? While there is much disagreement as to the efficiency aspects of unionism, it is instructive to consider some of the ways unions might affect efficiency both negatively and positively. We will consider the negative view first.

Negative View There are essentially three basic means by which unions might exert a negative impact on efficiency.

1 Featherbedding and Work Rules Some unions have undoubtedly diminished productivity growth by engaging in “make-work” or “featherbedding” practices and resisting the introduction of output-increasing machinery and equipment. These productivity-reducing practices often arise against a backdrop of technological change. Labor and management may agree to a crew size which is reasonable and appropriate at the time the agreement is concluded. But labor-saving technology may then emerge which renders the crew too large. The union is likely to resist the potential loss of jobs. For many years the Brotherhood of Locomotive Firemen and Engineers retained a fireman on train crews, even though his function was eliminated by the shift from steam to diesel engines.

Similarly, union painters sometimes eschewed the use of spray guns and in some instances limited the width of paint brushes. In more recent years, typographer unions resisted the introduction of computers in setting type. Historically, the musicians' union insisted on oversized orchestras for musical shows and required that a union standby orchestra be paid by employers using nonunion orchestras.

More generally, one can argue that unions are responsible for the establishment of work rules and practices which impede efficient production. For example, under seniority rules workers may be promoted in accordance with their employment tenure, rather than in terms of who can perform the available job with the greatest efficiency. Also, unions may impose jurisdictional restrictions on the kinds of jobs workers may perform. Sheet-metal workers or bricklayers may be prohibited from performing the simple carpentry work often associated with their jobs. Observance of such rules means, in this instance, that unneeded and underutilized carpenters must be available. Finally, it is often contended that unions constrain managerial preroga-

tives to establish work schedules, determine production targets, and to make freely the decisions contributing to productive efficiency.

2 Strikes A second way unions may adversely affect efficiency is through strikes. If union and management reach an impasse in their negotiations, a strike will result and the firm's production will cease for the strike's duration. The firm will forgo sales and profits and workers will sacrifice income.

Simple statistics on strike activity suggest that strikes are relatively rare and the associated aggregate economic losses are relatively minimal. In 1990, 687 major collective bargaining agreements—those covering 1000 or more workers—were negotiated. Strikes occurred in only 43 of these instances. Furthermore, many strikes last only a few days. As indicated earlier, the average amount of work-time lost each year because of strikes is only about one-fifth of 1 percent of total work-time. This loss is the equivalent of 4 hours per worker per year, which is less than 5 minutes per worker per week!

Note that economic costs associated with strikes may be greater or less than suggested by the amount of work-time lost. Costs may be greater if production of nonstruck firms is disrupted. An extended strike in the steel or rail transportation industries could have serious adverse repercussions for production and employment in many other industries and sectors of the economy.

On the other hand, costs may be less than implied by workdays lost by strikers as nonstruck firms increase their output to offset the loss of production by struck firms. While the output of General Motors will fall when its workers strike, car buyers may shift their demand to Ford and Chrysler which respond by increasing their employment and outputs. While GM and its employees are hurt by a strike, society as a whole may experience little or no decline in employment, real output, and income.

3 Labor Misallocation A third and more subtle avenue through which unions might adversely affect efficiency is the union wage advantage itself. In Figure 36-2 we have drawn (for simplicity's sake) identical labor demand curves for the unionized and nonunion sectors of the labor market for some particular kind of labor.⁵ If there were no union present initially, then the wage rate which would result from the competitive hire

⁵Technical note: Our discussion assumes pure competition in both product and resource markets.

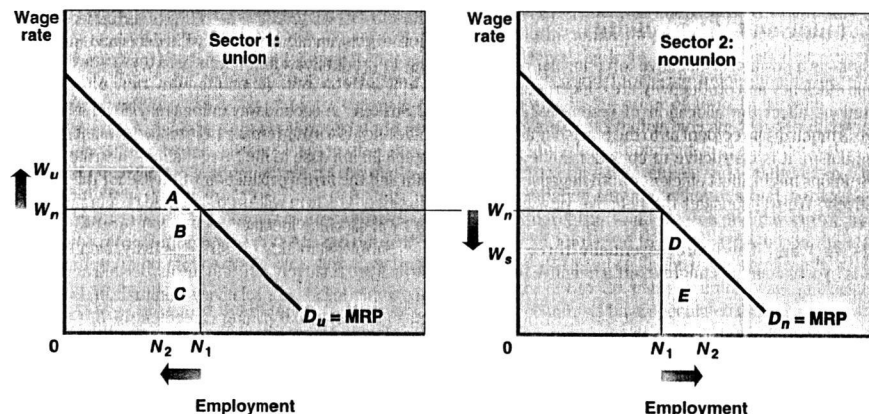


FIGURE 36-2 The effect of the union wage advantage on the allocation of labor

The higher wage W_u which the union achieves in sector 1 causes the displacement of N_1N_2 workers. The reemployment of these workers in nonunion sector 2 reduces the wage rate there from W_n to W_s . The associated loss of output in the union sector is area $A + B + C$, while the gain in the nonunion sector is only area $D + E$. Hence, the net loss of output is equal to area B . This suggests that the union wage advantage has resulted in the misallocation of labor and a decline in economic efficiency.

of labor would be, say, W_n . We now assume a union comes into being in sector 1 and succeeds in increasing the wage rate from W_n to W_u . As a consequence, N_1N_2 workers lose their jobs in the union sector. Assume they all move to nonunion sector 2 where they secure employment. This increase in labor supply in the nonunion sector depresses the wage rate from W_n to W_s .

Recall that the labor demand curves reflect the marginal revenue products (MRPs) of workers or, in other words, the contributions which workers make to the domestic output. This means that the shaded areas $A + B + C$ in the union sector represents the decrease in domestic output caused by the N_1N_2 employment decline in that sector. This $A + B + C$ area is the sum of the MRPs—the total contribution to domestic output—of the workers displaced by the W_n to W_u wage increase achieved by the union. The reemployment of these workers in nonunion sector 2 results in an increase in domestic output indicated by the shaded areas $D + E$. Because area $A + B + C$ exceeds area $D + E$, there is a net loss of domestic output. More precisely, because $A = D$ and $C = E$, the net loss attributable to the union wage advantage is equal to area B . Since the same amount of employed labor is now

producing a smaller output, labor is clearly being misallocated and inefficiently used.

Viewed from a slightly different perspective, after the shift of N_1N_2 workers from the union to the nonunion sector has occurred, workers will be paid a wage rate equal to their MRPs in both sectors. But the MRPs of the union workers will be higher than the MRPs of the nonunion workers. The economy will always benefit from a larger domestic output when any given type of labor is reallocated from a relatively low MRP use to a relatively high MRP use. But, given the union's presence and its ability to maintain the W_u wage rate in its sector, this reallocation from sector 2 to 1 will not occur.

Attempts to estimate the output loss due to the allocative inefficiency associated with union wage gains suggest that the loss is relatively small. One pioneering study assumed a 15 percent union wage advantage and estimated that approximately 0.14 percent—only about one-seventh of 1 percent—of the domestic output was lost! Similarly, a more recent estimate indicates that union wage gains cost the economy 0.2 to 0.4 percent of domestic product. In 1991 this cost would amount to about \$11 to \$23 billion or \$45.00 to \$90.00 per person.

Positive View Other economists take the position that on balance unions make a positive contribution to productivity and efficiency.

1 Managerial Performance: The Shock Effect The *shock effect* is the idea that a wage increase, imposed by a union in this instance, may induce affected firms to adopt improved production and personnel methods and become more efficient. One may carry Figure 36-2's analysis of labor misallocation one step further and argue that the union wage advantage will prompt union firms to *accelerate* the substitution of capital for labor (Chapter 27) and *hasten* the search for cost-reducing (productivity-increasing) technologies. When faced with higher production costs due to the union wage advantage, employers will be pushed to reduce costs by using more machinery and by seeking improved production techniques using less of both labor and capital per unit of output. In fact, if the product market is reasonably competitive, a unionized firm with labor costs 10 to 15 percent higher than nonunion competitors will simply not survive unless productivity can be raised. In short, union wage pressure may generate managerial actions which increase national productivity.

2 Reduced Worker Turnover Unions may also contribute to rising productivity within firms through their effects on worker turnover and worker security. Unions function as a **collective voice** for members in resolving disputes and improving working conditions. That is, if a group of workers is dissatisfied with its conditions of employment, it has two potential means of response. These are the "exit mechanism" and the "voice mechanism."

The **exit mechanism** simply refers to the use of the labor market—leave or exit your present job in search of a better one—as a means of reacting to "bad" employers and "bad" working conditions.

In contrast, the **voice mechanism** involves communication by workers with the employer to improve working conditions and resolve worker grievances. It might be risky for *individual* workers to express their dissatisfaction to employers because employers may retaliate by firing them as "troublemakers." But unions can provide workers with a *collective* voice to communicate problems and grievances to management and to press for their satisfactory resolution.

More specifically, unions may help reduce worker turnover in two ways.

1 Unions provide the voice mechanism as a substi-

tute for the exit mechanism. Unions are effective in correcting job dissatisfactions which would otherwise be "resolved" by workers through the exit mechanism of changing jobs.

2 The union wage advantage is a deterrent to job changes. Higher wages make unionized firms more attractive places to work. Several studies suggest that the decline in quit rates attributable to unionism is substantial, ranging from 31 to 65 percent.

A lower quit rate increases efficiency in several ways. First, lower turnover means a more experienced and, hence, more productive labor force. Second, fewer quits reduce the firm's recruitment, screening, and hiring costs. Finally, reduced turnover makes employers more willing to invest in the training (and therefore the productivity) of their workers. If a worker quits or "exits" at the end of, say, a year's training, the employer will get no return from the higher worker productivity attributable to that training. Lower turnover increases the likelihood that employers will receive a return on any training they provide, thereby making them more willing to upgrade their labor forces.

3 Seniority and Informal Training Much productivity-increasing training is transmitted informally. More-skilled workers may explain their functions to less-skilled workers on the job, during lunch, or during a coffee break. However, a more-skilled senior worker may want to conceal his or her knowledge from less-skilled junior workers if the latter can become competitive for the former's job. Because of union insistence on the primacy of seniority in such matters as promotion and layoff, worker security is enhanced. Given this security, senior workers will be more willing to pass on their job knowledge and skills to new or subordinate workers. This informal training enhances the quality and productivity of the firm's work force.

Mixed Research Findings A relatively large number of studies have measured the impact of unionization on productivity. These studies attempt to control for differences in labor quality, the amount of capital equipment used per worker, and other factors aside from unionization which might contribute to productivity differences. Unfortunately, evidence from these studies is inconclusive. For every study which finds a positive union effect on productivity, another study using different methodology or data concludes that there is a negative effect. Hence, at present there is no generally accepted conclusion regarding the overall impact of unions on labor productivity.

Distribution of Earnings

Labor unions envision themselves as institutions which enhance economic equality. Do unions in fact reduce the inequality with which earnings are distributed? The most convincing evidence suggests that unions do reduce earnings inequality.

Increasing Inequality Some economists employ Figure 36-2's analysis of labor misallocation to conclude that unions increase earnings inequality. They contend that, in the absence of the union, competition would bring wages into equality at W_u in these two sectors or submarkets. But the higher union wage realized in sector 1 displaces workers who seek reemployment in the nonunion sector. In so doing they depress nonunion wages. Instead of wage equality at W_u , we have higher wage rates of W_u for union workers and lower wages of W_u for nonunion workers. The impact of the union is clearly to increase earnings inequality. Furthermore, the fact that unionization is more extensive among the more highly skilled, higher-paid blue-collar workers than among less skilled, lower-paid blue-collar workers also suggests that the obtaining of a wage advantage by unions increases dispersion of earnings.

Promoting Equality There are other aspects of union wage policies which suggest that unionism promotes greater, not less, equality in the distribution of earnings.

1 Uniform Wages within Firms In the absence of unions employers are apt to pay different wages to individual workers on the same job. These wage differences are based on perceived differences in job performance, length of job tenure, and, perhaps, favoritism. Unions traditionally seek uniform wage rates for all workers performing a particular job. In short, while nonunion firms tend to assign wage rates to *individual workers*, unions—in the interest of worker allegiance and solidarity—seek to assign wage rate to *jobs*. To the extent that unions are successful, wage and earnings differentials based on supervisory judgments of individual worker performance are eliminated. An important side effect of this standard-wage policy is that wage discrimination against blacks, other minorities, and women is likely to be less when a union is present.

2 Uniform Wages among Firms In addition to seeking standard wage rates for given occupational classes

within firms, unions also seek standard wage rates among firms. The rationale is that the existence of substantial wage differences among competing firms may undermine the ability of unions to sustain and enhance wage advantages. For example, if one firm in a four-firm oligopoly is allowed to pay significantly lower wages to its union workers, the union is likely to find it difficult to maintain the union wage advantage in the other three firms. In particular, during a recession high-wage firms are likely to put great pressure on the union to lower wages to the level of the low-wage firm. To avoid this kind of problem unions seek to "take wages out of competition" by standardizing wage rates among firms, thereby reducing the degree of wage dispersion.

What is the *net* effect of unionism on the distribution of earnings? Although the issue remains controversial, one authoritative study concludes that the wage effects indicated in Figure 36-2 *increase* earnings inequality by about 1 percent, but the standardization of wage rates within and among firms *decreases* inequality by about 4 percent. The net result is a 3 percent decline in earnings inequality due to unionism. Because only a small proportion of the labor force is unionized, this 3 percent reduction in inequality is substantial.

Unions and Inflation*

We now examine the complicated and controversial question of whether unions can increase the average level of money wages and generate cost-push or, more specifically, wage-push inflation.

Two Models We have explored two general models of inflation, namely, the demand-pull and the cost-push models (Figures 9-6b and c and Figure 9-8). The demand-pull model suggests that, given aggregate supply, an increase in aggregate demand will result in a higher price level. Whether due to, say, an increase in the money supply or an increase in investment, the cause or impetus for inflation arises on the demand side of product markets which then increases the derived demands for labor and pulls up nominal wages. The important point is that in the demand-pull theory of

*This section presupposes that the reader has taken a course in macroeconomic principles.

inflation wage increases are an *effect* or symptom of inflation, not a *cause*. Wage increases do *not* cause inflation but are rather the *result* of excess aggregate demand. Wage increases simply *transmit* inflation, but do not initiate it.

In comparison, cost-push models allow union wage determination to play a causal role in inflation. Specifically, we know from equation (1) in Chapter 17 that if nominal-wage increases exceed increases in labor productivity, then unit labor costs will rise. Given that labor costs comprise about three-fourths of total production costs, product prices will rise roughly in accord with the increase in unit labor costs. In terms of Figure 9-8 a decrease in aggregate supply from AS_1 to AS_2 results in a higher price level. Some economists contend that union-inspired nominal-wage increases in excess of productivity increases can be an important cause of the indicated leftward shift of the aggregate supply curve.

Tentative Conclusions Which view is correct? While there is no universally accepted conclusion, most experts downgrade union wage-setting as a causal force in inflation. We know from our experience in the early 1960s that union wage determination can be compatible with price level stability. And one can argue with considerable credibility that the major episodes of rapid inflation in the United States were started either by expansions of aggregate demand or major supply shocks which had little or nothing to do with wage increases. Hence, the "great inflation" of the 1970s was rooted in increases in government military spending in the late 1960s, on the one hand, and supply shocks associated with the OPEC oil cartel and crop shortages, on the other.

More important perhaps is the fact that the cost-push model indicates that the decrease in aggregate supply which accompanies a union-induced increase in unit labor costs causes declines in output and increases in unemployment which act to restrain union wage demands (Figure 9-8). This suggests that rising unit labor costs could not generate continuing inflation unless accommodating monetary and fiscal policies gave rise to increases in aggregate demand to offset the falling output and rising unemployment which wage inflation would create (Figure 17-7b). Hence, the most reasonable judgment is that unions do *not* appear to be an initiating cause of inflation. Stated differently, unions do *not* seem to cause initial bursts of inflation or major increases in the rate of existing inflation independently of other causes.

QUICK REVIEW 36-2

- 1 Union workers receive wage rates 10 to 15 percent higher than comparable nonunion workers.
- 2 Union work rules, strikes, and the misallocation of labor associated with the union wage advantage are ways unions may reduce efficiency.
- 3 Unions may enhance productivity through the shock effect, by reducing worker turnover, and by providing the worker security prerequisite to informal on-the-job training.
- 4 On balance, unions probably reduce wage inequality by achieving wage uniformity within and among firms.
- 5 Most economists do not regard unions as an independent cause of inflation.

With this survey of unionism and collective bargaining complete, we now consider two additional factors affecting American labor markets—the problem of discrimination, and the controversial immigration issue.

DISCRIMINATION

In Chapter 35 we noted that blacks, Hispanics, and women bear a disproportionately large burden of poverty. The low incomes received by these groups are a consequence of the operation of the labor market. Thus, it is important that we consider the labor market aspects of discrimination.

Economic discrimination occurs when female or minority workers, who have the same abilities, education, training, and experience as white male workers, are accorded inferior treatment with respect to hiring, occupational access, promotion, or wage rate. Discrimination also occurs when females or minorities are denied access to education and training. Table 36-1 provides casual evidence which suggests the presence of racial discrimination. Similar data imply discrimination on the basis of gender. For example, the weekly earnings of full-time female workers is only about 70 percent that of males.

Dimensions of Discrimination

As Table 36-1 and our definition both suggest, discrimination may take several forms. Our discussion is in terms of racial and gender discrimination, but these

TABLE 36-1 Selected measures of discrimination and inequality of opportunity, 1990

Selected measure	Whites	Blacks
Income		
Median income of families	\$36,915	\$21,423
Percent of households in poverty	10.7	31.9
Percent of families with incomes of \$75,000 or more	13.7	4.8
Unemployment rate (percent of civilian labor force)		
All males	4.8	11.8
All females	4.6	10.8
Teenage† males	14.2	32.1
Teenage† females	12.6	30.0
Education		
Percent of population 25 years and over completing 4 years of high school or more	78.4	64.6
Percent of population 25 years and over completing 4 years of college or more	21.8	11.8
Occupational distribution (percent of total civilian employment)		
Managerial and professional occupations	27.1	16.0
Service occupations	12.0	22.9

†Males and females, 16–19 years old.

Sources: Statistical Abstract of the United States, 1991; Economic Report of the President, 1991; and Employment and Earnings, February 1991.

remarks also generally apply to discrimination based on age or ethnic background.

1 Wage discrimination occurs when black and other minority workers are paid less than whites for doing the same work. This kind of discrimination is of declining importance because of its explicitness and the fact that it clearly violates Federal law. But, as this chapter's Last Word demonstrates, wage discrimination can sometimes be very subtle and difficult to detect.

2 Employment discrimination means that unemployment is concentrated among minorities. Blacks are frequently the last hired and the first fired. Hence, the unemployment rate for blacks has been roughly double that for whites (Table 36-1).

3 Human-capital discrimination occurs when investments in education and training are lower for blacks than for whites. The smaller amount (Table 36-1) and inferior quality of the education received by blacks have cost them the opportunity to increase their productivity and qualify for better jobs. Unfortunately, a vicious circle seems to exist here. Many blacks are poor because they have acquired little human capital. Being poor, blacks have less financial ability to invest in education and training. They also have less economic motivation to invest in human capital. Facing the very

real possibility of wage, employment, and occupational discrimination, blacks tend to receive a lower rate of return on their investments in education and training.

4 Occupational discrimination means that minority workers have been arbitrarily restricted or prohibited from entering the more desirable, higher-paying occupations. Black executives and salespeople, not to mention electricians, bricklayers, and plumbers, are relatively few and far between. Historically, many craft unions effectively barred blacks from membership and hence from employment.

Occupational Segregation: The Crowding Model

This latter form of discrimination—**occupational segregation**—is particularly apparent in our economy. Women are disproportionately concentrated in a limited number of occupations such as nursing, public school teaching, secretarial and clerical jobs, and retail clerks. Blacks are crowded into a limited number of low-paying jobs such as laundry workers, cleaners and servants, hospital orderlies, and other manual jobs.

Assumptions The character and income consequences of occupational discrimination can be revealed

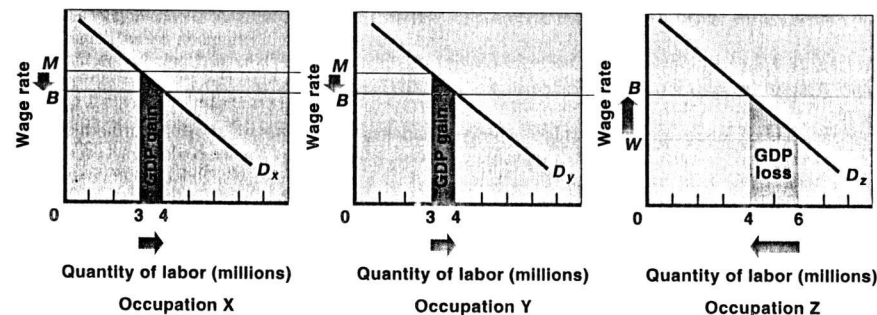


FIGURE 36-3 The simple economics of occupational discrimination

By crowding women into one occupation, men enjoy high wage rates of OM in occupations X and Y while women receive low wages of OW in occupation Z. The abandonment of discrimination will equalize wage rates at OB and result in a net increase in domestic output.

through a simple supply and demand model similar to that used to analyze the efficiency consequences of unions. We make the following simplifying assumptions.

- 1 The labor force is equally divided between male and female (or white and black) workers. Let's say there are 6 million male and 6 million female workers.
- 2 The economy is comprised of three occupations, each having identical labor demand curves, as shown in Figure 36-3.
- 3 Men and women (whites and blacks) have identical labor force characteristics; each of the three occupations could be filled equally well by men or women.

Effects of Crowding Suppose now that, as a consequence of irrational discrimination, the 6 million women are excluded from occupations X and Y and crowded into occupation Z. Men distribute themselves equally among occupations X and Y so there are 3 million male workers in each occupation and the resulting common wage rate for men is OM . (Assuming no barriers to mobility, any initially different distribution of males between X and Y would result in a wage differential which would prompt labor shifts from low- to high-wage occupation until wage equality was realized.) Note that women, on the other hand, are crowded into occupation Z and, because of this occupational segregation, receive a much lower wage rate OW . Given the reality of discrimination, this is an "equilibrium" situation. Women *cannot*, because of discrimination, reallocate themselves to occupations X and Y in the pursuit of higher wage rates.

Eliminating Discrimination But now assume that through legislation or sweeping changes in social attitudes, discrimination disappears. Women, attracted by higher wage rates, will shift from Z to X and Y. Specifically, 1 million women will shift into X and another 1 million into Y, leaving 4 million workers in Z. At this point 4 million workers will be in each occupation and wage rates will be equal to OB in all three occupations. Wage equality eliminates the incentive for further reallocations of labor.

This new, nondiscriminatory equilibrium is clearly to the advantage of women, who now receive higher wages, and to the disadvantage of men, who now receive lower wages. Women were initially harmed through discrimination to the benefit of men; the termination of discrimination corrects that situation.

There is also a net gain to society. Recall that the labor demand curve reflects labor's marginal revenue product (Chapter 27) or, in other words, labor's contribution to the domestic output.⁷ Hence, the gray areas for occupations X and Y show the *increases* in domestic output—the market value of the marginal or extra output—realized by adding 1 million women workers in each of those two occupations. Similarly, the orange area for occupation Z shows the *decline* in domestic output caused by the shifting of the 2 million women workers from occupation Z. We note that the sum of the two additions to domestic output exceeds the sub-

⁷Technical note: This assumes pure competition in product and resource markets.

traction from domestic output when discrimination is ended. Women workers are reallocating themselves from occupation Z, where their contribution to domestic output (their MRP) is relatively low, to alternative employments in X and Y, where their contributions to domestic output (their MRPs) are relatively high. Conclusion: *Society gains from a more efficient allocation of resources when discrimination is abandoned.* Discrimination influences the distribution of a diminished domestic output. That is, discrimination places the nation on a point inside of its production possibilities curve.

Costs of Discrimination

Given the diverse types of discrimination, the economic costs of discrimination are difficult to estimate. However, one estimate is that if economic and social policies were successful in lowering the black unemployment rate to the level of the white rate, and if education and training opportunities were made available to the black labor force so that the average productivity of black labor became equal to that of white workers, the total output of the economy would rise by about 4 percent. For example, in 1991 the economic cost of racial discrimination alone would be about \$227 billion.

Addenda

We must consider two important additions to our discussion of discrimination.

Comparable Worth Doctrine The first involves public policy. The reality of pervasive occupational segregation has given rise to the issue of comparable worth. Legislation such as the Equal Pay Act of 1963 which forced employers to pay equal wages to men and women performing the same jobs was of no help to many women because occupational segregation limited their access to jobs held by men. The essence of the **comparable worth doctrine** is that female secretaries, nurses, and clerks should receive the same salaries as male truck drivers or construction workers if the levels of skill, effort, and responsibility in these disparate jobs are comparable. The basic advantage of comparable worth is that it is a means of quickly correcting perceived pay inequities.

While the concept of comparable worth has considerable appeal, there are a number of important objections. For example, any comparison of the relative worth of various jobs is necessarily subjective and therefore arbitrary, opening the door to endless contro-

versies and lawsuits. Second, wage setting by administrative or bureaucratic judgment, rather than supply and demand, does not bode well for long-run efficiency. To the extent that the calculated worth of specific jobs varies from their market or equilibrium value, worker shortages or surpluses will develop. Furthermore, increasing the wages of women could attract even more females to traditionally "women jobs" and prolong occupational segregation.

Nondiscriminatory Factors Not all the average income differentials found between blacks and whites and males and females are necessarily due to discrimination. Most researchers agree, for example, that some part of the male-female earnings differential is attributable to factors other than discrimination. For example, the work-life cycle of married women who have children historically has involved a continuous period of work until birth of the first child. Then there is a five- to ten-year period of nonparticipation or partial participation in the labor force related to childbearing and child care, followed by a more continuous period of work experience when the mother is in her late thirties or early forties. The net result is that, on the average, married women have accumulated much less labor force experience than men in the same age group. Hence, on the average females are less productive workers and are therefore paid a lower average wage rate.

Furthermore, family ties apparently provide married women with less geographical mobility in job choice than males. In fact, married women may give up good positions to move with husbands who accept jobs elsewhere. And some married women may put convenience of job location and flexibility of working hours ahead of occupational choice. Again, women may have purposely crowded into such occupations as nursing and elementary school teaching because such occupations have the greatest carryover value for productive activity within the home. Finally, in the past decades more women have entered the labor force than have men. This large increase in the supply of female workers has acted as a drag on women's wages and earnings.

All this implies that some portion of the male-female earnings differential is due to considerations other than discrimination by gender. It also suggests that the male-female wage gap will narrow in the future, now that more women are attending college, working through their childbearing years, and pursuing higher-paying professional jobs.

QUICK REVIEW 36-3

1 *Discrimination may mean a paying different wages to equally qualified workers, b higher unemployment rates for minorities, c less education and training for women and minorities, and d the concentration of minorities and women in a limited number of occupations.*

2 *The crowding model demonstrates how a men can increase their wages at the expense of women, and b occupational segregation diminishes the domestic output.*

3 *Comparable worth means that females in one occupation should receive the same wages as males in another occupation if the levels of skill, effort, responsibility and working conditions are comparable.*

IMMIGRATION

The immigration issue has long been clouded in controversy and misunderstanding. Should more or fewer people be allowed to migrate to the United States? How should the much-publicized problem of illegal entrants be handled? We will illuminate this problem by (1) briefly summarizing United States' immigration history and policy, (2) presenting a bare-bones model of the economic effects of immigration, and (3) embellishing this simple model by considering some of the more subtle costs and benefits associated with the international movement of labor.

History and Policy

During the first 140 years of our history as an independent nation, immigration to the United States was virtually unimpeded. There is little question that the great infusion of foreign labor into our labor-scarce country was a major contributing factor to our nation's economic growth. But the great flood of immigrants which came to the United States in the quarter-century prior to World War I was sharply curtailed by the war itself and by a series of restrictive immigration laws enacted in the 1920s. However, after World War II, immigration policy was liberalized and the annual inflows of legal immigrants were roughly 250,000 in the 1950s, 320,000 in the 1960s, and 500,000 or more during most of the 1970s and early 1980s.

These data are very imperfect, however, because they do not include illegal immigrants. Estimates suggest that, in recent years, as many as 500,000 illegal

aliens may enter the United States each year, most coming from Mexico, the Caribbean, and Latin America. Despite this large annual influx of illegals, the total number of illegal aliens in the United States may only be about 3½ to 5 million (estimates vary from 2 to 12 million). Many illegal aliens come to the United States for a year or so to earn a "grubstake" and then return to their native countries.

Current legislation increases the number of legal immigrants to 700,000 per year. The legislation stresses family reunification by allowing United States citizens to bring in immediate relatives—spouses, children, and parents. But there has also been a substantial increase in the number of visas made available to highly skilled professionals such as researchers, engineers, and scientists. In addition, some 10,000 visas have been earmarked for wealthy immigrants who are willing to invest at least \$1 million in the United States. Emphasis is clearly on the kinds of immigrants who are likely to make significant contributions to American economic growth.

Immigration and immigration policy have been highly controversial, focusing largely on illegal immigrants. Public concern over illegal immigration gave rise to the Immigration Reform and Control Act of 1986, popularly known as the **Simpson-Rodino Act**. It has three major provisions.

1 Amnesty The law provides amnesty and grants legal status to undocumented individuals who have lived in the United States since 1982. Qualified workers receive work authorization cards and after five years of continuous residence become eligible to apply for citizenship.

2 Employer Sanctions Employers who knowingly hire illegal immigrants are subject to fines and possible imprisonment for repeated offenses.

3 Temporary Farm Labor The law allows temporary migrants or "guest workers" to enter the country to harvest perishable crops.

Economics of Immigration

We can gain some insight into the economic effects of immigration by employing a variation of the crowding model of discrimination (Figure 36-3). In Figure 36-4 we portray the demand for labor in the United States as D_u in the left diagram and the demand for labor in Mexico as D_m in the right diagram. The demand for

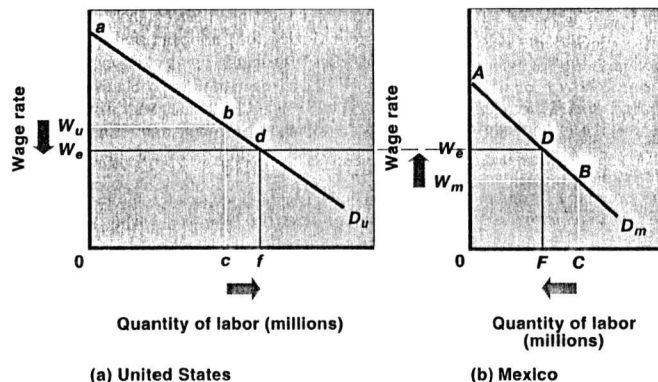


FIGURE 36-4 The simple economics of immigration

The migration of labor to high-income United States (a) from low-income Mexico (b) will increase the domestic output, reduce the average level of wages, and increase business incomes in the United States, while having the opposite effects in Mexico. The United States' domestic output gain of cbd exceeds Mexico's domestic output loss of fdb ; hence, there is a net increase in world output.

labor is greater in the United States, presumably because of the presence of more capital equipment and more advanced technologies which enhance the productivity of labor. (Recall from Chapter 27 that the labor demand curve is based on the marginal revenue product of labor.) Conversely, we assume that machinery and equipment are scarce in Mexico and that technology is less sophisticated; hence, labor demand is weak. We also assume that the premigration labor forces of the United States and Mexico are Oc and OC respectively, and that full employment exists in both countries.

Wage Rates and World Output If we further assume that (1) migration is costless; (2) it occurs solely in response to wage differentials; and (3) is unimpeded by legislation in either country, workers will migrate from Mexico to the United States until wage rates in the two countries are equal at W_e . In this case some $FC (= fc)$ million workers will have migrated from Mexico to the United States before equilibrium is achieved. Although the average level of wage rates falls from W_u to W_e in the United States, the domestic output (the sum of the marginal revenue products of the labor force) increases from $Oabc$ to $Oadf$. In Mexico, average wage rates rise from W_m to W_e , but domestic output declines

from $OABC$ to $OADF$.⁸ Observing that the domestic output gain of cbd in the United States exceeds the fdb loss in Mexico, we conclude that the world's real output has increased.

Just as elimination of the barrier of sex or racial discrimination enhances economic efficiency within a country, so the elimination of legislative barriers to the international flow of labor increases worldwide economic efficiency. The world gains because freedom to migrate moves people to countries where they can make a larger contribution to world production. To repeat: Migration involves an efficiency gain. It enables the world to produce a larger real output with a given amount of resources.

Income Shares Our model also suggests that this flow of immigrants will enhance business or capitalist incomes in the United States and reduce them in Mexico. We have just noted that the before-immigration

⁸What happens to the wage bill (wage rate multiplied by the number of workers) in each of the two countries depends on the elasticity of labor demand. If the demand for labor is elastic in the $W_u W_e$ wage range in the United States, the absolute size of the wage bill will increase. Conversely, if labor demand is inelastic in the $W_u W_e$ wage range, the absolute size of the wage bill will decline. A similar application of the total revenue (earnings) test for elasticity applies to Mexico.

domestic output in the United States is $Oabc$. The total wage bill is $OW_u bc$, that is, the wage rate multiplied by the number of workers. The remaining triangular area $W_u ab$ is "business" or capitalist income. The same reasoning applies to Mexico.

Unimpeded immigration will increase business income from $W_u ab$ to $W_e ad$ in the United States and reduce it from $W_m AB$ to $W_e AD$ in Mexico. Business benefits from immigration in the United States; Mexican businesses are hurt by emigration. This is what we would expect intuitively; America is receiving "cheap" labor, Mexico is losing "cheap" labor. This conclusion is consistent with the historical fact that American employers have often actively recruited immigrants.

Complications and Modifications

Our model includes a number of simplifying assumptions and also omits several relevant considerations. Let's therefore release some of the more critical assumptions and introduce omitted factors, observing how our conclusions are affected.

1 Cost of Migration The international movement of workers is not costless. Costs are not only the explicit or out-of-pocket costs of geographically moving oneself and one's possessions, but also the implicit or opportunity cost of lost income during the period of movement and reestablishing oneself in the host country. Still more subtle costs are involved in adapting to a new culture, language, climate, and so forth. All such costs must be estimated by the potential immigrant and weighed against the expected benefits of higher wages in the host country. If benefits are estimated to exceed costs, it is rational to migrate. If costs exceed benefits, one should not migrate.

In Figure 36-4 the existence of migration costs means that the flow of labor from Mexico to the United States will not occur to the extent that wages are equalized. Wages will remain higher in the United States than in Mexico. Furthermore, the world gain from migration will be reduced.

2 Remittances and Backflows Many migrants view their moves as temporary. Their plan is to move to a wealthier country, accumulate some desired level of wealth through hard work and frugality, and return home to establish their own enterprises. During their period in the host country, migrants frequently make sizable remittances to their families at home. This causes a redistribution of the net gain from migration

between the countries involved. In Figure 36-4 remittances by Mexican workers in the United States to their relatives would cause the gain in United States' domestic output to be less than that shown and the loss of Mexican domestic output to also be less than that shown.

Actual backflows—the return of migrants to their home countries—might also alter gains and losses through time. For example, if some of the Mexican workers who migrated to the United States acquired substantial labor-market or managerial skills and then returned home, their enhanced human capital might then make a substantial contribution to economic development in Mexico. Evidence suggests, however, that migrant workers who acquire skills in the receiving country tend not to return home. In fact, at various times the United States has been a beneficiary of "brain drains" as professional and other highly skilled workers have left western Europe and other nations for higher wages and better job opportunities in the United States.

3 Full Employment versus Unemployment Our model assumes full employment in both the sending and receiving country. Mexican workers presumably leave low-paying jobs to more-or-less immediately take higher-paying jobs in the United States. However, in many cases the factor that "pushes" immigrants from their homelands is not low wages, but chronic unemployment and underemployment. Many less developed countries are characterized by overpopulation and surplus labor; workers are either unemployed or so grossly underemployed that their marginal revenue product is zero.

Again, allowance for this possibility affects our discussion of gains and losses. Specifically, Mexico would gain (not lose!) by having such workers emigrate. These unemployed workers are making no contribution to Mexico's domestic output and must be sustained by transfers from the rest of the labor force. The remaining Mexican labor force will be better off by the amount of the transfers after the unemployed workers have migrated to the United States. Conversely, if the Mexican immigrant workers are unable to find jobs in the United States and are sustained through transfers from employed American workers, then the after-tax income of native American workers will decline.

4 Fiscal Aspects What impacts do immigrants have on tax revenues and government spending in the receiving country? Although evidence is scanty and

690

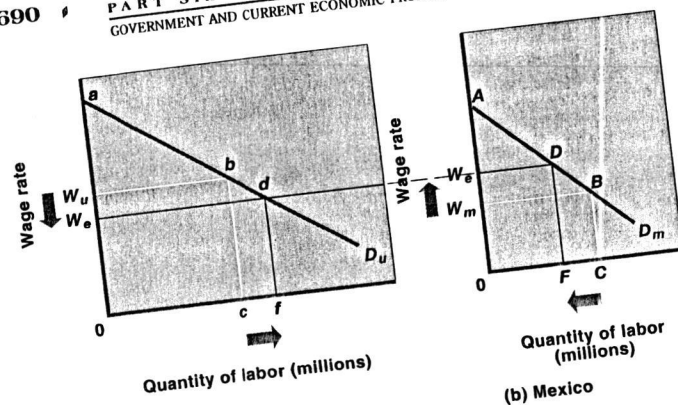
PART SIX
GOVERNMENT AND CURRENT ECONOMIC PROBLEMS

FIGURE 36-4 The simple economics of immigration. The migration of labor to high-income United States (a) will increase the domestic output, reduce the average level of wages, and increase business incomes in the United States, while having the opposite effects in Mexico. The United States' domestic output gain of $cbdf$ exceeds Mexico's domestic output loss of $FDBC$; hence, there is a net increase in world output.

FIGURE

The high workers there for while the equal to tion of la

of labor comes in the w . N_1N_2 work assume they secure em the nonuni to W_s .

Recall marginal re other words the domestic $A + B + C$ in domestic decline in tha of the MRP output—of th wage increase ment of these an increase in d areas $D + E$. $D + E$, there is precisely, beca tributable to the B . Since the sa

labor is greater in the United States, presumably because of the presence of more capital equipment and more advanced technologies which enhance the productivity of labor. (Recall from Chapter 27 that the labor demand curve is based on the marginal revenue product of labor.) Conversely, we assume that machinery and equipment are scarce in Mexico and that technology is less sophisticated; hence, labor demand is weak. We also assume that the premigration labor forces of the United States and Mexico are Oc and OC respectively, and that full employment exists in both countries.

Wage Rates and World Output If we further assume that (1) migration is costless; (2) it occurs solely in response to wage differentials; and (3) is unimpeded by legislation in either country, workers will migrate from Mexico to the United States until wage rates in the two countries are equal at W_e . In this case some $FC (= fe)$ million workers will have migrated from Mexico to the United States before equilibrium is achieved. Although the average level of wage rates falls from W_m to W_e in the United States, the domestic output (the sum of the marginal revenue products of the labor force) increases from $Oabc$ to $Oadf$. In Mexico, average wage rates rise from W_m to W_e , but domestic output declines

from $OABC$ to $OADF$.⁸ Observing that the domestic output gain of $cbdf$ in the United States exceeds the $FDBC$ loss in Mexico, we conclude that the world output has increased.

Just as elimination of the barrier of sex or discrimination enhances economic efficiency in a country, so the elimination of legislative barriers to international flow of labor increases worldwide economic efficiency. The world gains because free migration moves people to countries where they can make a larger contribution to world production. Migration involves an efficiency gain. It is the world to produce a larger real output with the same amount of resources.

Income Shares Our model also suggests that the flow of immigrants will enhance business or workers' incomes in the United States and reduce their incomes in Mexico. We have just noted that the before-immi-

⁸What happens to the wage bill (wage rate multiplied by the number of workers) in each of the two countries depends on the elasticity of the labor demand. If the demand for labor is elastic in the United States, the absolute size of the wage bill will increase. Conversely, if labor demand is inelastic in the United States, the absolute size of the wage bill will decline. A similar argument applies to Mexico.

domestic output in the United States is $Oabc$. The total wage bill is OW_ebc , that is, the wage rate multiplied by the number of workers. The remaining triangular area W_eab is "business" or capitalist income. The same reasoning applies to Mexico.

Unimpeded immigration will increase business income from W_eab to W_ead in the United States and reduce it from W_mAB to W_eAD in Mexico. Business profits from immigration in the United States; Mexican businesses are hurt by emigration. This is what we would expect intuitively; America is receiving "cheap" labor, or, Mexico is losing "cheap" labor. This conclusion is consistent with the historical fact that American employers have often actively recruited immigrants.

Implications and Modifications

Our model includes a number of simplifying assumptions and also omits several relevant considerations. We therefore release some of the more critical assumptions and introduce omitted factors, observing our conclusions are affected.

Cost of Migration The international movement of workers is not costless. Costs are not only the explicit out-of-pocket costs of geographically moving oneself and one's possessions, but also the implicit or opportunity cost of lost income during the period of movement. Establishing oneself in the host country. Still, subtle costs are involved in adapting to a new language, climate, and so forth. All such costs must be estimated by the potential immigrant and weighed against the expected benefits of higher wages in the host country. If benefits are estimated to exceed costs, it is rational to migrate. If costs exceed benefits, workers would not migrate.

Figure 36-4 the existence of migration costs would not alter the flow of labor from Mexico to the United States. The flow of labor from Mexico to the United States will not occur to the extent that wages are equalized. The flow of labor from Mexico to the United States will remain higher in the United States than in Mexico. Furthermore, the world gain from migration will be reduced.

Immigrants and Backflows Many migrants view their stay in the United States as temporary. Their plan is to move to their home country, accumulate some desired level of capital, and then return to their home country. During their stay in the United States, migrants frequently make investments to their families at home. This investment of the net gain from migration

between the countries involved. In Figure 36-4 remittances by Mexican workers in the United States to their relatives would cause the gain in United States' domestic output to be less than that shown and the loss of Mexican domestic output to also be less than that shown.

Actual backflows—the return of migrants to their home countries—might also alter gains and losses through time. For example, if some of the Mexican workers who migrated to the United States acquired substantial labor-market or managerial skills and then returned home, their enhanced human capital might then make a substantial contribution to economic development in Mexico. Evidence suggests, however, that migrant workers who acquire skills in the receiving country tend not to return home. In fact, at various times the United States has been a beneficiary of "brain drains" as professional and other highly skilled workers have left western Europe and other nations for higher wages and better job opportunities in the United States.

3 Full Employment versus Unemployment Our model assumes full employment in both the sending and receiving country. Mexican workers presumably leave low-paying jobs to more-or-less immediately take higher-paying jobs in the United States. However, in many cases the factor that "pushes" immigrants from their homelands is not low wages, but chronic unemployment and underemployment. Many less developed countries are characterized by overpopulation and surplus labor; workers are either unemployed or so grossly underemployed that their marginal revenue product is zero.

Again, allowance for this possibility affects our discussion of gains and losses. Specifically, Mexico would gain (not lose!) by having such workers emigrate. These unemployed workers are making no contribution to Mexico's domestic output and must be sustained by transfers from the rest of the labor force. The remaining Mexican labor force will be better off by the amount of the transfers after the unemployed workers have migrated to the United States. Conversely, if the Mexican immigrant workers are unable to find jobs in the United States and are sustained through transfers from employed American workers, then the after-tax income of native American workers will decline.

4 Fiscal Aspects What impacts do immigrants have on tax revenues and government spending in the receiving country? Although evidence is scanty and

LAST WORD

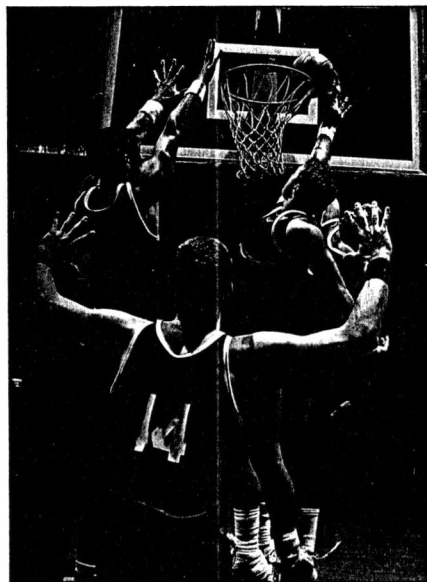
RACISM IN PROFESSIONAL BASKETBALL*

Although black players earn more than white players in the NBA, researchers have discovered evidence of wage discrimination against blacks.

Casual observation would suggest an absence of racial discrimination in the National Basketball Association (NBA). About three-fourths of all NBA players—and four-fifths of all starters—are black. Teams are highly integrated. There are more black coaches in the NBA than in other professional sports. Many of the most highly paid players are black. Indeed, raw salary data for 1985–1986 show that black players earned \$10,620 (2.7 percent) more than white players.

Yet recent research suggests that discrimination *does* exist. Sherer and Kahn* have adjusted 1985–1986 raw salary data for various measures of player performance (productivity) such as number of seasons played, games played per season, career points, field goal percentage, rebounds, assists per game, and so forth. These measures indicate that black players are superior to whites. Adjusting salaries to account for this superiority, Sherer and Kahn have concluded that black players earned about \$80,000 or 20 percent less than white players.

What is the source of this discrimination? Sherer and Kahn reject the notion of racist attitudes on the part of team owners. By rejecting talented black players, racist owners would find themselves with less successful teams, declining revenues, and franchises of lesser value. Furthermore, the fact that NBA teams are highly integrated suggests that fellow employees (white players) are not the source of discrimination. Sherer and Kahn find that team customers (fans) are the



source of NBA discrimination. Their research shows that home game attendance increases with the number of white players on the team. Specifically, they estimate that a team's revenue may increase by about \$115,000 to \$131,000 per season per additional white player, and suggest that both white players and team owners gain by serving fans' preferences to watch white players. Sherer and Kahn conclude that "As long as fans prefer to see white players, profit-oriented teams will make discriminatory salary offers."[†]

[†]Ibid., p. 60.

refugees from Haiti or Central America—may require several years of public or philanthropic aid to learn the language and become assimilated. But illegal aliens from, say, Mexico or the Caribbean are likely to be net taxpayers.

Several studies have found that very few illegals collect unemployment compensation, go on welfare, receive food-stamps, or use Medicaid. Some do use free public hospitals and send their children to pub-

lic schools, but the incremental costs involved are probably small. On the tax revenue side, it is clear that most illegals do have social security and federal income taxes withheld from their pay, although a sizable proportion apparently pay less than their legal obligation of the latter.

The very low incidence of social welfare payments to illegals is not a mystery. These payments are usually made only to the unemployed and most illegals are working. When they are not, fear of detection and deportation keeps them from applying to benefit programs. Thus, the direct social welfare costs of illegals are low.⁹

QUICK REVIEW 36-4

- *Current policy allows about 700,000 immigrants to enter the United States each year with preference being given to "priority workers" who have special education, training, and talents.*
- *A country receiving immigrants will experience a larger domestic output, a lower average level of wages, and an increase in business incomes. The opposite effects will occur in the donor country.*
- *The economic effects of immigration are complicated by the costs of migration, remittances, backflows of immigrants, employment conditions, and fiscal implications.*

CHAPTER SUMMARY

- 1 The growth of labor unions was slow and irregular until the 1930s due to court and employer hostility.
- 2 The AFL dominated the American labor movement from 1886 until the CIO was formed in 1936. Its philosophy was essentially that of Samuel Gompers—business unionism, political neutrality, and craft unionism.
- 3 Union growth was rapid in the 1930s and 1940s. The shift toward industrial unionism, triggered by the formation of the CIO in 1936, was a significant factor in this growth. Equally important were the pro-labor legislation passed by the Federal government in the 1930s and the wartime prosperity of the 1940s.
- 4 The Norris-La Guardia Act of 1932 rendered yellow-dog contracts unenforceable and sharply limited the use of injunctions in labor disputes. The Wagner Act of 1935—"labor's Magna Charta"—guaranteed labor the rights to organize and to bargain collectively with management.
- 5 The Taft-Hartley Act of 1947 brought about a shift from government-sponsored to government-regulated collective

⁹Walter Fogel, "Illegal Alien Workers in the United States," *Industrial Relations*, October 1977, p. 255. To the extent that illegal immigrants displace American workers they may impose indirect costs on our welfare and income-maintenance programs.

Economics and Beyond

Even our elementary consideration of the economic aspects of immigration makes it clear that the issues involved are complex. Much obviously depends on the character of the immigrants themselves and economic conditions in the receiving country. Many benefits would accrue to the United States from a liberal immigration policy when (1) immigrants are young, educated, and skilled; and (2) our economy is fully employed and experiencing robust growth. Benefits to Americans are less evident when (1) immigrants are unskilled and illiterate, and (2) our economy is stagnant and plagued by high unemployment. The "brain drain" to America in the 1960s and Castro's emptying out of Cuban prisons and mental hospitals in 1980 were highly contrasting episodes in American immigration history!

The immigration issue is also complicated by an assortment of essentially noneconomic issues. Do minority immigrants inflame racial problems? Does the concentration of migrants in urban areas such as New York and Miami generate social tensions and increase crime rates? Would America without fresh inflows of immigrants somehow not be America?

bargaining. The act **a** specifically outlaws certain "unfair practices" of unions, **b** regulates certain internal operations of unions, **c** controls the content of collective bargaining agreements, and **d** outlines a procedure for handling "national health and welfare" strikes.

6 The Landrum-Griffin Act of 1959 was designed to regulate the internal processes of unions—in particular the handling of union finances and the union's relationships with its members.

7 Unionism has declined relatively in the United States since the mid-1950s. Some labor economists attribute this to changes in the composition of domestic output and in the demographic structure of the labor force. Others contend that employers, recognizing that unionization results in lower profitability, have more aggressively sought to dissuade workers from being union members.

8 Labor and management "live together" under the terms of collective bargaining agreements. These work agreements cover four major topics: **a** union status and managerial prerogatives; **b** wages and hours; **c** seniority and job control; and **d** a grievance procedure.

9 Union workers currently enjoy wages which are 10 to 15 percent higher than comparable nonunion workers. There is little evidence to suggest that unions have been

subject to dispute, the consensus is that immigrants are probably net contributors to the fiscal system of the host country. They are disproportionately comprised of younger people—frequently, unattached young males—who are in the prime of life and who have received some schooling in their native country. Certainly, highly skilled migrants who already speak the host country's language are likely to be heavy net taxpayers. Less skilled workers—for example, political

*Peter D. Sherer and Lawrence M. Kahn, "Racial Differences in Professional Basketball Players' Compensation," *Journal of Labor Economics*, January 1988, pp. 40–61.

TABLE 37-1 Exports of goods and services as a percentage of gross domestic product, selected countries, 1990

Country	Exports	
	Percentage of GDP	
The Netherlands	57	
Germany	36	
New Zealand	27	
Canada	25	
United Kingdom	24	
France	22	
Italy	19	
Japan	11	

Source: IMF, *International Financial Statistics*, 1992.

domestic markets cannot produce with reasonable efficiency the variety of goods they want to consume. For such countries, exports are the route for obtaining goods they desire and therefore exports may run from 25 to 35 percent or more of their GDPs. Other countries—the United States, for example—have rich and diversified resource bases and vast internal markets and are therefore less dependent on world trade.

1 Volume For the United States and the world the volume of international trade has been increasing both absolutely and relatively. Table 37-2 reflects the substantial growth in the dollar volume of both American exports and imports over the past three decades. Since 1960 United States' exports and imports of goods and services have more than doubled as a percentage of our GDP. Exports and imports currently are each about 10 to 11 percent of GDP. Curiously, however, the United States accounts for a diminishing percentage of total world trade. In 1947 it supplied about one-third of the world's total exports compared to about one-seventh today. World trade has increased more rapidly

for other nations than it has for the United States. *But in terms of absolute volumes of imports and exports the United States is the world's leading trading nation.*

2 Dependence There can be no question as to the United States' dependence on the world economy. We are almost entirely dependent on other countries for bananas, cocoa, coffee, spices, tea, raw silk, nickel, tin, natural rubber, and diamonds. Casual observation suggests that imported goods compete strongly in many of our domestic markets: Japanese cameras and video recorders, French and Italian wines, English bicycles, and Japanese motorcycles and autos are a few cases in point. Foreign cars have made persistent gains in American markets and now account for about 35 percent of total sales in the United States. Even the great American pastime—baseball—relies heavily on imported gloves.

But world trade is a two-way street, and a host of American industries are highly dependent on foreign markets. Almost all segments of agriculture rely heavily on foreign markets—rice, wheat, cotton, and tobacco exports vary from one-fourth to more than one-half of total output. The chemical, aircraft, automobile, machine tool, coal, and computer industries are only a few of many American industries which sell significant portions of their output in international markets. Table 37-3 shows some of the major commodity exports and imports of the United States.

3 Trade Patterns An overall picture of the pattern of United States merchandise trade was given in Table 5-6. A quick review of that table provides the basis for several observations.

1 In 1990 our imports of goods from abroad were substantially in excess of our exports of goods.

2 The bulk of our export and import trade is with other developed nations, not with the less developed nations or the countries of eastern Europe.

TABLE 37-2 Trade in the U.S. economy, 1960–1991* (dollars in billions)

	1960		1975		1991	
	Amount	Percent of GDP	Amount	Percent of GDP	Amount	Percent of GDP
Exports of goods and services	\$25.3	4.9	\$136.3	8.6	\$592.5	10.4
Imports of goods and services	22.8	4.4	122.7	7.7	621.9	11.0
Net exports	2.4	0.5	13.6	0.9	–29.4	0.5

*Data are on a national income accounts basis.

Source: Department of Commerce.

TABLE 37-3 Principal commodity exports and imports of the United States, 1990 (in billions)

Exports	Amount	Imports	Amount
Chemicals	\$28.4	Petroleum	\$62.1
Computers	25.9	Automobiles	45.9
Consumer durables	21.0	Clothing	23.9
Aircraft	18.4	Computers	23.0
Grains	14.9	Household appliances	18.7
Semiconductors	13.3	Chemicals	14.3
Generating equipment	12.7	Semiconductors	12.2
Automobiles	10.9	Iron and steel	11.3
Nonferrous metals	10.9	Toys and sporting goods	9.7
Telecommunications	9.6	Telecommunications	9.4

3 Canada is our most important trading partner quantitatively. Twenty-two percent of our exports are sold to Canadians, who in turn provide us with 19 percent of our imports.

4 There is a sizable imbalance in our trade with Japan; our imports greatly exceed our exports.

5 Our dependence on foreign oil is reflected in the excess of imports over exports in our trade with the OPEC nations.

4 Level of Output Changes in net exports—the difference between the value of a nation's exports and imports—have multiple effects on the level of domestic output in roughly the same way as do fluctuations in the various types of domestic spending. A small change in the volume of American imports and exports can have magnified repercussions on the domestic levels of output, employment, and prices.

Unique Aspects

Aside from essentially quantitative considerations, world trade has certain unique characteristics.

1 Mobility Differences Though the difference is a matter of degree, the mobility of resources is considerably less among nations than it is within nations. American workers are free to move from Iowa to Idaho or from Maine to Minnesota. Crossing international boundaries is a different story.

Immigration laws, not to mention language and cultural barriers, severely restrict migration of labor between nations. Different tax laws, different governmental regulations, different business practices, and a

host of other institutional barriers limit migration of real capital over international boundaries.

International trade is a substitute for the international mobility of resources. If human and property resources do not move readily among nations, the movement of goods and services is an effective substitute.

2 Currency Differences Each nation uses a different currency. An American firm distributing Hondas or BMWs in the United States must buy yen or marks to pay the Japanese or German automobile manufacturers. The possible complications which may accompany this exchange of currencies are explored in Chapter 38.

3 Politics As we will note, international trade is subject to political interferences and controls which differ markedly in degree and kind from those applying to domestic trade.

THE ECONOMIC BASIS FOR TRADE

But why do nations trade? What is the basis for trade between nations? *International trade is a way nations can specialize, increase the productivity of their resources, and realize a larger total output than otherwise.* Sovereign nations, like individuals and regions of a nation, can gain by specializing in products they can produce with greatest relative efficiency and by trading for goods they cannot produce efficiently.

While the above rationale for world trade is correct, it in a sense begs the question. A better answer to

pletely in wheat (point *W* in Figure 37-2a) and Brazil completely in coffee (point *c* in Figure 37-2b).

Improved Options Now, instead of being constrained by its domestic production possibilities line and having to give up 1 ton of wheat for every ton of coffee it wants as it moves up its domestic production possibilities line from point *W*, the United States, through trade with Brazil, can get $1\frac{1}{2}$ tons of coffee for every ton of wheat it exports to Brazil as it moves up the trading possibilities line *WC'*.

Similarly, we can think of Brazil as starting at point *c*, and instead of having to move down its domestic production possibilities line and having to give up 2 tons of coffee for each ton of wheat it wants, it can now export just $1\frac{1}{2}$ tons of coffee for each ton of wheat it wants by moving down its *cw'* trading possibilities line.

Specialization and trade give rise to a new exchange ratio between wheat and coffee which is reflected in a nation's trading possibilities line. This new exchange ratio is superior for both nations to the self-sufficiency exchange ratio in the production possibilities line of each. By specializing in wheat and trading for Brazil's coffee, the United States can obtain *more than* 1 ton of coffee for 1 ton of wheat. Similarly, by specializing in coffee and trading for United States' wheat, Brazil can get 1 ton of wheat for *less than* 2 tons of coffee.

Added Output The crucial point is that by specializing according to comparative advantage and trading for those goods produced with the least relative efficiency domestically, both the United States and Brazil can realize combinations of wheat and coffee beyond their production possibilities boundaries. *Specialization according to comparative advantage results in a more efficient allocation of world resources, and larger outputs of both wheat and coffee are therefore available to the United States and Brazil.* To be more specific, suppose that at the $1W = 1\frac{1}{2}C$ terms of trade, the United States exports 10 tons of wheat to Brazil and Brazil in return exports 15 tons of coffee to the United States.

How do the new quantities of wheat and coffee available to the two nations compare with the optimal product-mixes that existed before specialization and trade? Point *A* in Figure 37-2a reminds us that the United States chose 18 tons of wheat and 12 tons of coffee originally. But, by producing 30 tons of wheat and no coffee, and by trading 10 tons of wheat for 15 tons of coffee, the United States can enjoy 20 tons of wheat and 15 tons of coffee. This new, superior combi-

nation of wheat and coffee is shown by point *A'* in Figure 37-2a. Compared with the nontrading figures of 18 tons of wheat and 12 tons of coffee, the United States' **gains from trade** are 2 tons of wheat and 3 tons of coffee. Similarly, we assumed Brazil's optimal product-mix was 4 tons of coffee and 8 tons of wheat (point *B*) before specialization and trade. Now, by specializing in coffee—producing 20 tons of coffee and no wheat—Brazil can realize a combination of 5 tons of coffee and 10 tons of wheat by exporting 15 tons of its coffee in exchange for 10 tons of American wheat. This new position is shown by point *B'* in Figure 37-2b. Brazil's gains from trade are 1 ton of coffee and 2 tons of wheat.

As a result of specialization and trade, both countries have more of both products. Table 37-4 is a summary statement of these figures and merits your careful study.

The fact that points *A'* and *B'* are economic positions superior to *A* and *B* is extremely important. Recall from Chapter 2 that a given nation can expand its production possibilities boundary by (1) expanding the quantity and improving the quality of its resources or (2) realizing technological progress. We have now discovered another means—international trade—by which a nation can circumvent the output constraint imposed by its production possibilities curve. The effects of international specialization and trade are tantamount to having more and better resources or discovering improved production techniques.

Increasing Costs

In formulating a straightforward statement of the principles underlying international trade, we have made several simplifications. Our discussion was purposely limited to two products and two nations to minimize verbiage, but multination and multiproduct examples yield similar conclusions. The assumption of constant costs, on the other hand, is a more substantive simplification. Let's therefore consider the significance of increasing costs (concave-from-the-origin production possibility curves) for our analysis.

Suppose, as in our previous constant-cost illustration, that the United States and Brazil are at positions on their production possibilities curves where their cost ratios are initially $1W = 1C$ and $1W = 2C$ respectively. As before, comparative advantage indicates that the United States should specialize in wheat and Brazil in coffee. But now, as the United States begins to expand wheat production, its $1W = 1C$ cost ratio will fall; it will have to sacrifice *more than* 1 ton of coffee to get 1

additional ton of wheat. Resources are no longer perfectly shiftable between alternative uses, as the constant-cost assumption implied. Resources less and less suitable to wheat production must be allocated to the American wheat industry in expanding wheat output, and this means increasing costs—the sacrifice of larger and larger amounts of coffee for each additional ton of wheat.

Similarly, Brazil, starting from its $1W = 2C$ cost ratio position, expands coffee production. But as it does, it will find that its $1W = 2C$ cost ratio begins to rise. Sacrificing a ton of wheat will free resources which are only capable of producing something *less than* 2 tons of coffee, because these transferred resources are less suitable to coffee production.

As the American cost ratio falls from $1W = 1C$ and Brazil's rises from $1W = 2C$, a point will be reached at which the cost ratios are equal in the two nations, for example, at $1W = 1\frac{1}{2}C$. At this point the underlying basis for further specialization and trade—differing cost ratios—has disappeared, and further specialization is therefore uneconomic. And most important, this point of equal cost ratios may be realized where the United States is still producing some coffee along with its wheat and Brazil is producing some wheat along with its coffee. *The primary effect of increasing costs is to make specialization less than complete.* For this reason we often find domestically produced products competing directly against identical or similar imported products within a particular economy.

The Case for Free Trade Restated

The compelling logic of the case for free trade is hardly new. Indeed, in 1776 Adam Smith asserted:

It is the maxim of every prudent master of a family, never to attempt to make at home what it will cost him more to make than to buy. The taylor does not attempt to make his own shoes, but buys them of the shoemaker. The shoemaker does not attempt to make his own clothes but employs a taylor. The farmer attempts to make neither the one nor the other, but employs those different artificers. All of them find it for their interest to employ their whole industry in a way in which they have some advantage over their neighbors, and to purchase with a part of its produce, or what is the same thing, with the price of a part of it, whatever else they have occasion for.¹

¹Adam Smith, *The Wealth of Nations* (New York: Modern Library, Inc., 1937), p. 424.

In modern jargon, the case for free trade comes down to this one potent argument. *Through free trade based on the principle of comparative advantage, the world economy can achieve a more efficient allocation of resources and a higher level of material well-being.* The resource mixes and technological knowledge of each country are different. Therefore, each nation can produce particular commodities at different real costs. Each nation should produce goods for which its opportunity costs are low relative to those of other nations and exchange these specialties for products for which its opportunity costs are high relative to those of other nations. If each nation does this, the world can realize fully the advantages of geographic and human specialization. That is, the world—and each free-trading nation—can obtain a larger real income from the fixed supplies of resources available to it. Protection—barriers to free trade—lessens or eliminates gains from specialization. If nations cannot freely trade, they must shift resources from efficient (low-cost) to inefficient (high-cost) uses to satisfy their diverse wants.

A side benefit of free trade is that it promotes competition and deters monopoly. The increased competition afforded by foreign firms forces domestic firms to adopt the lowest-cost production techniques. It also compels them to be innovative and progressive with respect to both product quality and production methods, thereby contributing to economic growth. And free trade gives consumers a wider range of products from which to choose. The reasons to favor free trade are essentially the same reasons which endorse competition. Therefore, it is not surprising that most economists embrace the case for free trade as an economically valid position.

QUICK REVIEW 37-1

- ♦ *World trade is increasingly important to the United States and other nations of the world.*
- ♦ *International trade enables nations to specialize, enhance the productivity of their resources, and obtain a larger output.*
- ♦ *The principle of comparative advantage states that total world output will be greatest when each good is produced by that nation having the lowest opportunity cost.*
- ♦ *Specialization is less than complete among nations because opportunity costs normally rise as more of a particular good is produced.*

TRADE BARRIERS

No matter how compelling the logic of the case for free trade, barriers to free trade do exist.

1 Tariffs are excise taxes on imported goods: they may be imposed for purposes of revenue or protection. **Revenue tariffs** are usually applied to products not produced domestically, for example, tin, coffee, and bananas in the case of the United States. Rates on revenue tariffs are typically modest and their purpose is to provide the Federal government with tax revenues.

Protective tariffs, on the other hand, are designed to shield domestic producers from foreign competition. Although protective tariffs are usually not high enough to prohibit importation of foreign goods, they put foreign producers at a competitive disadvantage in selling in domestic markets.

2 Import quotas specify the maximum amounts of commodities which may be imported in any period of time. Frequently, import quotas are more effective in retarding international commerce than tariffs. A given product might be imported in relatively large quantities despite high tariffs; low import quotas, on the other hand, completely prohibit imports once quotas are filled.

3 Nontariff barriers (NTBs) refer to licensing requirements, unreasonable standards pertaining to product quality and safety, or simply unnecessary bureaucratic red tape in customs procedures. Japan and the European countries frequently require their domestic importers of foreign goods to obtain licenses. By restricting the issuance of licenses, imports can be effectively restricted. Great Britain bars importation of coal in this way.

4 Voluntary export restrictions (VERs) are a relatively new trade barrier by which foreign firms "voluntarily" limit the amount of their exports to a particular country. VERs, which have the effect of import quotas, are agreed to by exporters in the hope of avoiding more stringent trade barriers. Thus Japanese auto manufacturers agreed to a VER on exports to the United States under the threat of higher U.S. tariffs or the imposition of low import quotas.

Motivations: Special-Interest Effect

If tariffs and quotas impede free trade and diminish economic efficiency, why do we have them? While nations as a whole gain from free international trade, particular industries and groups of resource suppliers can be hurt. In our comparative advantage example, spe-

cialization and trade adversely affected the American coffee industry and the Brazilian wheat industry. Such groups may seek to preserve or improve their economic positions by persuading the government to impose tariffs or quotas to protect them from the detrimental effects of free trade. Chapter 32's special-interest effect—or concept of rent-seeking activity—is highly relevant.

The direct beneficiaries of import relief or export subsidy are usually few in number, but each has a large individual stake in the outcome. Thus, their incentive for vigorous political activity is strong.

But the costs of such policies may far exceed the benefits. It may cost the public \$40,000–\$50,000 a year to protect a domestic job that might otherwise pay an employee only half that amount in wages and benefits. Furthermore, the costs of protection are widely diffused—in the United States, among 50 States and [254] million citizens. Since the cost borne by any one citizen is likely to be quite small, and may even go unnoticed, resistance at the grass-roots level to protectionist measures often is considerably less than pressures for their adoption.²

Also, the costs of protectionism are hidden because tariffs and quotas are embedded in the prices of goods. Thus policy makers face fewer political restraints in responding positively to demands for protectionism.

Later in this chapter we will consider the specific arguments and appeals made to justify protection.

Economic Impact of Tariffs

Simple supply and demand analysis is useful in examining the economic effects of protective tariffs. The D_d and S_d curves in Figure 37-3 show domestic demand and supply for a product in which the United States has a comparative disadvantage, for example, cassette recorders. (Disregard $S_d + Q$ for now.) Without world trade, the domestic price and output would be OP_d and Oq , respectively.

Assume now that the domestic economy is opened to world trade, and that the Japanese, who have a comparative advantage in cassette recorders and dominate the world market, begin to sell their recorders in the United States. We assume that with free trade the domestic price cannot differ from the lower world price, which here is OP_w . At OP_w domestic consumption is Od , domestic production is Oa , and the difference between the two, ad , reflects imports.

²Economic Report of the President, 1982, p. 177.

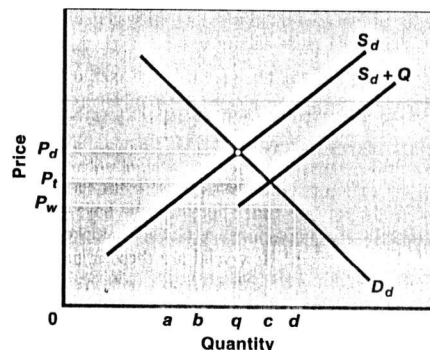


FIGURE 37-3 The economic effects of a protective tariff or an import quota

A tariff of $P_w P_t$ will reduce domestic consumption from Od to Oc . Domestic producers will be able to sell more output (Ob rather than Oa) at a higher price (OP_t rather than OP_w). Foreign exporters are injured because they sell less output (bc rather than ad) in the United States. The orange area indicates the amount of tariffs paid by American consumers. An import quota of bc units will have the same effects as the tariff, with one exception: the orange area will go to foreign producers rather than to the U.S. Treasury.

Direct Effects Suppose now that the United States imposes a tariff of $P_w P_t$ per unit on the imported recorders. This will raise the domestic price from OP_w to OP_t and will have several effects.

1 Decline in Consumption Consumption of recorders in the United States will decline from Od to Oc as the higher price moves buyers up their demand curve. The tariff prompts consumers to buy fewer recorders; that is, to reallocate a portion of their expenditures to less desired substitute products. American consumers are clearly injured by the tariff, since they pay $P_w P_t$ more for each of the Oc units which they now buy at price P_t .

2 Increased Domestic Production American producers—who are *not* subject to the tariff—will receive a higher price of OP_t per unit. Because this new price is higher than the pretariff or world price of OP_w , the domestic recorder industry will move up its supply curve S_d , increasing domestic output from Oa to Ob . Domestic producers will enjoy both a higher price and expanded sales. These effects explain the interest of domestic producers in lobbying for protective tariffs. From a social point of view, however, the expanded

domestic production of ab reflects the fact that the tariff permits domestic producers of recorders to bid resources away from other, more efficient, industries.

3 Decline in Imports Japanese producers will be hurt. Although the sales price of recorders is higher by $P_w P_t$, that increase accrues to the United States government and not to Japanese producers. The after-tariff world price, and thus the per unit revenue to Japanese producers, remains at OP_w , while the volume of United States imports (Japanese exports) falls from ad to bc .

4 Tariff Revenue Finally, note that the orange rectangle indicates the amount of revenue which the tariff yields. Specifically, total revenue from the tariff is determined by multiplying the tariff of $P_w P_t$ per unit by the number of imported recorders, bc . This tariff revenue is essentially a transfer of income from consumers to government and does not represent any net change in the nation's economic well-being. The result is that government gains this portion of what consumers lose.

Indirect Effects There are more subtle effects of tariffs which go beyond our supply and demand diagram. Because of diminished sales of recorders in the United States, Japan will now earn fewer dollars with which to buy American exports. That is, American export industries—industries in which the United States has a comparative advantage—will cut production and release resources. These are highly efficient industries, as evidenced by their comparative advantage and ability to sell goods in world markets. In short, *tariffs directly promote the expansion of relatively inefficient industries which do not have a comparative advantage and indirectly cause the contraction of relatively efficient industries which do have a comparative advantage*. This means that tariffs cause resources to be shifted in the wrong direction. This is not surprising. We know that specialization and unfettered world trade based on comparative advantage would lead to the efficient use of world resources and an expansion of the world's real output. The purpose and effect of protective tariffs are to reduce world trade. Therefore, aside from their specific effects on consumers and foreign and domestic producers, tariffs diminish the world's real output.

Economic Impact of Quotas

We noted earlier that an import quota is a legal limit placed on the amount of some product which can be imported each year. Quotas have the same economic

impact as a tariff with one big difference: While tariffs generate revenue for the United States government, a quota transfers that revenue to foreign producers.

Suppose in Figure 37-3 that, instead of imposing a tariff of $P_w P_i$ per unit, the United States prohibits any Japanese imports of recorders in excess of bc units. In other words, an import quota of bc recorders is imposed on Japan. Note that we have deliberately chosen the size of this quota to be the same amount as imports would be under a $P_w P_i$ tariff, so we are comparing "equivalent" situations. As a consequence of the quota, the supply of recorders is $S_d + Q$ in the United States. This is comprised of the domestic supply plus the constant amount $bc (=Q)$ which importers will provide at each domestic price.³

Most of the economic results are the same as with a tariff. Recorder prices are higher (P_i instead of P_w) because imports have been reduced from ad to bc . Domestic consumption of recorders is down from Od to Oc . American producers enjoy both a higher price (P_i rather than P_w) and increased sales (Ob rather than Oa).

The critical difference is that the price increase of $P_w P_i$ paid by American consumers on imports of bc —that is, the orange area—no longer goes to the United States Treasury as tariff (tax) revenue, but rather flows to those Japanese firms which have acquired the rights to sell recorders in the United States. For Americans, a tariff produces a better economic outcome than a quota, other things being the same. A tariff generates government revenue which can be used to cut other taxes or to finance public goods and services which benefit Americans. In contrast, the higher price created by quotas results in additional revenue for foreign producers.

It is relevant that in the early 1980s the American automobile industry with the support of its workers successfully lobbied for an import quota on Japanese autos. The Japanese government in turn apportioned this quota among its various auto producers. The restricted supply of Japanese cars in the American market allowed Japanese manufacturers to increase their prices and, hence, their profits. The American import quotas in effect provided Japanese auto manufacturers with a cartel-like arrangement which enhanced their profits. It is significant that when American import quotas were dropped in the mid-1980s, the Japanese gov-

³The $S_d + Q$ supply curve does not exist below price P_w because Japanese producers would not export recorders to the United States at any price below P_w when they can sell them to other countries at the world market prices of P_w .

ernment replaced them with its own system of export quotas for Japanese automakers.

THE CASE FOR PROTECTION: A CRITICAL REVIEW

Although free-trade advocates prevail in the classroom, protectionists sometimes dominate the halls of Congress. What arguments do protectionists make to justify trade barriers? How valid are these arguments?

Military Self-Sufficiency Argument

The argument here is not economic but of a political-military nature: Protective tariffs are needed to preserve or strengthen industries producing strategic goods and materials essential for defense or war. It plausibly contends that in an uncertain world, political-military objectives (self-sufficiency) must take precedence over economic goals (efficiency in the allocation of world resources).

Unfortunately, there is no objective criterion for weighing the relative worth of the increase in national security on the one hand, and the decrease in productive efficiency on the other, which accompany reallocation of resources toward strategic industries when such tariffs are imposed. The economist can only point out that certain economic costs are involved when tariffs are levied to enhance military self-sufficiency.

Although we might all agree that it is probably not a good idea to import our missile guidance systems from China, the self-sufficiency argument is nevertheless open to serious abuse. Virtually every industry can directly or indirectly claim a contribution to national security. Can you name an industry which did *not* contribute in some small way to World War II? Aside from abuses, are there not better ways than tariffs to provide for needed strength in strategic industries? When achieved through tariffs, self-sufficiency creates costs in the form of higher domestic prices on the output of the shielded industry. The cost of enhanced military security is apportioned arbitrarily among those consumers who buy the industry's product. A direct subsidy to strategic industries, financed out of general tax revenues, would more equitably distribute these costs.

Increase Domestic Employment

This "save American jobs" argument for tariffs becomes increasingly fashionable as an economy en-

counters a recession. It is rooted in macro analysis. Aggregate expenditures in an open economy are comprised of consumption expenditures (C) plus investment expenditures (I_p) plus government expenditures (G) plus net export expenditures (X_n). Net export expenditures consist of exports (X) minus imports (M). By reducing imports, M , aggregate expenditures will rise, stimulating the domestic economy by boosting income and employment. But there are important shortcomings associated with this policy.

1 Job Creation from Imports While imports may eliminate some American jobs, they create others. Imports may have eliminated jobs of American steel and textile workers in recent years, but others have gained jobs selling Hondas and imported electronics equipment. While import restrictions alter the composition of employment, they may actually have little or no effect on the volume of employment.

2 Fallacy of Composition All nations cannot simultaneously succeed in import restriction; what is true for *one* nation is not true for *all* nations. The exports of one nation must be the imports of another. To the extent that one country is able to stimulate its economy through an excess of exports over imports, another economy's unemployment problem is worsened by the resulting excess of imports over exports. It is no wonder that tariff and import quotas to achieve domestic full employment are termed "beggar my neighbor" policies. They achieve short-run domestic goals by making trading partners poorer.

3 Retaliation Nations adversely affected by tariffs and quotas are likely to retaliate, causing a competitive raising of trade barriers which will choke off trade to the end that all nations are worse off. The **Smoot-Hawley Tariff Act of 1930**, which imposed the highest tariffs ever enacted in the United States, backfired miserably. Rather than stimulate the American economy, this tariff act only induced a series of retaliatory restrictions by adversely affected nations. This caused a further contraction of international trade and lowered the income and employment levels of all nations.

4 Long-Run Feedbacks In the long run an excess of exports over imports is doomed to failure as a device for stimulating domestic employment. It is through American imports that foreign nations earn dollars with which to purchase American exports. In the long run a nation must import in order to export. The long-run impact of tariffs is not to increase domestic employ-

ment but at best to reallocate workers away from export industries and toward protected domestic industries. This shift implies a less efficient allocation of resources.

In summary, the argument that tariffs increase net exports and therefore create jobs is misleading:

Overall employment in an economy is determined by internal conditions and macroeconomic policies, not by the existence of trade barriers and the level of trade flows. The United States created [more than 18] million payroll jobs over the course of the [1982–1990] economic expansion, a period of U.S. trade deficits and relatively open U.S. markets. During the same period the European Community (EC) created virtually no net new jobs, even though they experienced trade surpluses. The same level of employment can be obtained in the total absence of free trade as when trade is completely free. But without foreign trade a nation will be worse off economically because, in effect, it will throw away part of its productive capability—the ability to convert surplus goods into other goods through foreign trade.⁴

Diversification for Stability

Closely related to the increase-domestic-employment argument for tariff protection is the diversification-for-stability argument. The point here is that highly specialized economies—for example, Saudi Arabia's oil economy or Cuba's sugar economy—are highly dependent on international markets for their incomes. Wars, cyclical fluctuations, and adverse changes in the structure of industry will force large and frequently painful readjustments on such economies. Tariff and quota protection is therefore allegedly needed to promote greater industrial diversification and consequently less dependence on world markets for just one or two products. This will help insulate the domestic economy from international political developments, depressions abroad, and from random fluctuations in world supply and demand for one or two particular commodities, thereby providing greater domestic stability.

There is some truth in this argument. There are also serious qualifications and shortcomings.

- 1 The argument has little or no relevance to the United States and other advanced economies.
- 2 The economic costs of diversification may be great; for example, one-crop economies may be highly inefficient in manufacturing.

⁴Economic Report of the President, 1988, p. 131. Updated

Infant-Industry Argument

The infant-industry argument contends that protective tariffs are needed to allow new domestic industries to establish themselves. Temporarily shielding young domestic firms from the severe competition of more mature and therefore currently more efficient foreign firms will give infant industries a chance to develop and become efficient producers.

This argument for protection rests on an alleged exception to the case for free trade. The exception is that all industries have not had, and in the presence of mature foreign competition, will never have, the chance to make long-run adjustments in the direction of larger scale and greater efficiency in production. Tariff protection for infant industries will therefore correct a current misallocation of world resources now perpetuated by historically different levels of economic development between domestic and foreign industries.

Counterarguments Though the infant-industry argument has logical validity, these qualifying points must be noted.

- 1 In the less developed nations it is very difficult to determine which industries are the infants capable of achieving economic maturity and therefore deserving of protection.
- 2 Protective tariffs may persist even after industrial maturity has been realized.
- 3 Most economists feel that if infant industries are to be subsidized, there are better means than tariffs for doing it. Direct subsidies, for example, have the advantage of making explicit which industries are being aided and to what degree.

Strategic Trade Policy In recent years the infant-industry argument has taken a modified form in advanced economies. The contention is that government should use trade barriers strategically to reduce the risk of product development borne by domestic firms, particularly products involving advanced technology. Firms protected from foreign competition can grow more rapidly and therefore achieve greater economies of scale than unprotected foreign competitors. Thus, the protected firms can eventually dominate world markets because of lower costs. Supposedly, dominance of world markets will enable the domestic firms to return high profits to the home nation. These profits allegedly will exceed the domestic sacrifices caused by trade barriers. Also, specialization in high-technology industries supposedly is beneficial because technology

advances achieved in one domestic industry often can be transferred to other domestic industries.

Japan and South Korea, in particular, have been accused of using this form of **strategic trade policy**. The problem with this strategy and therefore this argument for tariffs is that the nations put at a disadvantage by strategic trade policies tend to retaliate with tariffs of their own. The outcome may be higher tariffs worldwide, reductions in world trade, and loss of the gains from specialization and exchange.

Protection Against "Dumping"

The protection-against-dumping argument for tariffs contends that tariffs are needed to protect American firms from foreign producers which "dump" excess goods onto the American market at less than cost. Two reasons have been suggested as to why foreign firms might wish to sell in America at below cost.

- 1 These firms may use **dumping** to drive out American competitors, obtain monopoly power, and then raise prices. The long-term economic profits resulting from this strategy may more than offset the earlier losses which accompany the dumping.
- 2 Dumping may be a form of price discrimination—charging different prices to different customers. The foreign seller may find it can maximize its profits by charging a high price in its monopolized domestic market while unloading its surplus output at a lower price in the United States. The surplus output may be needed to obtain the overall per unit cost saving associated with large-scale production.

Because dumping is a legitimate concern, it is prohibited under American trade law. Where dumping occurs and is shown to injure American firms, the Federal government imposes tariffs called "antidumping duties" on the specific goods. But relative to the number of goods exported to the United States, documented cases of dumping are few. Dumping therefore does *not* justify widespread, permanent tariffs. Furthermore, allegations of dumping require careful investigation to determine their validity. Foreign producers often argue that dumping allegations and antidumping duties are an American method of restricting legitimate trade. The fact is that some foreign firms can produce certain goods at substantially less cost than American competitors, and what on the surface may seem to be dumping often is comparative advantage at work. If abused, the antidumping law can increase the price of imports and restrict competition in the American market. This reduced competition allows Ameri-

can firms to raise prices at consumers' expense. And even where true dumping does occur, American consumers gain from the lower-priced product—at least in the short term—much as they gain from a price war among American producers.

Cheap Foreign Labor

The cheap-foreign-labor argument holds that domestic firms and workers must be shielded from the ruinous competition of countries where wages are low. If protection is not provided, cheap imports will flood American markets and the prices of American goods—along with the wages of American workers—will be pulled down and our domestic living standards reduced.

This argument can be rebutted at several levels. The logic of the argument would suggest that it is *not* mutually beneficial for rich and poor persons to trade with one another. However, that is not the case. A low-income farm worker may pick lettuce or tomatoes for a rich landowner and both may benefit from the transaction. And don't American consumers gain when they buy a Taiwanese vest pocket radio for \$12 as opposed to a qualitatively similar American-made radio selling for \$20?

Also, recall that gains from trade are based on comparative advantage. Looking back at Figure 37-1, suppose the United States and Brazil have labor forces of exactly the same size. Noting the positions of the production possibilities curves, we observe that American labor is absolutely more productive because our labor force can produce more of either good. Because of this greater productivity, we can expect wages and living standards to be higher for American labor. Conversely, Brazil's less-productive labor will receive lower wages.

The cheap-foreign-labor argument would suggest that, to maintain our standard of living, America should not trade with low-wage Brazil. Suppose we follow this suggestion. Will wages and living standards rise in the United States as a result? The answer is "No." To obtain coffee America will now have to reallocate a portion of its labor from its relatively efficient wheat industry to its relatively inefficient coffee industry. As a result, the average productivity of American labor will fall as will real wages and living standards. In fact, the labor forces of *both* countries will have diminished standards of living because without specialization and trade they will have less output available to them. Compare column 4 with column 1 in Table 37-4 or points A' and B' with A and B in Figure 37-2 to confirm this point.

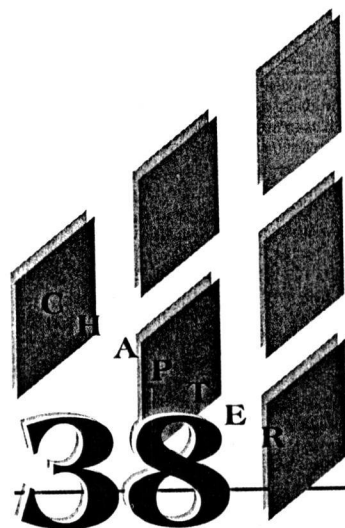
A Summing Up

The arguments for protection are numerous, but they are not weighty. Under proper conditions, the infant-industry argument stands as a valid exception, justifiable on economic grounds. And on political-military grounds, the self-sufficiency argument can be used to validate protection. Both arguments, however, are susceptible to severe abuses, and both neglect alternative means of fostering industrial development and military self-sufficiency. Most other arguments are semi-emotional appeals in the form of half-truths and outright fallacies. These arguments note only the immediate and direct consequences of protective tariffs. They ignore the fact that in the long run a nation must import in order to export.

There is also compelling historical evidence suggesting that free trade has led to prosperity and growth and that protectionism has had the opposite effects. Several examples follow.

- 1 The United States Constitution forbids individual states from levying tariffs, making America a huge free-trade area. Economic historians acknowledge this is an important positive factor in the economic development of our nation.
- 2 Great Britain's movement toward freer international trade in the mid-nineteenth century was instrumental in its industrialization and growth in that century.
- 3 As we will see, the creation of the Common Market in Europe after World War II has eliminated tariffs among member nations. Economists agree that creation of this free-trade area has been an important ingredient in the western European prosperity of recent decades.
- 4 More generally, the trend toward tariff reduction since the mid-1930s has been a stimulus to post-World War II expansion of the world economy.
- 5 We have already noted that the high tariffs imposed by our Smoot-Hawley Act of 1930 and the retaliation which it engendered worsened the Great Depression of the 1930s.
- 6 Studies of less developed countries overwhelmingly suggest that those which have relied on import restrictions to protect their domestic industries have realized slow growth in comparison to those pursuing more open economic policies.⁵

⁵Examples are from *Economic Report of the President 1985*, pp. 115-117.



Exchange Rates, the Balance of Payments, and Trade Deficits

If you take an American dollar to the bank and ask to exchange it for United States currency, you will get a puzzled look. If you persist, you may get in exchange another dollar. One American dollar can buy exactly one American dollar. But, as of January 24, 1992, one United States dollar could buy 5340 Turkish lira, 1.34 Australian dollars, .56 British pounds, 1.16 Canadian dollars, 5.48 French francs, 1.61 German marks, 123.93 Japanese yen, or 5.84 Swedish krona. What explains this seemingly haphazard array of exchange rates?

In Chapter 37 we examined comparative advantage as the underlying economic basis of world trade and discussed the effects of barriers to free trade. In this chapter we first introduce the monetary or financial aspects of international trade. How are currencies of different nations exchanged when import and export transactions occur? Second, we analyze and interpret a nation's international balance of payments. What is meant by a "favorable" or "unfavorable" balance of trade? Third, the kinds of exchange rate systems which trading nations have used are explained and evaluated. In this discussion we examine the polar extremes of freely flexible and fixed exchange rates and then survey actual systems which have existed historically. Finally, we explore the balance of trade deficits the United States has encountered over the past decade.

FINANCING INTERNATIONAL TRADE

A basic feature distinguishing international from domestic payments is that two different national currencies are exchanged. When American firms export goods to British firms, the American exporter wants to be paid in dollars. But British importers have pounds sterling. The problem, then, is to exchange pounds for

dollars to permit the American export transaction to occur.

This problem is resolved in *foreign exchange markets* where dollars can be used to purchase British pounds, Japanese yen, German marks, Italian lira, and so forth, and vice versa. Sponsored by major banks in New York, London, Zurich, Tokyo, and elsewhere, foreign exchange markets facilitate American exports and imports.

American Export Transaction

Suppose an American exporter agrees to sell \$30,000 worth of computers to a British firm. Assume that the *rate of exchange*—the rate or price at which pounds can be exchanged for, or converted into, dollars, and vice versa—is \$2 for £1. This means that the British importer must pay £15,000 to the American exporter. Let's summarize what occurs in terms of simple bank balance sheets (Figure 38-1).

a To pay for the American computers, the British buyer draws a check on its demand deposit in a London bank for £15,000. This is shown by the -£15,000 demand deposit entry in the righthand side of the balance sheet of the London bank.

b The British firm then sends this £15,000 check to the American exporter. But the American exporting firm must pay its employees and materials suppliers, as well as its taxes, in dollars, not pounds. So the exporter sells the £15,000 check or draft on the London bank to a large American bank, probably in New York City, which is a dealer in foreign exchange. The American firm is given a \$30,000 demand deposit in the New York bank in exchange for the £15,000 check. Note the new demand deposit entry of +\$30,000 in the New York bank.

c What does the New York bank do with the £15,000? It deposits it in a correspondent London bank for future sale. Thus, +£15,000 of demand deposits appear in the liabilities column of the balance sheet of the London bank. This +£15,000 (\$30,000) is an asset as viewed by the New York bank. To simplify, we assume that the correspondent bank in London is the

same bank from which the British importer obtained the £15,000 draft.

Note these salient points.

1 *American exports create a foreign demand for dollars, and the satisfaction of this demand generates a supply of foreign monies—pounds, in this case—held by American banks and available to American buyers.*

2 *The financing of an American export (British import) reduces the supply of money (demand deposits) in Britain and increases the supply of money in the United States by the amount of the purchase.*

American Import Transaction

But why would the New York bank be willing to give up dollars for pounds sterling? As just indicated, the New York bank is a dealer in foreign exchange; it is in the business of buying—for a fee—and, conversely, in selling—also for a fee—pounds for dollars.

Having just explained that the New York bank would buy pounds with dollars in connection with an American export transaction, we will now examine how it would sell pounds for dollars in financing an American import (British export) transaction. Suppose that an American retail concern wants to import £15,000 worth of woolens from a British mill. Again, simple commercial bank balance sheets summarize our discussion (Figure 38-2).

a Because the British exporting firm must pay its obligations in pounds rather than dollars, the American importer must exchange dollars for pounds. It does this by going to the New York bank and purchasing £15,000 for \$30,000—perhaps the American importer

FIGURE 38-1 Financing a U.S. export transaction

American export transactions create a foreign demand for dollars. The satisfaction of this demand increases the supplies of foreign monies held by American banks.

LONDON BANK	
Assets	Liabilities and net worth
	Demand deposit of British importer -£15,000(a)
	Deposit of New York bank +£15,000(c)

NEW YORK BANK	
Assets	Liabilities and net worth
Deposit in London bank +£15,000(c) (\$30,000)	Demand deposit of American exporter +\$30,000(b)

LONDON BANK		NEW YORK BANK	
Assets	Liabilities and net worth	Assets	Liabilities and net worth
	Demand deposit of British exporter +£15,000(b)	Deposit in London bank -£15,000(a) (\$30,000)	Demand deposit of American importer -\$30,000(a)
	Deposit of New York bank -£15,000(a)		

FIGURE 38-2 Financing a U.S. import transaction
American import transactions create an American demand for foreign monies. The satisfaction of that demand reduces the supplies of foreign monies held by American banks.

purchases the same £15,000 which the New York bank acquired in the previous American export transaction. In Figure 38-2, this purchase reduces the American importer's demand deposit in the New York bank by \$30,000 and the New York bank gives up its £15,000 deposit in the London bank.

b The American importer sends its newly purchased check for £15,000 to the British firm, which deposits it in the London bank. Note the +£15,000 deposit in the liabilities and net worth column of Figure 38-2.

We find that:

1 American imports create a domestic demand for foreign monies (pounds sterling, in this case) and that fulfillment of this demand reduces the supplies of foreign monies held by American banks.

2 An American import transaction increases the money supply in Britain and reduces the money supply in the United States.

By combining these two transactions, a further point comes into focus. American exports (computers) make available, or "earn," a supply of foreign monies for American banks, and American imports (British woolens) create a demand for these monies. In a broad sense, any nation's exports finance or "pay for" its imports. Exports provide the foreign currencies needed to pay for imports. From Britain's point of view, its exports of woolens earn a supply of dollars, which are then used to meet the demand for dollars associated with Britain's imports of computers.

Postscript: Although our examples are confined to the exporting and importing of goods, we will find that demands for and supplies of pounds also arise from transactions involving services and the payment of in-

terest and dividends on foreign investments. Thus Americans demand pounds not only to finance imports, but also to purchase insurance and transportation services from the British, to vacation in London, to pay dividends and interest on British investments in the United States, and to make new financial and real investments in Britain.

THE INTERNATIONAL BALANCE OF PAYMENTS

We now explore the wide variety of international transactions which create a demand for and generate a supply of a given currency. This spectrum of international trade and financial transactions is reflected in the United States' international balance of payments. A nation's balance of payments statement records all transactions which take place between its residents (including individuals, businesses, and governmental units) and the residents of all foreign nations. These transactions include merchandise exports and imports, tourist expenditures, purchases and sales of shipping and insurance services, interest and dividends received or paid abroad, purchases and sales of financial or real assets abroad, and so forth. The United States' balance of payments shows the balance between all the payments the United States receives from foreign countries and all the payments which we make to them. A simplified balance of payments for the United States in 1990 is shown in Table 38-1. Let's analyze this accounting statement to see what it reveals about our international trade and finance.

TABLE 38-1 The United States' balance of payments, 1990 (in billions)

Current account	
(1) U.S. merchandise exports	\$+390
(2) U.S. merchandise imports	-498
(3) Balance of trade	\$-108
(4) U.S. exports of services	+133
(5) U.S. imports of services	-107
(6) Balance on goods and services	-82
(7) Net investment income	+12
(8) Net transfers	-22
(9) Balance on current account	-92
Capital account	
(10) Capital inflows to the U.S.	+117*
(11) Capital outflows from the U.S.	-59
(12) Balance on capital account	+58
(13) Current and capital account balance	-34
(14) Official reserves	+34
	\$ 0

*Includes a \$64 billion statistical discrepancy which is believed to be comprised primarily of unaccounted capital inflows.

Source: Survey of Current Business, December 1991.

Current Account

The top portion of Table 38-1 summarizes the United States' trade in currently produced goods and services and is called the **current account**. Items 1 and 2 show American exports and imports of merchandise (goods) respectively in 1990. We have designated American exports with a *plus* sign and our imports with a *minus* sign because American merchandise exports (and other export-type transactions) are **credits** in that they create or earn supplies of foreign exchange. As we saw in our discussion of how international trade is financed, any export-type transaction obligating foreigners to make "inpayments" to the United States generates supplies of foreign monies in American banks.

Conversely, American imports (and other import-type transactions) are **debits**; they use up foreign exchange. Again, our earlier discussion of trade financing indicated that American imports obligate Americans to make "outpayments" to the rest of the world which draw down available supplies of foreign currencies held by American banks.

Trade balance Items 1 and 2 in Table 38-1 tell us that in 1990 our merchandise exports of \$390 billion did not earn enough foreign monies to finance our merchandise imports of \$498 billion. Specifically, the merchan-

dise balance of trade or, more simply, the **trade balance** refers to the difference between a country's merchandise exports and merchandise imports. If exports exceed imports, then a *trade surplus* or "favorable balance of trade" is being realized. If imports exceed exports, then a *trade deficit* or "unfavorable balance of trade" is occurring. We note in item 3 that in 1990 the United States incurred a trade deficit of \$108 billion.

Balance on Goods and Services Item 4 reveals that the United States not only exports autos and computers, but also sells transportation services, insurance, and tourist and brokerage services to residents of foreign countries. These service sales or "exports" totaled \$133 billion in 1990. Item 5 indicates that Americans buy or "import" similar services from foreigners. These service imports were \$107 billion in 1990.

The **balance on goods and services**, shown in Table 38-1 as item 6, is the difference between our exports of goods and services (items 1 and 4) and our imports of goods and services (items 2 and 5). In 1990 our exports of goods and services fell short of our imports of goods and services by \$82 billion.

Balance on Current Account Item 7 reflects that historically the United States has been a net international lender. Over time we have invested more abroad than

foreigners have invested in the United States. Thus net investment income represents the excess of interest and dividend payments which foreigners have paid us for the services of our exported capital over what we paid in 1990 in interest and dividends for their capital invested in the United States. Table 38-1 shows that, on balance, our net investment income earned us \$12 billion worth of foreign currencies for "exporting" the services of American money capital invested abroad.

Item 8 reflects net transfers, both public and private, from the United States to the rest of the world. Included here is American foreign aid, pensions paid to Americans living abroad, and remittances of immigrants to relatives abroad. These \$22 billion of transfers are "outpayments" and exhaust available supplies of foreign exchange. As it has been facetiously put, net transfers entail the importing of "goodwill" or "thank-you notes."

By taking all transactions in the current account into consideration we obtain the **balance on current account** shown by item 9 in Table 38-1. In 1990 the United States realized a current account deficit of \$92 billion. This means that our current account import transactions (items 2, 5, and 8) created a demand for a larger dollar amount of foreign currencies than our export transactions (items 1, 4, and 7) supplied.

Capital Account

The **capital account** reflects capital flows in the purchase or sale of real and financial assets which occurred in 1990. For example, Honda or Nissan might acquire an automobile assembly plant in the United States. Or, alternatively, the investments may be of a financial nature, for example, an Arabian oil sheik might purchase GM stock or Treasury bonds. In either event such transactions generate supplies of foreign currencies for the United States. They are therefore credit or inpayment items, designated with a plus sign. The United States is exporting stocks and bonds and thereby earning foreign exchange. Item 10 in Table 38-1 shows that such transactions amounted to \$117 billion in 1990.

Conversely, Americans invest abroad. General Electric might purchase a plant in Hong Kong or Singapore to assemble pocket radios or telephones. Or an American might buy stock in an Italian shoe factory. Or an American bank might finance construction of a meat processing plant in Argentina. These transactions have a common feature; they all use up or exhaust supplies

of foreign currencies. We therefore attach a minus sign to remind us that these are debit or outpayment transactions. The United States is importing stocks, bonds, and IOUs from abroad. Item 11 in Table 38-1 reveals that \$59 billion of these transactions occurred in 1990. When items 10 and 11 are combined, the **balance on the capital account** was a *plus* \$58 billion—the United States enjoyed a capital account surplus of \$58 billion in 1990.

Interrelationships

The current and capital accounts are interrelated; they are essentially reflections of one another. The current account *deficit* means that American exports of goods and services were not sufficient to pay for our imports of goods and services.¹ How did we finance the difference? The answer is that the United States must either borrow from abroad or give up ownership of some of its assets to foreigners as reflected in the capital account.

A simple analogy is useful here. Suppose in a given year your expenditures exceed your earnings. How will you finance your "deficit"? You might sell some of your assets or borrow. You might sell some real assets (your car or stereo) or perhaps some financial assets (stocks or bonds) which you own. Or you might obtain a loan from your family or a bank.

Similarly, when a nation incurs a deficit in its current account, its expenditures for foreign goods and services (its imports) exceed the income received from the international sales of its own goods and services (its exports). It must somehow finance that current account deficit by selling assets and by borrowing, that is, by going into debt. And that is what is reflected in the capital account surplus. Our capital account surplus of \$58 billion (item 12) indicates that in 1990 the United States "sold off" real assets (buildings, farmland) and received loans from the rest of the world in that amount to help finance our current account deficit of \$92 billion.

Recap: A nation's current account deficit will be financed essentially by a net capital inflow in its capital account. Conversely, a nation's current account *surplus* would be accompanied by a net capital *outflow* in its capital account. The excess earnings from its current account surplus will be used to purchase real assets of, and make loans to, other nations.

¹We ignore transfer payments (item 8) in making this statement.

Official Reserves

The central banks of nations hold quantities of foreign currencies called **official reserves** which are added to or drawn on to settle any *net* differences in current and capital account balances. In 1990 the surplus in our capital account was considerably less than the deficit in our current account so we had a \$34 billion net deficit on the combined accounts (item 13). That is, the United States earned less foreign monies in all international trade and financial transactions than it used. This deficiency of earnings of foreign currencies was subtracted from the existing balances of foreign monies held by our central banks. The *plus* \$34 billion of official reserves shown by item 14 in Table 38-1 represents this reduction of our stocks of foreign currencies. The plus sign indicates this is a credit or "export-type" transaction which represents a supply of foreign exchange.

Frequently the relationship between the current and capital account is just the opposite of that shown in Table 38-1. That is, the current account deficit is less than the capital account surplus. Hence, our central banks would experience an increase in their holdings of foreign currencies. This would show as a *minus* item in the balance of payments; it is a debit or "import-type" transaction because it represents a use of foreign exchange.

The important point here is that the three components of the balance of payments statement—the current account, the capital account, and the official reserves account—must sum to zero. Every unit of foreign exchange used (as reflected in our "minus" outpayment or debit transactions) in our international transactions must have a source (our "plus" inpayment or credit transactions).

Payments Deficits and Surpluses

Although the balance of payments must always sum to zero, economists and political officials frequently speak of **balance of payments deficits and surpluses**. In doing so they are referring to the "current and capital account balance" shown as item 13 in Table 38-1. If this is a negative item, a balance of payments deficit is being realized as was the case for the United States in 1990. In 1990 the United States earned less foreign monies from all its trade and financial transactions than it used. The United States did not "pay its way" in world trade and finance and therefore depleted its official reserves

of foreign monies. If the current and capital account balance were positive, then the United States would be faced with a balance of payments surplus. The United States would have earned sufficient foreign exchange from its export-type transactions to pay for its import-type transactions. As we have just seen, it would add to its stocks of foreign monies—that is, increase its official reserve holdings.

A decrease in official reserves (shown by a positive official reserves item in Table 38-1) measures a nation's balance of payments deficit; an increase in official reserves (shown by a negative official reserves item) measures its balance of payments surplus.

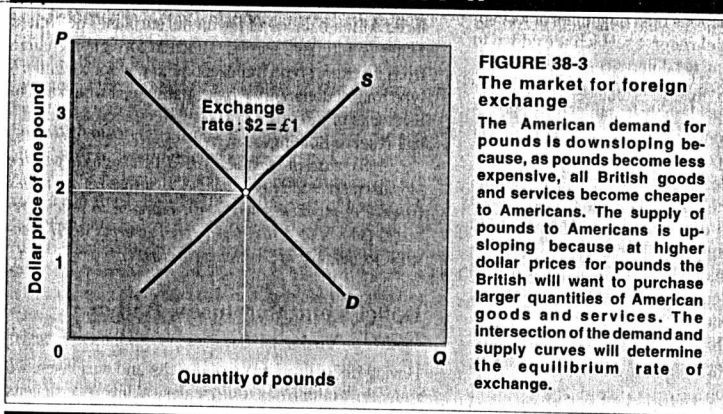
Deficits and Surpluses: Bad or Good?

Having defined a variety of deficits and surpluses, we must now inquire as to their desirability. Are deficits bad, as the term implies? Is a surplus desirable, as that word suggests? The answer to both questions is "not necessarily." A large merchandise trade deficit such as the United States has been incurring in recent years is regarded by many as "unfavorable" or "adverse," as it suggests American producers are losing their competitiveness in world markets. Our industries seem to be having trouble selling their goods abroad and are simultaneously facing strong competition from imported goods. On the other hand, a trade deficit is *favorable* from the vantage point of American consumers who are currently receiving more goods as imports than they are forgoing as exports.

Similarly, the desirability of a balance of payments deficit or surplus depends on (1) the events causing them and (2) their persistence through time. For example, the large payments deficits imposed on the United States and other oil-importing nations by OPEC's dramatic runup of oil prices in 1973–1974 and 1979–1980 were very disruptive in that they forced the United States to invoke policies to curtail oil imports.

Also, any nation's official reserves are limited. Persistent or long-term payments deficits, which must be financed by drawing down those reserves, would ultimately deplete reserves. In this case that nation would have to undertake policies to correct its balance of payments. These policies might require painful macroeconomic adjustments, trade barriers and similar restrictions, or changing the international value of its currency.

KEY GRAPH



QUICK REVIEW 38-1

- American exports create a demand for dollars and a supply of foreign currencies; American imports create a demand for foreign currencies and a supply of American dollars.
- The current account balance is a nation's exports of goods and services less its imports of goods and services plus its net investment income and net transfers.
- The capital account balance is a nation's capital inflows less its capital outflows.
- A balance of payments deficit occurs when the sum of the balances on current and capital accounts is negative; a balance of payments surplus arises when the sum of the balances on current and capital accounts is positive.

EXCHANGE RATE SYSTEMS AND BALANCE OF PAYMENTS ADJUSTMENTS

Both the size and persistence of a nation's balance of payments deficits and surpluses and the kind of adjustments it must make to correct these imbalances depend on the system of exchange rates being used. There are two polar options: (1) a system of flexible or floating exchange rates where the rates at which national currencies exchange for one another are deter-

mined by demand and supply, and (2) a system of rigidly fixed exchange rates by which governmental intervention in foreign exchange markets or some other mechanism offsets the changes in exchange rates which fluctuations in demand and supply would otherwise cause.

Freely Floating Exchange Rates

Freely floating exchange rates are determined by the unimpeded forces of demand and supply. Let's examine the rate, or price, at which American dollars might be exchanged for, say, British pounds sterling. As indicated in Figure 38-3 (Key Graph), the demand for pounds will be downsloping; the supply of pounds, upsloping.

The downsloping demand for pounds shown by *D* indicates that, if pounds become less expensive to Americans, British goods will become cheaper to Americans. Americans will demand larger quantities of British goods and therefore larger amounts of pounds to buy those goods.

The supply of pounds is upsloping, as *S*, because, as the dollar price of pounds rises (that is, the pound price of dollars falls), the British will purchase more American goods. At higher and higher dollar prices for pounds, the British can get more American dollars and therefore more American goods per pound. Thus, American goods become cheaper to the British, induc-

ing them to buy more of these goods. When the British buy American goods, they supply pounds to the foreign exchange market because they must exchange pounds for dollars to purchase our goods.

The intersection of the supply and demand for pounds will determine the dollar price of pounds. In this instance the equilibrium rate of exchange is \$2 to £1.

Depreciation and Appreciation An exchange rate determined by free-market forces can and does change frequently. When the dollar price of pounds increases, for example, from \$2 for £1 to \$3 for £1, the value of the dollar has depreciated relative to the pound. Currency depreciation means that it takes more units of a country's currency (dollars) to buy a single unit of some foreign currency (pounds).

Conversely, when the dollar price of pounds decreases—from \$2 for £1 to \$1 for £1—the value of the dollar has appreciated relative to the pound. Currency appreciation means that it takes fewer units of a country's currency (dollars) to buy a single unit of some foreign currency (pounds).

In our American-British illustrations, when the dollar depreciates the pound necessarily appreciates and vice versa. When the exchange rate between dollars and pounds changes from \$2 = £1 to \$3 = £1, it now takes more dollars to buy £1 and the dollar has depreciated. But it now takes fewer pounds to buy a dollar. At the initial rate it took £½ to buy \$1; at the new rate it only takes £⅓ to buy \$1. The pound has appreciated relative to the dollar. If the dollar depreciates vis-à-vis the pound, the pound appreciates vis-à-vis the dollar. Conversely, if the dollar appreciates vis-à-vis the pound, the pound depreciates vis-à-vis the dollar. These relationships are summarized in Figure 38-4.

Determinants of Exchange Rates Why are the demand for and the supply of pounds located as they are in Figure 38-3? What forces will cause the demand and supply curves for pounds to change, thereby causing the dollar to appreciate or depreciate?

Changes in Tastes Any change in consumer tastes or preferences for the products of a foreign country will alter the demand for, or supply of, that nation's currency and change its exchange rate. If American technological advances in computers make them more attractive to British consumers and businesses, then they will supply more pounds in exchange markets in purchasing more American computers and the dollar

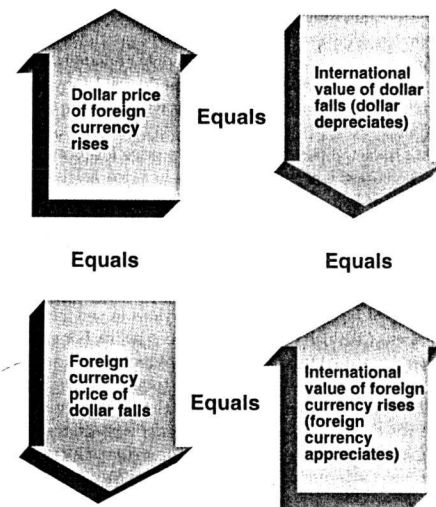


FIGURE 38-4 Currency appreciation and depreciation

An increase in the dollar price of foreign currency is equivalent to a decline in the international value of the dollar (dollar depreciates). An increase in the dollar price of foreign currency also implies a decline in the foreign currency price of dollars. That is, the international value of foreign currency rises relative to the dollar (foreign currency appreciates).

will appreciate. Conversely, if British tweeds become more fashionable in the United States, our demand for pounds will increase and the dollar will depreciate.

Relative Income Changes If the growth of a nation's national income is more rapid than other countries', its currency is likely to depreciate. A country's imports vary directly with its level of income. As incomes rise in the United States, American consumers buy more domestically produced goods and also more foreign goods. If the United States' economy is expanding rapidly and the British economy is stagnant, American imports of British goods—and therefore U.S. demand for pounds—will increase. The dollar price of pounds will rise, meaning the dollar has depreciated.

Relative Price Changes If the domestic price level rises rapidly in the United States and remains constant in Britain, American consumers will seek out relatively

low-priced British goods, increasing the demand for pounds. Conversely, the British will purchase fewer American goods, reducing the supply of pounds. This combination of an increase in the demand for, and a reduction in the supply of, pounds will cause the dollar to depreciate.

In fact, differences in relative price levels among nations—which reflect changes in price levels over time—help explain persistent differences in exchange rates. In 1992 an American dollar could buy .56 British pounds, 124 Japanese yen, or 5340 Turkish lira. One reason for these differences is that the prices of British goods and services in pounds were far lower than the prices of Japanese goods and services in yen and the prices of Turkish goods and services in lira. For example, the same market basket of products costing \$500 in the United States might cost 250 pounds in England, 67,500 yen in Japan, and 2,500,000 lira in Turkey. Generally, the higher the prices of a nation's goods and services in terms of its own currency, the greater the amount of that currency which can be obtained with an American dollar.

Taken to its extreme, this **purchasing power parity theory** holds that differences in exchange rates *equate* the purchasing power of various currencies. That is, the exchange rates among national currencies perfectly adjust in such a way as to equal the ratios of the nations' price levels. For example, if a market basket of goods costs \$100 in the United States and £50 in Great Britain, the exchange rate should be $\$2 = \pounds 1$. Thus, a dollar spent on goods sold in Britain, Japan, Turkey, and other nations supposedly will have equal purchasing power. In practice, however, exchange rates depart significantly from purchasing power parity, even over long periods. Nevertheless, relative price levels are clearly a major determinant of exchange rates.

Relative Real Interest Rates Suppose the United States restricts the growth of its money supply (tight money policy), as it did in the late 1970s and early 1980s, to control inflation. As a result, *real* interest rates—nominal interest rates adjusted for the rate of inflation—were high in the United States compared to most other nations. Consequently, British individuals and firms found the United States an attractive place to make financial investments. This increase in the demand for American financial assets meant an increase in the supply of British pounds and the dollar therefore appreciated in value.

Speculation Suppose it is widely anticipated that the American economy will (a) grow faster than the Brit-

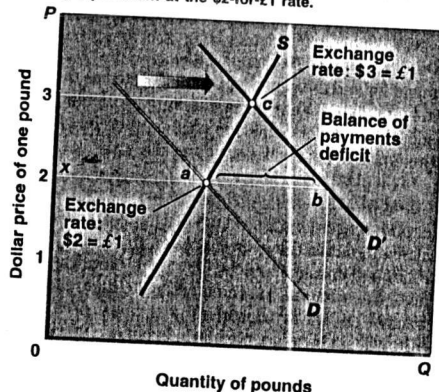
ish economy, (b) experience more rapid inflation than the British economy, and (c) have lower future real interest rates than Britain. All these expectations would lead one to believe that in the future the dollar will depreciate and, conversely, the pound will appreciate. Holders of dollars will thus attempt to convert them into pounds, increasing the demand for pounds. This conversion causes the dollar to depreciate and the pound to appreciate. A self-fulfilling prophecy arises: The dollar depreciates and the pound appreciates because speculators act on the supposition that these changes in currency values will in fact happen.

Flexible Rates and the Balance of Payments

Proponents of flexible exchange rates argue that such rates have a compelling virtue: *They automatically adjust so as eventually to eliminate balance of payments deficits or surpluses.* We can explain this by looking at S and D in Figure 38-5 which restate the demand for, and supply of, pounds curves from Figure 38-3. The equilibrium exchange rate of $\$2 = \pounds 1$ correctly suggests there is no balance of payments deficit or surplus. At the $\$2 = \pounds 1$ exchange rate the quantity of pounds demanded by Americans to import British goods, buy

FIGURE 38-5 Adjustments under flexible exchange rates, fixed exchange rates, and the gold

Under flexible rates an American trade deficit at the \$2-for-£1 rate would be corrected by an increase in the rate to \$3 for £1. Under fixed rates the *ab* shortage of pounds would be met out of international monetary reserves. Under the gold standard the deficit would cause changes in domestic price and income levels which would shift the demand for pounds (*D'*) to the left and the supply (*S*) to the right, sustaining equilibrium at the \$2-for-£1 rate.



British transportation and insurance services, and pay interest and dividends on British investments in the United States equals the amount of pounds supplied by the British in buying American exports, purchasing services from Americans, and making interest and dividend payments on American investments in Britain. In brief, there would be no change in official reserves in Table 38-1.

Now suppose tastes change and Americans decide to buy more British automobiles. Or assume that the American price level has increased relative to Britain, or that interest rates have fallen in the United States compared to Britain. Any or all of these changes will cause the American demand for British pounds to increase from D to, say, D' in Figure 38-5.

We observe that *at the initial* $\$2 = \pounds 1$ exchange rate an American balance of payments deficit has been created in the amount ab . That is, at the $\$2 = \pounds 1$ rate there is a shortage of pounds in the amount ab to Americans. American export-type transactions will earn xa pounds, but Americans will want xb pounds to finance import-type transactions. Because this is a free competitive market, the shortage will change the exchange rate (the dollar price of pounds) from $\$2 = \pounds 1$ to, say, $\$3 = \pounds 1$; that is, the dollar has *depreciated*.

At this point it must be emphasized that *the exchange rate is a very special price which links all domestic (United States') prices with all foreign (British) prices*. Specifically, the dollar price of a foreign good is found by multiplying the foreign product price by the exchange rate in dollars per unit of the foreign currency. At an exchange rate of \$2 = £1, a British Triumph automobile priced at 9000 will cost an American \$18,000 (= 9000 × \$2).

A change in the exchange rate therefore alters the prices of all British goods to Americans and all American goods to potential British buyers. Specifically, the change in the exchange rate from $\$2 = \pounds 1$ to $\$3 = \pounds 1$ will alter the relative attractiveness of American imports and exports in such a way as to restore equilibrium in the balance of payments of the United States. From the American point of view, as the dollar price of pounds changes from $\$2$ to $\$3$, the Triumph priced at $\pounds 9000$, which formerly cost an American $\$18,000$, now costs $\$27,000$ ($= 9000 \times 3$). Other British goods will also cost more to Americans, and American imports of British goods and services will decline. Graphically, this is shown as a move from point *b* toward point *c* in Figure 38-5.

Conversely, from Britain's standpoint the exchange rate, that is, the pound price of dollars, has fallen (from £2 to £1 for \$1). The international value of

the pound has *appreciated*. The British previously got only \$2 for £1; now they get \$3 for £1. American goods are therefore cheaper to the British, and American exports to Great Britain will rise. In Figure 38-5 this is shown by the move from point *a* toward point *c*.

The two adjustments described—a decrease in American imports from Great Britain and an increase in American exports to Great Britain—are precisely those needed to correct the American balance of payments deficit. (You should reason through the operation of freely fluctuating exchange rates in correcting an initial American balance of payments *surplus* in its trade with Great Britain.)

In summary, the free fluctuation of exchange rates in response to shifts in the supply of, and demand for, foreign monies automatically corrects balance of payments deficits and surpluses.

Disadvantages Even though freely fluctuating exchange rates automatically work eventually to eliminate payments imbalances, they may involve several significant problems:

1 Uncertainty and Diminished Trade The risks and uncertainties associated with flexible exchange rates may discourage the flow of trade. Suppose an American automobile dealer contracts to purchase ten Triumph cars for £90,000. At the current exchange rate of, say \$2 for £1, the American importer expects to pay \$180,000 for these automobiles. But if in the three-month delivery period the rate of exchange shifts to \$3 for £1, the £90,000 payment contracted by the American importer will now be \$270,000.

This unheralded increase in the dollar price of pounds may easily turn the potential American importer's anticipated profits into substantial losses. Aware of the possibility of an adverse change in the exchange rate, the American importer may not be willing to assume the risks involved. The American firm therefore may confine its operations to domestic automobiles, with the result that international trade does not occur in this item.

The same rationale applies to investment. Assume that, when the exchange rate is \$3 to £1, an American firm invests \$30,000 (or £10,000) in a British enterprise. It estimates a return of 10 percent, that is, it anticipates earnings of \$3000 or £1000. Suppose these expectations prove correct in that the British firm earns £1000 the first year on the £10,000 investment. But suppose that during the year, the value of the dollar *appreciates* to \$2 = £1. The absolute return is now only \$2000 (rather than \$3000) and the rate of return falls from the

anticipated 10 percent to only 6% percent ($= \$2000/\$30,000$). Investment is inherently risky. The added risk posed by adverse changes in exchange rates may persuade the potential American investor to avoid overseas ventures.²

2 Terms of Trade A nation's terms of trade will be worsened by a decline in the international value of its currency. For example, an increase in the dollar price of pounds will mean that the United States must export more goods and services to finance a given level of imports from Britain.

3 Instability Freely fluctuating exchange rates may also have destabilizing effects on the domestic economy as wide fluctuations stimulate and then depress those industries producing internationally traded goods. If the American economy is operating at full employment and the international value of its currency depreciates as in our illustration, the results will be inflationary for two reasons. Foreign demand for American goods will increase, that is, the net exports component of aggregate expenditures will increase and cause demand-pull inflation. Also, prices of all American imports will increase. Conversely, appreciation of the dollar would lower exports and increase imports, causing unemployment.

From the vantage point of policy, acceptance of floating exchange rates may complicate the use of domestic fiscal and monetary policies in seeking full employment and price stability. This is especially so for nations whose exports and imports are large relative to their GDPs (Table 37-1).

Fixed Exchange Rates

At the other extreme nations have often fixed or "pegged" their exchange rates to circumvent the disadvantages associated with floating rates. To analyze the implications and problems associated with fixed rates,

²At some cost and inconvenience a trader can circumvent part of the risk of unfavorable exchange rate fluctuations by "hedging" in the "futures market" for foreign exchange. For example, our American auto importer can purchase the needed pounds at the current \$2 for £1 exchange rates to be made available three months in the future when the British cars are delivered. Unfortunately, this does not eliminate entirely exchange rate risks. Suppose the dollar price of pounds falls (the dollar appreciates) in the three-month delivery period and a competing importing firm did not hedge its foreign exchange purchase. This means the competitor will obtain its shipment of Triumphs at a lower price and will be able to undersell our original importer.

assume that the United States and Britain agree to maintain a \$2 = £1 exchange rate.

The basic problem is that a governmental proclamation that a dollar will be worth so many pounds does *not* mandate stability of the demand for, and supply of, pounds. As demand and supply shift over time, government must intervene directly or indirectly in the foreign exchange market if the exchange rate is to be stabilized.

In Figure 38-5 suppose the American demand for pounds increases from D to D' and an American payments deficit of ab arises. This means that the American government is committed to an exchange rate ($\$2 = £1$) which is below the equilibrium rate ($\$3 = £1$). How can the United States prevent the shortage of pounds—reflecting an American balance of payments deficit—from driving the exchange rate up to the equilibrium level? The answer is to alter market demand or supply or both so that they continue to intersect at the $\$2 = £1$ rate of exchange. There are several means for achieving this.

1 Use of Reserves The most desirable means of pegging an exchange rate is to manipulate the market through the use of official reserves. International monetary reserves are stocks of foreign monies owned by a particular government. How do reserves originate? Let's assume that in the past the opposite market condition prevailed in which there was a surplus, rather than a shortage, of pounds, and the United States government had acquired that surplus. That is, at some earlier time the United States government spent dollars to buy surplus pounds which were threatening to reduce the $\$2 = £1$ exchange rate to, say, $\$1 = £1$. By now selling part of its reserve of pounds, the United States government could shift the supply of pounds curve to the right so that it intersects D' at b in Figure 38-5, thereby maintaining the exchange rate at $\$2 = £1$.

Historically nations have used gold as "international money" or, in other words, as reserves. Thus, in our example the United States government might sell some of the gold it owns to Britain for pounds. The pounds thus acquired could be used to augment the supply earned through American trade and financial transactions to shift the supply of pounds to the right to maintain the $\$2 = £1$ exchange rate.

It is critical that the amount of reserves be enough to accomplish the required increase in the supply of pounds. This is *not* a problem if deficits and surpluses occur more or less randomly and are of approximately

equivalent size. That is, last year's balance of payments surplus with Britain will increase the United States' reserve of pounds and this reserve can be used to "finance" this year's deficit. But if the United States encounters persistent and sizable deficits for an extended period, the reserves problem can become critical and force the abandonment of a system of fixed exchange rates. Or, at least, a nation whose reserves are inadequate must resort to less appealing options to maintain exchange rate stability. Let's consider these other options.

2 Trade Policies One set of policy options includes measures designed to control the flows of trade and finance directly. The United States might try to maintain the $\$2 = £1$ exchange rate in the face of a shortage of pounds by discouraging imports (thereby reducing the demand for pounds) and by encouraging exports (thereby increasing the supply of pounds). Imports can be reduced by imposing tariffs or import quotas. Similarly, special taxes may be levied on the interest and dividends Americans receive for foreign investments. Also, the United States government might subsidize certain American exports and thus increase the supply of pounds.

The fundamental problem with these policies is that they reduce the volume of world trade and distort its composition or pattern away from that which is economically desirable. Tariffs, quotas, and the like can be imposed only at the sacrifice of some portion of the economic gains or benefits attainable from a free flow of world trade based on comparative advantage. These effects should not be underestimated; the imposition of trade barriers can elicit retaliatory responses from other nations which are adversely affected.

3 Exchange Controls: Rationing Another option is exchange controls or rationing. Under exchange controls the United States government would handle the problem of a pound shortage by requiring that all pounds obtained by American exporters be sold to it. Then, in turn, the government allocates or rations this short supply of pounds (xa in Figure 38-5) among various American importers who demand the quantity xb . In this way the American government would restrict American imports to the amount of foreign exchange earned by American exports. American demand for British pounds in the amount ab would be unfulfilled. Government eliminates a balance of payments deficit by restricting imports to the value of exports.

There are many objections to exchange controls.

1 Like trade controls—tariffs, quotas, and export subsidies—exchange controls distort the pattern of international trade away from that based on comparative advantage.

2 The process of rationing scarce foreign exchange necessarily involves discrimination among importers. Serious problems of equity and favoritism are implicit in the rationing process.

3 Controls impinge on freedom of consumer choice. Americans who prefer Mazdas may be forced to buy Mercurys. The business opportunities of some American importers will necessarily be impaired because imports are being constrained by government.

4 There are likely to be enforcement problems. The market forces of demand and supply indicate there are American importers who want foreign exchange badly enough to pay *more* than the $\$2 = £1$ official rate; this sets the stage for extralegal or "black market" foreign exchange dealings.

4 Domestic Macro Adjustments A final means of maintaining a stable exchange rate is to use domestic fiscal and monetary policies to eliminate the shortage of pounds. In particular, restrictive fiscal and monetary measures will reduce the United States' national income relative to Britain's. Because American imports vary directly with our national income, our demand for British goods, and therefore for pounds, will be restrained.

To the extent that these contractionary policies reduce our price level relative to Britain's, American buyers of consumption and investment goods will divert their demands from British to American goods, also restricting the demand for pounds. Finally, a restrictive (tight) money policy will increase United States' interest rates compared to Britain and reduce American demand for pounds to make financial investments in Britain.

From Britain's standpoint lower prices on American goods and higher American interest rates will increase British imports of American goods and stimulate British financial investment in the United States. Both developments will increase the supply of pounds. The combination of a decrease in the demand for and an increase in the supply of pounds will eliminate the initial American payments deficit. In Figure 38-5 the new supply and demand curves will intersect at some new equilibrium point on the ab line where the exchange rate persists at $\$2 = £1$.

This means of maintaining pegged exchange rates is hardly appealing. The "price" of exchange rate stabil-

ity for the United States is falling output, employment, and price levels—in other words, a recession. Achieving a balance of payments equilibrium and realizing domestic stability are both important national economic objectives; but to sacrifice the latter for the former is to let the tail wag the dog.

QUICK REVIEW 38-2

1 In a system where exchange rates are free to float, they are determined by the demand for, and supply of, individual national currencies.

2 Determinants of freely floating exchange rates—factors which shift currency supply and demand curves—include changes in tastes, changes in relative national incomes, relative price level changes, relative real interest rate changes, and speculation.

3 Under a system of fixed exchange rates, nations set their exchange rates and then maintain them by buying or selling reserves of foreign currencies, incurring inflation or recession, establishing trade barriers, or employing exchange controls.

INTERNATIONAL EXCHANGE RATE SYSTEMS

There have been three different exchange rate systems which nations have employed in recent history.

The Gold Standard: Fixed Exchange Rates

Over the 1879–1934 period—except for the World War I years—an international monetary system known as the gold standard prevailed. The gold standard provided for fixed exchange rates. A look at its operation and ultimate downfall is instructive as to the functioning and some of the advantages and problems associated with fixed-rate systems. Currently a number of economists advocate fixed exchange rates and a few even call for a return to the international gold standard.

Conditions A nation is on the gold standard when it fulfills three conditions:

- 1 It must define its monetary unit in terms of a certain quantity of gold.
- 2 It must maintain a fixed relationship between its stock of gold and its domestic money supply.
- 3 It must allow gold to be freely exported and imported.

If each nation defines its monetary unit in terms of gold, the various national currencies will have a fixed relationship to one another. For example, suppose the United States defines a dollar as being worth 25 grains of gold and Britain defines its pound sterling as being worth 50 grains of gold. This means that a British pound is worth $\frac{1}{2}$ dollars or, simply, £1 equals \$2.

Gold flows Now, ignoring costs of packing, insuring, and shipping gold between countries, under the gold standard the rate of exchange would not vary from this \$2-for-£1 rate. No one in the United States would pay more than \$2 for £1, because you could always buy 50 grains of gold for \$2 in the United States, ship it to Britain, and sell it for £1. Nor would the English pay more than £1 for \$2. Why should they, when they could buy 50 grains of gold in England for £1, send it to the United States, and sell it for \$2?

In practice the costs of packing, insuring, and shipping gold must be taken into account. But these costs would only amount to a few cents per 50 grains of gold. If these costs were 3 cents for 50 grains of gold, Americans wanting pounds would pay up to \$2.03 for a pound rather than buy and export 50 grains of gold to get the pound. Why? Because it would cost them \$2 for the 50 grains of gold plus 3 cents to send it to England to be exchanged for £1. This \$2.03 exchange rate, above which gold would begin to flow out of the United States, is called the gold export point.

Conversely, the exchange rate would fall to \$1.97 before gold would flow into the United States. The English, wanting dollars, would accept as little as \$1.97 in exchange for £1, because from the \$2 which they could get by buying 50 grains of gold in England and reselling it in the United States, 3 cents must be subtracted to pay shipping and related costs. This \$1.97 exchange rate, below which gold would flow into the United States, is called the gold import point.

Our conclusion is that under the gold standard the flow of gold between nations would result in exchange rates which for all practical purposes are fixed.

Domestic Macro Adjustments Figure 38-5 helps explain the kinds of adjustments the gold standard would entail. Here, initially the demand for and the supply of pounds are D and S respectively and the resulting intersection point at a coincides with the fixed exchange rate of \$2 = £1 which results from the “in gold” definitions of the pound and the dollar. Now suppose for some reason American preferences for British goods increase, shifting the demand for pounds curve

to D' . In Figure 38-5 there is now a shortage of pounds equal to ab , implying an American balance of payments deficit.

What will happen? Remember that the rules of the gold standard prohibit the exchange rate from moving from the fixed \$2 = £1 relationship; the rate can *not* move up to a new equilibrium of \$3 = £1 at point c as it would under freely floating rates. Instead, the exchange rate would rise by a few cents to the American gold export point at which gold would flow from the United States to Britain.

Recall that the gold standard requires participants to maintain a fixed relationship between their domestic money supplies and their quantities of gold. Therefore, the flow of gold from the United States to Britain would bring about a contraction of the money supply in America and an expansion of the money supply in Britain. Other things being equal, this will reduce aggregate demand and, therefore, lower real domestic output, employment, and the price level in the United States. Also, the reduced money supply will boost American interest rates.

The opposite occurs in Britain. The inflow of gold increases the money supply, causing aggregate demand, national income, employment, and the price level to all increase. The increased money supply will also lower interest rates in Britain.

In Figure 38-5 declining American incomes and prices will reduce our demand for British goods and services and therefore reduce the American demand for pounds. Lower relative interest rates in Britain will make it less attractive for Americans to invest there, also reducing the demand for pounds. For all these reasons the D' curve will shift to the left.

Similarly, higher incomes and prices in Britain will increase British demand for American goods and services and higher American interest rates will encourage the British to invest more in the United States. These developments all increase the supply of pounds available to Americans, shifting the S curve of Figure 38-5 to the right.

In short, domestic macroeconomic adjustments in America and Britain, triggered by the international flow of gold, will produce new demand and supply for pound curves which intersect at some point on the horizontal line between points a and b .

Note the critical difference in the adjustment mechanisms associated with freely floating exchange rates and the fixed rates of the gold standard. With floating rates the burden of the adjustment is on the exchange rate itself. In contrast, the gold standard in-

volves changes in the domestic money supplies of participating nations which in turn precipitate changes in price levels, real domestic output and employment, and interest rates.

Although the gold standard boasts the advantages of stable exchange rates and the automatic correction of balance of payments deficits and surpluses, its basic drawback is that nations must accept domestic adjustments in such distasteful forms as unemployment and falling incomes, on the one hand, or inflation, on the other. In using the gold standard nations must be willing to submit their domestic economies to painful macroeconomic adjustments. Under this system a nation's monetary policy would be determined largely by changes in the demand for and supply of foreign exchange. If the United States, for example, was already moving toward recession, the loss of gold under the gold standard would reduce its money supply and intensify the problem. Under the international gold standard nations would have to forgo independent monetary policies.

Demise The worldwide Great Depression of the 1930s signaled the end of the gold standard. As domestic outputs and employment plummeted worldwide, the restoration of prosperity became the primary goal of afflicted nations. Protectionist measures such as the United States' Smoot-Hawley Tariff were enacted as nations sought to increase net exports and stimulate their domestic economies. And each nation was fearful that its economic recovery would be aborted by a balance of payments deficit which would lead to an outflow of gold and consequent contractionary effects. Indeed, nations attempted to devalue their currencies in term of gold to make their exports more attractive and imports less attractive. These devaluations undermined a basic condition of the gold standard and the system broke down.

The Bretton Woods System

Not only did the Great Depression of the 1930s lead to the downfall of the gold standard, it also prompted erection of trade barriers which greatly impaired international trade. World War II was similarly disruptive to world trade and finance. Thus, as World War II drew to a close the world trading and monetary systems were in shambles.

To lay the groundwork for a new international monetary system, an international conference of Allied nations was held at Bretton Woods, New Hampshire, in

1944. Out of this conference evolved a commitment to an *adjustable-peg system* of exchange rates, sometimes called the **Bretton Woods system**. The new system sought to capture the advantages of the old gold standard (fixed exchange rates), while avoiding its disadvantages (painful domestic macroeconomic adjustments).

Furthermore, the conference created the **International Monetary Fund (IMF)** to make the new exchange rate system feasible and workable. This international monetary system, emphasizing relatively fixed exchange rates and managed through the IMF, prevailed with modifications until 1971. The IMF continues to play a basic role in international finance and in recent years has performed a major role in ameliorating debt problems of the less developed countries.

IMF and Pegged Exchange Rates Why did the Bretton Woods adjustable-peg system evolve? We have noted that during the depressed 1930s, various countries resorted to the practice of **devaluation**—devaluing³ their currencies to try to stimulate domestic employment. For example, if the United States was faced with growing unemployment, it might devalue the dollar by *increasing* the dollar price of pounds from \$2.50 for £1 to, say, \$3 for £1. This action would make American goods cheaper to the British and British goods dearer to Americans, increasing American exports and reducing American imports. The resulting increase in net exports, abetted by the multiplier effect, would stimulate output and employment in the United States.

But the problem is that every nation can play the devaluation game, and most gave it a whirl. The resulting rounds of competitive devaluations benefited no one; on the contrary, they actually contributed to further demoralization of world trade. Nations at Bretton Woods therefore agreed that the postwar monetary system must provide for overall exchange rate stability whereby disruptive currency devaluations could be avoided.

What was the adjustable-peg system of exchange rates like? First, as with the gold standard, each IMF member was obligated to define its monetary unit in terms of gold (or dollars), thereby establishing par

rates of exchange between its currency and the currencies of all other members. Each nation was further obligated to keep its exchange rate stable vis-à-vis any other currency.

But how was this obligation to be fulfilled? The answer, as we saw in our discussion of fixed exchange rates, is that governments must use international monetary reserves to intervene in foreign exchange markets. Assume, for example, that under the Bretton Woods system the dollar was “pegged” to the British pound at \$2 = £1. Now suppose in Figure 38-5 that the American demand for pounds temporarily increases from D to D' so that a shortage of pounds of ab arises at the pegged rate. How can the United States keep its pledge to maintain a \$2 = £1 rate when the new market or equilibrium rate would be at \$3 = £1? The United States could supply additional pounds in the exchange market, shifting the supply of pounds curve to the right so that it intersects D' at b and thereby maintains the \$2 = £1 rate of exchange.

Where would the United States obtain the needed pounds? Under the Bretton Woods system there were three main sources.

1 Reserves The United States might currently possess pounds in a “stabilization fund” as the result of the opposite exchange market condition existing in the past. That is, at some earlier time the United States government may have spent dollars to purchase surplus pounds which were threatening to reduce the \$2 = £1 exchange rate to, say, \$1 = £1.

2 Gold Sales The United States government might sell some of the gold it holds to Britain for pounds. The proceeds would then be offered in the exchange market to augment the supply of pounds.

3 IMF Borrowing The needed pounds might be borrowed from the IMF. Nations participating in the Bretton Woods system were required to make contributions to the IMF on the basis of the size of their national income, population, and volume of trade. Thus, if necessary, the United States could borrow pounds on a short-term basis from the IMF by supplying its own currency as collateral.

Fundamental Imbalances: Adjusting the Peg A fixed-rate system such as Bretton Woods functions well so long as a nation's payments deficits and surpluses occur more or less randomly and are approximately equal in size. If a nation's payments surplus last year

allows it to add a sufficient amount to its international monetary reserves to finance this year's payments deficit, no problems will arise. But what if the United States, for example, encountered a “fundamental imbalance” in its international trade and finance and was confronted with persistent and sizable payments deficits? In this case it is evident that the United States would eventually run out of reserves and be unable to maintain its fixed exchange rate.

Under the Bretton Woods system, a fundamental payments deficit was corrected by devaluation, that is, by an “orderly” reduction in the nation's pegged exchange rate. Also, the IMF allowed each member nation to alter the value of its currency by 10 percent without explicit permission from the Fund to correct a deeply rooted or “fundamental” balance of payments deficit. Larger exchange rate changes required the sanction of the Fund's board of directors. By requiring approval of significant rate changes, the Fund guarded against arbitrary and competitive currency devaluation prompted by nations seeking a temporary stimulus to their domestic economies. In our illustration, devaluing the dollar would increase American exports and lower American imports, correcting its persistent payments deficits.

The objective of the adjustable-peg system was to realize a world monetary system which embraced the best features of both a fixed exchange rate system (such as the old international gold standard) and a system of freely fluctuating exchange rates. By reducing risk and uncertainty, short-term exchange rate stability—pegged exchange rates—would presumably stimulate trade and lead to the efficient use of world resources. Periodic exchange rate adjustments—adjustments of the pegs—made in an orderly fashion through the IMF, and on the basis of permanent or long-run changes in a country's payments position, provided a mechanism by which persistent international payments imbalances could be resolved by means other than painful changes in domestic levels of output and prices.

Demise of the Bretton Woods System Under the Bretton Woods system gold and the dollar came to be accepted as international reserves. The acceptability of gold as an international medium of exchange was derived from its role under the international gold standard of an earlier era. The dollar became acceptable as international money for two reasons.

1 The United States emerged from World War II as the free world's strongest economy.

2 The United States had accumulated large quantities of gold and between 1934 and 1971 maintained a policy of buying gold from, and selling gold to, foreign monetary authorities at a fixed price of \$35 per ounce. Thus the dollar was convertible into gold on demand; the dollar came to be regarded as a substitute for gold and therefore “as good as gold.”

But the role of the dollar as a component of international monetary reserves contained the seeds of a dilemma. Consider the situation as it developed in the 1950s and 1960s. The problem with gold as international money was a quantitative one. The growth of the world's money stock depends on the amount of newly mined gold, less any amounts hoarded for speculative purposes or used for industrial and artistic purposes. Unfortunately, the growth of the gold stock lagged behind the rapidly expanding volume of international trade and finance. Thus the dollar came to occupy an increasingly important role as an international monetary reserve.

Economies of the world acquire dollars as reserves as the result of United States' balance of payments deficits. With the exception of some three or four years, the United States incurred persistent payments deficits throughout the 1950s and 1960s. These deficits were financed in part by drawing down American gold reserves. But for the most part United States' deficits were financed by growing foreign holdings of American dollars which were “as good as gold” until 1971.

As the amount of dollars held by foreigners soared and as our gold reserves dwindled, other nations inevitably began to question whether the dollar was really “as good as gold.” The ability of the United States to maintain the convertibility of the dollar into gold became increasingly doubtful, and, therefore, so did the role of the dollar as generally accepted international monetary reserves. Hence, the dilemma: “. . . to preserve the status of the dollar as a reserve medium, the payments deficit of the United States had to be eliminated; but elimination of the deficit would mean a drying up of the source of additional dollar reserves for the system.”⁴ The United States had to reduce or eliminate its payments deficits to preserve the dollar's status as an international medium of exchange. But success in this endeavor would limit the expansion of international reserves or liquidity and restrict the growth of international trade and finance.

⁴Delbert A. Snider, *Introduction to International Economics*, 7th ed. (Homewood, Ill.: Richard D. Irwin, Inc., 1979), p. 352.

³A note on terminology is in order. We noted earlier in this chapter that the dollar has *appreciated* (depreciated) when its international value has increased (decreased) as the result of changes in the demand for, or supply of, dollars in foreign exchange markets. The terms *revalue* and *devalue* are used to describe an increase or decrease, respectively, in the international value of a currency which occurs as the result of governmental action.

This problem came to a head in the early 1970s. Faced with persistent and growing United States' payments deficits, President Nixon suspended the dollar's convertibility into gold on August 15, 1971. This suspension abrogated the policy to exchange gold for dollars at \$35 per ounce, which had existed for thirty-seven years. This new policy severed the link between gold and the international value of the dollar, thereby "floating" the dollar and allowing its value to be determined by market forces. The floating of the dollar withdrew American support from the old Bretton Woods system of fixed exchange rates and sounded the death knell for that system.

The Managed Float

The system of exchange rates which has since evolved is not easily described; it can probably best be labeled a system of managed floating exchange rates. It is recognized that changing economic conditions among nations require continuing changes in exchange rates to avoid persistent payments deficits or surpluses; exchange rates must be allowed to float. But short-term changes in exchange rates—perhaps accentuated by purchases and sales by speculators—disrupt and discourage the flow of trade and finance. Thus, it is generally agreed that the central banks of the various nations should buy and sell foreign exchange to smooth out such fluctuations in rates. That is, central banks should "manage" or stabilize short-term speculative variations in their exchange rates.

These characteristics were formalized by a leading group of IMF nations in 1976. Thus, ideally, the managed floating system will have not only the needed long-term exchange rate flexibility to correct fundamental payments imbalances, but also sufficient short-term stability of rates to sustain and encourage international trade and finance.

Actually, the current exchange rate system is more complicated than the previous paragraphs suggest. While the major currencies—German marks, American and Canadian dollars, Japanese yen, and the British pound—fluctuate or float in response to changing demand and supply conditions, most of the European Common Market nations are attempting to peg their currencies to one another. Furthermore, many less developed nations peg their currencies to the dollar and allow their currencies to fluctuate with it. Finally, some nations peg the value of their currencies to a "basket" or group of other currencies.

How well has the managed floating system worked? It has both proponents and critics.

Pros Proponents argue that the system has functioned well—far better than anticipated—during its relatively brief existence.

1 Trade Growth In the first place, fluctuating exchange rates did not lead to the diminution of world trade and finance that skeptics had predicted. In real terms world trade has grown at approximately the same rate under the managed float as it did during the decade of the 1960s under the fixed exchange rates of the Bretton Woods system.

2 Managing Turbulence Proponents argue that the managed float has weathered severe economic turbulence which might well have caused a fixed exchange regime to have broken down. Such dramatic events as worldwide agricultural shortfalls in 1972–1974, extraordinary oil-price increases in 1973–1974 and again in 1979–1980, worldwide stagflation in 1974–1976 and 1981–1983, and large U.S. budget deficits in the 1980s, all generated substantial international trade and financial imbalances. Flexible rates facilitated international adjustments to these developments, whereas the same events would have put unbearable pressures on a fixed-rate system.

Cons But there is still considerable sentiment in favor of a system characterized by greater exchange rate stability. Those favoring stable rates see problems with the current system.

1 Volatility and Adjustment Critics argue that exchange rates have been excessively volatile under the managed float. This volatility, it is argued, has occurred even when underlying economic and financial conditions of particular nations have been stable. Perhaps more importantly, the managed float has not readily resolved balance of payments imbalances as flexible rates are presumably capable of doing. Thus the United States has run persistent trade deficits in recent years, while Germany and Japan have had persistent surpluses. Changes in the international values of the dollar, mark, and yen have not yet corrected these imbalances.

2 A "Nonsystem"? Skeptics feel that the managed float is basically a "nonsystem"; the rules and guidelines circumscribing the behavior of each nation as to its exchange rate are not sufficiently clear or constraining to make the system viable in the long run. Nations will inevitably be tempted to intervene in foreign exchange markets, not merely to smooth out short-term

or speculative fluctuations in the value of their currencies, but to prop up their currency if it is chronically weak or to manipulate the value of their currency to achieve domestic stabilization goals. In brief, there is fear that in time there may be more "managing" and less "floating" of exchange rates, and this may be fatal to the present loosely defined system.

An example of more "managing" and less "floating" of exchange rates occurred in February 1987 when the "Group of Seven" industrial nations (G-7 nations)—the United States, West Germany, Japan, Britain, France, Italy, and Canada—agreed to take actions to stabilize the value of the dollar. In the previous two years the dollar had declined rapidly because of a sizable U.S. trade deficit. Although the U.S. trade deficit remained large, it was felt that a further depreciation of the dollar might be disruptive to economic growth in several G-7 economies. The G-7 nations thus bought large quantities of dollars to prop up the dollar's value. Since 1987 the G-7 nations have periodically intervened in foreign exchange markets to help stabilize the value of the dollar. Do these actions represent an admission by the industrial economies that the system of flexible exchange rates is seriously flawed?

The jury is still out on the managed float and no clear assessment has been reached: "Flexible rates have neither attained their proponents' wildest hopes nor confirmed their opponents' worst fears. But they have seen the major industrial economies through [two decades] mined with major disturbances to the international economy."⁵

QUICK REVIEW 38-3

Under the gold standard (1889–1934), nations fixed exchange rates by valuing their currencies in terms of gold, by tying their stocks of money to gold, and by allowing gold to flow between nations when balance of payment deficits and surpluses occurred.

The Bretton Woods, or adjustable-peg, system of exchange rates (1944–1971) fixed or pegged short-run exchange rates, but permitted orderly long-run adjustments of the pegs.

The managed floating system of exchange rate (1971–present) relies on foreign exchange markets to establish equilibrium exchange rates, but permits central banks to buy and sell foreign currencies to manage or stabilize short-term speculative changes in exchange rates.

⁵Richard E. Caves and Ronald W. Jones, *World Trade and Payments*, 3d ed. (Boston: Little, Brown and Company, 1981), p. 471.

RECENT UNITED STATES' TRADE DEFICITS

As shown in Figure 38-6, the United States had large trade deficits in the 1980s and early 1990s. Specifically, our merchandise trade deficit jumped from \$25 billion in 1980 to \$180 billion in 1987, then fell to \$74 billion in 1991. In 1980 the United States had a current account surplus of \$2 billion; by 1987 this had changed to a \$160 billion deficit. By 1991 the current account deficit had narrowed to \$92 billion.

What caused these large trade deficits? What were their effects? Why have they recently diminished?

Causes of the Trade Deficits

It is generally agreed that three major factors contributed to the large trade deficits of the 1980s and early 1990s.

The Rise of the Dollar As Figure 38-7 indicates, there was a pronounced rise in the international value of the dollar between 1980 and 1985. Here the value of the dollar is compared to ten other major currencies (weighted by the amount of trade we carry on with each country). By the end of 1984 the dollar was about 65 percent above its 1980 average value and at the highest level since floating exchange rates were adopted in the early 1970s. A strong or appreciated dollar means that foreign monies are cheaper to Americans and, conversely, dollars are more expensive to foreigners. As a result, foreign goods are cheap to Americans and our imports rise. Conversely, American goods are expensive to foreigners and our exports fall.

But why did the value of the dollar surge between 1980 and 1985? The basic answer is that real interest rates in the United States—nominal interest rates less the rate of inflation—rose in the United States compared to foreign countries. High real interest rates made the United States a very attractive place for foreigners to invest. As a result, the demand for dollars to make such investments increased, causing the dollar to appreciate in value.

Real interest rates were relatively high in America for two reasons.

1 The large Federal budget deficits of the 1980s are cited by many economists as a basic cause of high interest rates. Simply put, government borrowing to finance its deficits increased the domestic demand for money and boosted interest rates.

2 In 1979 the United States shifted to a tighter money policy in its efforts to control inflation. This

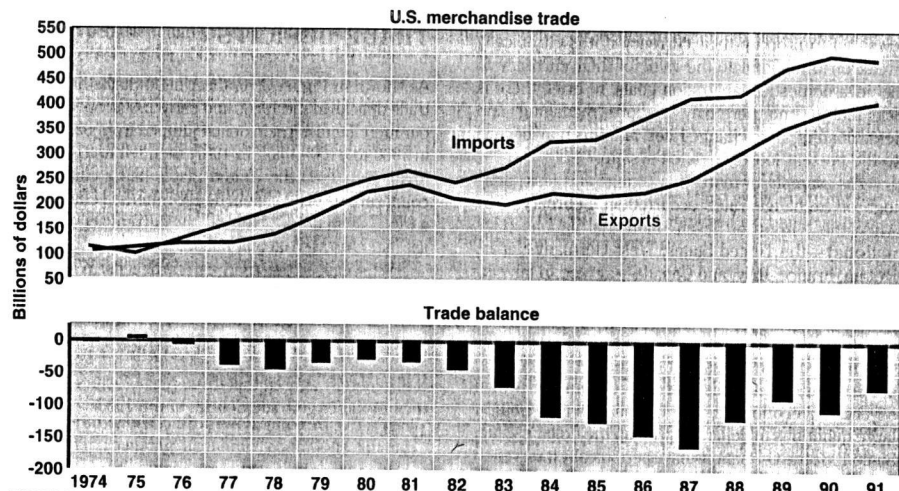


FIGURE 38-6 United States merchandise exports and imports and the trade balance

In recent years American trade deficits have been persistently large.

action increased interest rates directly by reducing the supply of money relative to its demand. Indirectly the lower rate of inflation kept the demand of foreign investors for dollars high because lower inflation means a higher *real* rate of return on investments in the United States.

By 1985 the value of the dollar had reached record heights relative to other currencies. Two factors then began to interact to reduce the dollar's value sharply over the next two years.

1 Five industrial nations—the United States, West Germany, Great Britain, France, and Japan—collectively decided to nudge the dollar downward to help correct the massive U.S. trade deficit and the trade surpluses in Japan and other nations. These five nations agreed to increase the supply of dollars in foreign exchange markets to reduce the dollar's value.

2 The demand for foreign currency in the United States rose sharply because more foreign money was needed to pay for the expanding volume of imports. This increase in the demand for yen, francs, and other foreign currencies increased the value of these currencies relative to the dollar. As shown in Figure 38-7, the value of the dollar declined sharply relative to other currencies over the 1985–1987 period.

Despite the sharp decline in the dollar between 1985 and 1987, the American trade imbalance stubbornly persisted. The major reason was that Japanese and other foreign importers did not immediately increase their dollar prices of products by as much as the decline in the international value of the dollar. Instead of increasing their prices, major importers accepted lower per unit profits on their goods. Therefore, imports to the United States for a time continued to rise, offsetting increases in American exports. Also, recall that in 1987 the G-7 nations agreed to halt the decline in the value of the dollar. Only in the second half of 1988 did the American trade deficit finally begin to shrink.

Rapid American Growth A second cause of the large trade deficits of the 1980s and 1990s is that the United States experienced a more rapid recovery from the 1980–1982 world recession than did its major trading partners. For example, American growth was about double that of Europe in 1983 and nearly triple the European rate in 1984. Although the gap in growth rates narrowed, the American growth rate continued to outpace the European rate between 1985 and 1990. This is significant because, like domestic consumption, a nation's purchases of foreign goods (its imports) vary di-

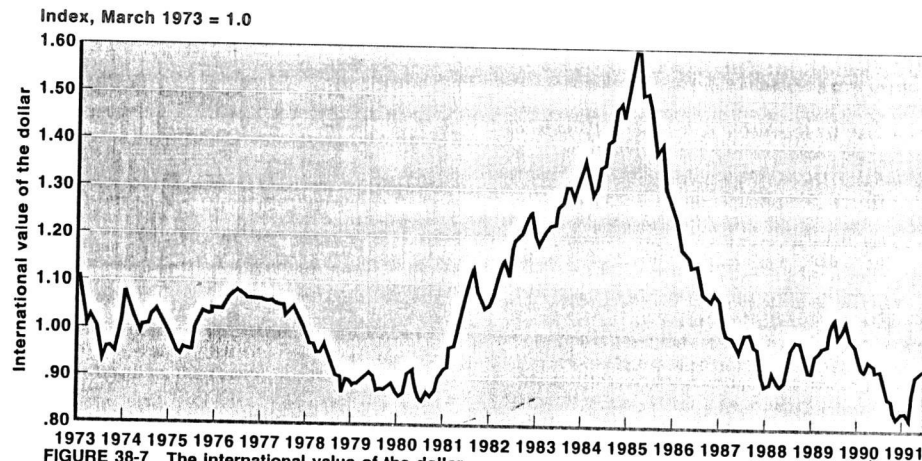


FIGURE 38-7 The international value of the dollar

Between 1980 and 1985 the value of the dollar increased greatly relative to other major currencies, tending to increase our imports and decrease our exports. The dollar fell sharply from 1985 through 1987 but trade deficits continued into the 1990s.

rectly with the level of domestic income. Because our national income expanded relatively rapidly, our imports also expanded rapidly. The slower growth of foreign national incomes meant their imports (our exports) grew slowly.

Exports to Less Developed Countries A third factor contributing to the large trade deficits was a falloff in our exports to the less developed countries (LDCs). An important source of the LDCs' external debt problem was their need to finance large international trade deficits by borrowing from the industrially advanced nations. As part of rescheduling and restructuring their debts in the 1980s, the less developed countries agreed to lessen their trade deficits. Thus they reduced their imports by using more restrictive monetary and fiscal policies to restrain the growth of their national incomes. In so doing their demands for imported goods declined. Part of those import reductions involved American goods, that is, United States exports. Many LDCs also *devalued* their currencies or, in other words, lowered the exchange rate value of their currencies by governmental decree. Devaluation restricted their imports and stimulated their exports. Thus the LDCs bought less from, and sold more to, the United States.

Effects of U.S. Trade Deficits

What have been the effects or consequences of our foreign trade deficits?

Dampened Aggregate Demand A trade deficit—more specifically, negative net exports—reduces aggregate demand and therefore, unless offset by other spending, diminishes the levels of real domestic output and employment via the multiplier effect. While this was a factor in keeping our level of employment below the full-employment rate for much of the 1980s, it also helped restrain inflation. A strong, appreciated dollar lowers the prices of all imported goods. Furthermore, a surging volume of imports exerts downward pressure on the prices of domestic goods that compete with those imports.

The constraining effect of a trade deficit is concentrated on industries which are highly dependent on export markets or are most competitive with imports. Some of the problems faced by American farmers, automobile manufacturers, and steel producers in the 1980s, for example, were related to the strong dollar and the associated trade deficits. These difficulties contributed greatly to the upsurge in political pressure for protectionist policies discussed in Chapter 37. They

LAST WORD

BUY AMERICAN: THE GLOBAL REFRIGERATOR

Humorist Art Buchwald pokes fun at those who suggest we could end our trade deficits by buying American consumer products.

"There is only one way the country is going to get on its feet," said Baleful.

"How's that?" I asked, as we drank coffee in his office at the Baleful Refrigerator Company.

"The consumer has to start buying American," he said, slamming his fist down on the desk. "Every time an American buys a foreign refrigerator it costs one of my people his job. And every time one of my people is out of work it means he or she can't buy refrigerators."

"It's a vicious circle," I said.

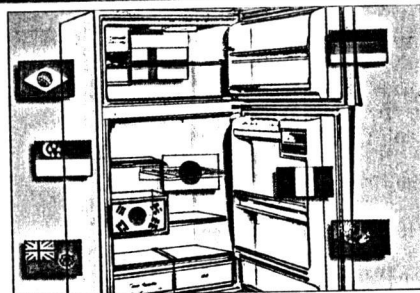
Baleful's secretary came in. "Mr. Thompson, the steel broker is on the phone."

My friend grabbed the receiver. "Thompson, where is that steel shipment from Japan that was supposed to be in last weekend? . . . I don't care about weather. We're almost out of steel, and I'll have to close down the refrigerator assembly line next week. If you can't deliver when you promise, I'll find myself another broker."

"You get your steel from Japan?" I asked Baleful.

"Even with shipping costs, their price is still lower than steel made in Europe. We used to get all our sheets from Belgium, but the Japanese are now giving them a run for their money."

The buzzer on the phone alerted Baleful. He listened for a few moments and then said, "Excuse me, I have a call from Taiwan. Mark Four? Look, R&D designed a new push-button door handle and we're going to send the specs to you. Tell Mr. Chow if his people send us a sample of one and can make it for us at the



same price as the old handle, we'll give his company the order."

A man came in with a plastic container and said, "Mr. Baleful, you said you wanted to see one of these before we ordered them. They are the containers for the ice maker in the refrigerator."

Baleful inspected it carefully and banged it on the floor a couple of times. "What's the price on it?"

"Hong Kong can deliver it at \$2 a tray, and Dong-Fu Plastics in South Korea said they can make it for \$1.70."

"It's just a plastic tray. Take the South Korea bid. We'll let Hong Kong supply us with the shelves for the freezer. Any word on the motors?"

"There's a German company in Brazil that just came out with a new motor, and it's passed all our tests, so Johnson has ordered 50,000."

"Call Cleveland Motors and tell them we're sorry, but the price they quoted us was just too high."

"Yes, sir," the man said and departed.

occurring, it is clearly beneficial to American consumers. After all, a trade deficit means that Americans are currently receiving more goods and services as imports from the rest of the world than we are sending to the rest of the world as exports. Trade deficits augment our domestic living standards during the period in which they occur.

A related consequence of our recent trade deficits is that in 1985 the United States' status changed from that of a net creditor to that of a *net debtor* for the first time since 1914. That is, the United States now owes foreigners more than they owe this country. Recall that current account deficits are financed primarily by net

The secretary came in again and said, "Harry telephoned and wanted to let you know the defroster just arrived from Finland. They're unloading the box cars now."

"Good. Any word on the wooden crates from Singapore?"

"They're at the dock in Hoboken."

"Thank heaven. Cancel the order from Boise Cascade."

"What excuse should I give them?"

"Tell them we made a mistake in our inventory, or we're switching to plastic. I don't care what you tell them."

Baleful turned to me. "Where were we?"

"You were saying that if the consumer doesn't start buying American, this country is going to be in a lot of trouble."

"Right. It's not only his patriotic duty, but his livelihood that's at stake. I'm going to Washington next week to tell the Senate Commerce Committee that if they don't get off the stick, there isn't going to be a domestic refrigerator left in this country. We're not going to stay in business for our health."

"Pour it to them," I urged him.

Baleful said, "Come out with me into the showroom."

I followed him. He went to his latest model, and opened the door. "This is an American refrigerator made by the American worker, for the American consumer. What do you have to say to that?"

"It's beautiful," I said. "It puts foreign imports to shame."

Source: Art Buchwald, "Being Bullish on Buying American." Reprinted by permission. We discovered this article in *Master Curriculum Guide in Economics: Teaching Strategies for International Trade* (New York: Joint Council on Economic Education, 1988).

capital inflows to the United States. When our exports are insufficient to pay for our imports, we finance the difference by borrowing from foreigners or, in other words, by going into debt. The financing of our recent large trade deficits has caused foreigners to accumulate a larger volume of claims against American assets than we have accumulated against foreign assets. The U.S. foreign debt burden climbed to \$721 billion in 1990, making us the largest debtor nation in the world.

One implication of net debtor status is that we can no longer look forward to a net inflow of dividend and interest payments (see item 7 in Table 38-1's balance of payments) to help cover deficits in our merchandise

and services trade. A second implication is that more of our corporations are foreign-owned.

The above comments on the economic effects of our trade deficit for the United States economy can be reversed as far as our industrialized trading partners are concerned. The current accounts of Japan and Germany, for example, tended to move toward surplus. These countries experienced an expansionary-inflationary stimulus and unusual growth in their export-dependent industries. They also increased their holdings of American debt.

QUICK REVIEW 38-4

In the 1980s and early 1990s the United States experienced large trade deficits, caused by a strong dollar, relatively rapid American growth prior to the 1990-1991 recession, and reduced purchases of our exports by less developed nations.

These large deficits had a contractionary, anti-inflationary impact, hurt export-dependent industries, and resulted in the United States becoming a debtor nation; they also temporarily enhanced America's standard of living.

Reducing the Trade Deficit

Two kinds of policies for reducing large trade deficits are most often cited: reduction of the Federal budget deficit and measures to accelerate economic growth abroad.

Reduction of the Budget Deficit Many economists agree that the most critical cause of our continuing trade deficits has been our large annual Federal budget deficits. It is argued that a reduction in the size of our Federal budget deficit will lower the real interest rate in the United States compared to other nations. In other words, a reduction in the government's demand for funds to finance its deficits will lower domestic interest rates and thus make financial investments in the United States less attractive to foreigners. The demand for dollars by foreigners will decline and the dollar will depreciate. Given a depreciated dollar, our exports will increase and imports will fall, correcting our trade deficit. This scenario is exactly opposite to the one shown earlier in Figure 18-2.

Would not a "managed" depreciation of the dollar by the G-7 nations produce a decline in the U.S. trade deficit, even without a reduction in our budget deficit? Perhaps so, but this point may be moot. Our trading partners have not been interested in allowing the dollar

also generated interest in industrial policies designed to provide special help for allegedly "key" industries deemed critical to American industrial preeminence.

Increased American Indebtedness A trade deficit is also considered "unfavorable" because it must be financed by increased American indebtedness to foreigners. A trade deficit means we must borrow from the rest of the world to finance that deficit. This failure to "pay our way" in international trade is usually interpreted as a sign of domestic economic weakness and, hence, undesirable. However, economists point out that, at the time a trade or current account deficit is

to fall appreciably below its 1988 level, unless we reduce our budget deficit. In effect, these nations contend that the United States must "get its fiscal house in order" to achieve a better balance of international trade.

Economic Growth Abroad The American trade deficit can also be reduced if nations abroad speed up their rates of economic growth. Higher levels of foreign national income increase the demand for American exports. The G-7 group of industrial nations has recognized the importance of economic growth in the nations which have trade surpluses as a way to reduce these surpluses and lower the American trade deficit. In the late 1980s, the governments of Japan and Germany established expansionary fiscal and monetary policies to bolster national income and increase the demand for goods produced in America.

Other "Remedies" There are several other possible "remedies" to the persistent United States' trade deficits.

Easy Money Policy Under appropriate circumstances, an easy money policy lowers real interest rates and reduces a trade deficit. The process works as follows. The decline in interest rates reduces the international demand for dollars, which results in a depreciation of the dollar. Dollar depreciation raises our exports and lowers our imports (Table 15-4).

Protective Tariffs Protective tariffs can be used to reduce imports, but this strategy results in the loss of the gains from specialization and international trade. Furthermore, it may not be successful: Tariffs which reduce our imports foster retaliatory tariffs abroad which reduce our exports. Trade deficits do not disappear in this circumstance; instead, all trading partners suffer declines in their living standards.

Recession Recessions in the United States reduce disposable income and thus spending on all goods, in-

cluding imports. Because exports are largely unaffected, the decline in imports trims the trade deficit. This is precisely what happened in the United States during the recession of 1990-1991. But recession is an undesirable way to reduce trade deficits; it imposes higher economic costs (lost output) on society than the costs associated with the trade deficit itself. Also, unless the fundamental causes of the deficits have in the meanwhile been remedied, imports and thus trade deficits again rise when the economy begins to recover from recession.

Increased American Competitiveness The American trade deficits can be reduced by lowering the costs of, and improving the quality of, American goods and services relative to foreign goods. Cost-saving production technologies, development of improved products, and more efficient management techniques each can contribute to a decline in the trade deficit by lowering United States demand for imported goods and increasing foreign demand for American goods.

Direct Foreign Investment Ironically, our persistent trade deficit has set off a chain of events which has begun to feed back to reduce the trade deficit itself. The vast accumulation of American dollars in foreign hands has enabled foreign individuals and firms to buy American factories or to build new plants in the United States. Furthermore, the fall in the value of the dollar has provided an incentive for foreign firms to produce in the United States rather than in their own nations.

In short, the trade deficit has given rise to an increase in *direct foreign investment* in the form of plant and equipment. Foreign-owned factories are beginning to turn out increasing volumes of goods that otherwise would have been imported. Hondas and Mazdas, produced in American factories, have replaced Hondas and Mazdas formerly imported from Japan. Other examples abound. The upshot is that the American trade deficit may shrink as imports are replaced with goods produced in foreign-owned factories in the United States.

CHAPTER SUMMARY

1 American exports create a foreign demand for dollars and make a supply of foreign exchange available to Americans. Conversely, American exports simultaneously create a demand for foreign exchange and make a supply of dollars available to foreigners. Generally, a nation's exports earn the foreign currencies needed to pay for its imports.

2 The balance of payments records all international trade and financial transactions taking place between a given nation and the rest of the world. The trade balance compares merchandise exports and imports. The balance on goods and services compares exports and imports of both goods and services. The current account balance considers not

only goods and services transactions, but also net investment income and net transfers.

3 A deficit on the current account will be largely offset by a surplus on the capital account. Conversely, a surplus on the current account will be largely offset by a deficit on the capital account. A balance of payments deficit occurs when the sum of the current and capital accounts is in deficit. A payments deficit is financed by drawing down official reserves. A balance of payments surplus occurs when the sum of the current and capital accounts is in surplus. A payments surplus results in an increase in official reserves. The desirability of a balance of payments deficit or surplus depends on its causes and its persistence over time.

4 Flexible or floating exchange rates are determined by the demand for and supply of foreign currencies. Under floating rates a currency will depreciate or appreciate as a result of changes in tastes, relative income changes, relative price changes, relative changes in real interest rates, and speculation.

5 Maintenance of fixed exchange rates requires adequate reserves to accommodate periodic payments deficits. If reserves are inadequate, nations must invoke protectionist trade policies, engage in exchange controls, or endure undesirable domestic macroeconomic adjustments.

6 Historically, the gold standard provided exchange rate stability until its disintegration during the 1930s. Under this system, gold flows between nations precipitated sometimes painful changes in price, income, and employment levels in bringing about international equilibrium.

7 Under the Bretton Woods system exchange rates were pegged to one another and were stable. Participating nations were obligated to maintain these rates by using stabilization funds, gold, or borrowings from the IMF. Persistent or "fundamental" payments deficits could be resolved by IMF-sanctioned currency devaluations.

8 Since 1971 a system of managed floating exchange rates has been in use. Rates are generally set by market forces, although governments intervene with varying frequency to alter their exchange rates.

9 Between 1980 and 1991 the United States experienced large international trade deficits. Causes include a rapidly appreciating dollar between 1980 and 1985; b relatively rapid expansion of the American economy prior to the recession of 1990-1991; and c curtailed purchases of our exports by the less developed countries.

10 The effects of large trade deficits have been manifold. They have had a contractionary, anti-inflationary effect on our domestic economy. American export-dependent industries have experienced declines in output, employment, and profits, thereby generating political pressures for protection. The United States has become the world's largest debtor nation. However, the trade deficit has meant a current increase in the living standards of American consumers.

11 Two solutions to the trade deficit are a reduction of the budget deficit and b faster economic growth abroad. Other "remedies" are an easy money policy, protective tariffs, recession, improved U.S. competitiveness, and direct foreign investment.

TERMS AND CONCEPTS

balance of payments	capital account	flexible or floating	points
current account	balance on the capital	exchange rates	Bretton Woods system
credits	account	depreciation and	International Monetary
debits	official reserves	appreciation	Fund
trade balance	balance of payments	purchasing power	devaluation
balance on goods and	deficits and	parity	managed floating
services	surpluses	gold standard	exchange rates
balance on current	fixed exchange rates	gold import and export	G-7 nations
account			

QUESTIONS AND STUDY SUGGESTIONS

1 Explain how an American automobile importer might finance a shipment of Toyotas from Japan. Demonstrate how an American export of machinery to Italy might be financed. Explain: "American exports earn supplies of foreign monies which Americans can use to finance imports."

2 "A rise in the dollar price of yen necessarily means a fall in the yen price of dollars." Do you agree? Illustrate and elaborate: "The critical thing about exchange rates is that

they provide a direct link between the prices of goods and services produced in all trading nations of the world." Explain the purchasing power parity theory of exchange rates.

3 The Swedish auto company Saab imports car components from Germany and exports autos to the United States. In 1990 the dollar depreciated, and the German mark appreciated, relative to the Swedish krona. Speculate as to how this hurt Saab—twice.

weakening their incentive to produce, and is typically not conducive to capital improvements. At the other extreme is the absurd arrangement whereby each family owns and farms a minute fragment of land far too small for the application of modern agricultural technology. An important complication to the problem of land reform lies in the fact that political considerations sometimes push reform in that direction which is least defensible on economic grounds. For many nations, land reform may well be the most acute institutional problem to be resolved in initiating the process of economic development.

Examples: Land reform in South Korea undermined the political control of the landed aristocracy and made way for the development of strong commercial and industrial middle classes, all to the benefit of the country's economic development. In contrast, the prolonged dominance of the landed aristocracy in the Philippines has helped stifle the development of that economy.⁸

QUICK REVIEW 39-1

- 1 About three-fourths of the world's population lives in the LDCs of Africa, Asia, and Latin America.
- 2 Natural resource scarcities and inhospitable climates restrict growth in many LDCs.
- 3 The LDCs are characterized by overpopulation, high unemployment rates, underemployment, and low labor productivity.
- 4 Low saving rates, capital flight, weak infrastructures, and lack of investors impair capital accumulation.
- 5 Sociocultural and institutional factors are often serious impediments to growth.

THE VICIOUS CIRCLE: A SUMMING UP

Many of the characteristics of LDCs just described are simultaneously causes and consequences of their poverty. These countries are caught in a vicious circle of poverty. They stay poor because they are poor! Consider Figure 39-1. The fundamental feature of an LDC is low per capita income. Being poor, a family has little ability or incentive to save. Furthermore, low incomes mean low levels of demand. Thus, there are few avail-

⁸Mrinal Datta-Chaudhuri, "Market Failure and Government Failure," *Journal of Economic Perspectives*, Summer, 1990, p. 36

able resources, on the one hand, and no strong incentives, on the other, for investment in physical or human capital. This means labor productivity is low. And, since output per person is real income per person, it follows that per capita income is low.

Many experts feel that the key to breaking out of this vicious circle is to increase the rate of capital accumulation, to achieve a level of investment of, say, 10 percent of the national income. But Figure 39-1 reminds us that the real villain for many LDCs—rapid population growth—may be waiting in the wings to undo the potentially beneficial effects of this higher rate of capital accumulation. For example, using hypothetical figures, suppose that initially an LDC is realizing no growth in its real GDP. But now it somehow manages to increase its saving and investment to 10 percent of its GDP. As a result, its real GDP begins to grow at, say, 2.5 percent per year. Given a stable population, real GDP per capita will also grow at 2.5 percent per year. If this persists, the standard of living will double in about 28 years. But what if population grows at the Latin American rate of 2.5 percent per year? Then real income per person is unchanged and the vicious circle persists.

More optimistically, if population can be kept constant or constrained to some growth rate significantly below 2.5 percent, then real income per person will rise. This implies the possibility of still further enlargement in the flows of saving and investment, continued advances in productivity, and the continued growth of per capita real income. In short, if a process of self-sustaining expansion of income, saving, investment, and productivity can be achieved, the self-perpetuating vicious circle of poverty can be transformed into a self-regenerating, beneficent circle of economic progress. The trick is to make effective those policies and strategies which will accomplish this transition.

ROLE OF GOVERNMENT

Economists do not agree on the appropriate role of government in seeking economic growth.

A Positive Role

One view is that, at least during initial stages of development, government should play a major role. The reasons for this stem in large part from the character of the obstacles facing LDCs.

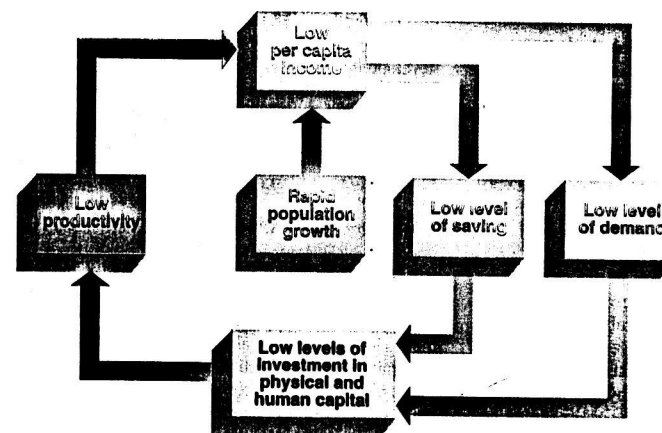


FIGURE 39-1 The vicious circle of poverty

Low per capita incomes make it extremely difficult for poor nations to save and invest, a condition that perpetuates low productivity and low incomes. Furthermore, rapid population growth may quickly absorb increases in per capita real income and thereby may negate the possibility of breaking out of the poverty circle.

1 Law and Order Some of the poorest countries are plagued by widespread banditry and intertribal warfare which divert both attention and resources from the task of development. A strong and stable national government is needed to establish domestic law and order and to achieve peace and unity.

2 Lack of Entrepreneurship The absence of a sizable and vigorous entrepreneurial class, ready and willing to accumulate capital and initiate production, indicates that in many cases private enterprise is intrinsically not capable of spearheading the growth process.

3 Infrastructure Many obstacles to economic growth center on deficiencies of public goods and services, or, in other words, an inadequate infrastructure. Sanitation and basic medical programs, education, irrigation and soil conservation projects, and construction of highways and transportation-communication facilities are all essentially nonmarketable goods and services yielding widespread spillover benefits. Government is the sole institution in a position to provide these goods and services in required quantities.

4 Forced Saving and Investment Government action may also be required to break through the saving-investment dilemma which impedes capital formation in LDCs.

It may well be that only governmental fiscal action can provide a solution by forcing the economy to accumulate capital. There are two alternatives. One is to force the economy to save by increasing taxes. These tax revenues can then be channeled into priority investment projects. The problems of honestly and efficiently administering the tax system and achieving a relatively high degree of compliance with tax laws are frequently very great.

The other alternative is to force the economy to save through inflation. Government can finance capital accumulation by creating and spending new money or by selling bonds to banks and spending the proceeds. The resulting inflation is the equivalent of an arbitrary tax on the economy.

There are serious arguments against the advisability of saving through inflation. In the first place, inflation tends to distort the composition of investment away from productive facilities to such items as luxury

LAST WORD

FAMINE IN AFRICA

The roots of Africa's persistent famines include both natural and human causes.

Famine in Africa is not uncommon. The world was shocked during the 1984–1985 Ethiopian famine by pictures of fly-tormented, emaciated children with bloated bellies. Despite an outpouring of aid from the rich nations, that famine caused 1 million deaths. A number of African nations—including Ethiopia, Sudan, Angola, Liberia, Mozambique, and Malawi—are persistently threatened by famine. Various estimates put from 5 to 20 million Africans at risk. This tragic situation is especially ironic because most African countries were self-sufficient in food at the time they became independent nations; they are now heavily dependent on imported foodstuffs for survival.

The immediate and much-publicized cause of this catastrophe is drought. But the ultimate causes of Africa's declining ability to feed itself are more complex and multifaceted, an interplay of natural and human conditions. In addition to a lack of rainfall, these include chronic civil strife, rapid population growth, widespread soil erosion, and counterproductive public policies.

1 Civil Strife Regional rebellions and prolonged civil wars have devastated some African nations. Both Ethiopia and the Sudan, for example, have been plagued by decades of civil strife. Not only do these conflicts divert precious resources from civilian uses, but they greatly complicate the ability of wealthy nations to provide famine and developmental aid. Ethiopia's government has often denied food aid to areas occupied by antigovernment forces. Donated food is



frequently diverted to the army and denied to starving civilians.

2 Population Growth The hard fact is that in Africa population is growing more rapidly than is food production. Specifically, population is increasing at about 3 percent per year while food output is growing at only 2 percent per year. This grim arithmetic suggests de-

clining living standards, hunger, and malnutrition. The World Bank reports that during the 1980s the per capita incomes of the sub-Saharan nations fell to about three-quarters of the level reached by the end of the 1970s.

3 Ecological Degradation But apart from the simple numbers involved, population growth has apparently contributed to the ecological degradation of Africa. Given population pressures and the increasing need for food, marginal land has been deforested and put into crop production. In many cases trees which have served as a barrier to the encroachment of the desert have been cut for fuel, allowing the fragile topsoil to be blown away by desert winds. The ultimate scarcity of wood which has accompanied deforestation has forced the use of animal dung for fuel, thereby denying its traditional use as fertilizer. Furthermore, traditional fallow periods have been shortened, resulting in overplanting and overgrazing and, in simple terms, a wearing out of the soil. Deforestation and land overuse have reduced the capacity of the land to absorb moisture, diminishing its productivity and its ability to resist drought. Some authorities feel that the diminished ability of the land to absorb water reduces the amount of moisture which evaporates into the clouds to return ultimately as rainfall. All of this is complicated by the fact that there are few facilities for crop storage. Thus, even when crops are good, it is difficult to accumulate a surplus for future lean years. A large percentage of domestic farm output in some parts of Africa is lost to rats, insects, and spoilage.

4 Public Policies and Debt Ill-advised public policies have contributed to Africa's famines. In the first place, African governments have generally neglected investment in agriculture in favor of industrial develop-

ment and military strength. It is estimated that African governments on the average spend four times as much on armaments as they do on agriculture. Over 40 percent of Ethiopia's budget is for the support of an oppressive military. Second, many African governments have followed the policy of establishing the prices of agricultural commodities at low levels to provide cheap food for growing urban populations. This low-price policy has diminished the incentives of farmers to increase productivity. While foreign aid has helped to ease the effects of Africa's food-population problems, most experts reject aid as a long-term solution. Indeed, experience suggests that aid in the form of foodstuffs can only provide temporary relief and may undermine the realization of long-run local self-sufficiency. Foreign food aid, it is contended, merely treats symptoms and not causes.

All of this is made more complex by the fact that the sub-Saharan nations are burdened with relatively large and growing external debts. The IMF reports that the aggregate debt of these nations rose from \$21 billion in 1976 to \$127 billion in 1990. As a condition of further aid, these nations have had to invoke austerity programs which have contributed to declines in their per capita incomes. One tragic consequence is that many of these nations have cut back on social service programs for children. A UNICEF spokesman has indicated that 3 million children died in 1987 worldwide "because they didn't have 50 cents worth of vaccine in them."

To summarize: the famine confronting much of Africa is partly a phenomenon of nature and in part self-inflicted. Drought, overpopulation, ecological deterioration, and errant public policies have all been contributing factors. This complex of causes implies that hunger and malnutrition in Africa may persist long after the rains return.

CHAPTER SUMMARY

- Most of the nations of the world are less developed (low per capita income) nations. While some LDCs have been achieving quite rapid growth rates in recent years, others have realized little or no growth at all.
- Initial scarcities of natural resources and the limited possibility of augmenting existing supplies may impose a serious limitation on a nation's capacity to develop.
- The presence of large and rapidly growing populations in most LDCs contributes to low per capita incomes. In particular, increases in per capita incomes frequently induce

rapid population growth, to the end that per capita incomes again deteriorate to near subsistence levels.

4 Most LDCs suffer from both unemployment and underemployment. Labor productivity is low because of insufficient investment in physical and human capital.

5 In many LDCs both the saving and investment aspects of capital formation are impeded by formidable obstacles. In some of the poorest LDCs the savings potential is very low. Many LDC savers have chosen to transfer their funds to the IACs rather than invest domestically. The absence of a vig-

orous entrepreneurial class and the weakness of investment incentives are also serious impediments to capital accumulation.

6 Appropriate social and institutional changes and, in particular, the presence of "the will to develop" are essential ingredients in economic development.

7 The vicious circle of poverty brings together many of the obstacles to growth, saying in effect that "poor countries stay poor because of their poverty." Low incomes inhibit saving and accumulation of physical and human capital, making it difficult to increase productivity and incomes. Rapid population growth can offset otherwise promising attempts to break the vicious circle.

8 The nature of the obstacles to growth—the absence of an entrepreneurial class, the dearth of infrastructure, the saving-investment dilemma, and the presence of social-institutional obstacles to growth—suggests the need for government action in initiating the growth process. However, the corruption and maladministration which are quite common to the public sectors of the LDCs suggest that government may be relatively ineffective as an instigator of growth.

9 Advanced nations can assist in development by reducing trade barriers and by providing both public and private capital.

10 Rising energy prices, declining export prices, depreciation of the dollar, and concern about LDCs' creditworthiness

LAST WORD

OBITUARY: THE SOVIET UNION

Contradictions of the world's first communist state killed Marx's vision of a free and prosperous society.

The Soviet Union is dead. It was 74.

Marketed as utopia, run by slogan and fear, it blended genuine achievement with elaborate facade. It created a system whose top priority seemed to be concealing its own failings.

Foreigners, intimidated by its military prowess and obsessive secrecy, frequently overestimated its strength. Its own citizens, bombarded by buoyant propaganda as they went about their harsh existence, sometimes had no idea if their lives were really growing better or worse.

Irony and artifice were everywhere. The Soviet Union led the world in production of steel, oil, tractors and locomotives, all as it moved inexorably to economic ruin.

Construction crews competed to build whole apartment buildings in a month, a week, a day. Yet the Soviet dream of abolishing "communal apartments," where three or four families had to share a tiny kitchen and toilet, was never achieved.

The Soviets were a nuclear super-power that projected military and diplomatic strength around the world. Yet at home, medical care was poor, citizens' diets poorer, and shoddy goods were the norm.

And a nation that claimed to publish more books and newspapers than any other had a ruthless system of censorship, political control and suppression of free ideas.

What went wrong? Why did a nation rich in natural resources—with a literate, educated work force and one-sixth of the world's land mass to stretch out in—fail to build the vibrant, prosperous, free society that was supposed to be a beacon to the world?

Most Soviets blame, first of all, "scientific socialism," the shaky and untested economic model that Lenin's Communists forced on backward Russia.



It was imposed from above and preserved through brute force. There were huge economic advances at first, but accomplished through intimidation as much as economic logic. After World War II, the "planned economy" operated as a continent-wide shell game, with resources wastefully rushed here and there to maintain an illusion of economic progress.

Since the economy essentially did not work, the leaders who depended on it had to find other ways to preserve their strength and build national pride.

They created a genuine center of accomplishment in the Soviet military, which was denied no human or financial resource. The Soviets also excelled in areas of high technology, mathematics and space; brought electricity, communications and industry to backward zones of their nation; and provided an example of quick, forced economic development that many in the Third World admired and attempted to emulate.

For anyone who was unimpressed by these accomplishments, the Soviets also created a system of terror that silenced political dissidents, religious activists, nationalist agitators and anyone else whose cause might be more appealing than communism.

But perhaps most important of all, the leaders built their strength on a suffocating cradle-to-grave so-

hold. During this critical period the Soviet economy may further deteriorate. In addition, the abortive August 1991 coup indicates clearly that economic collapse can threaten to reverse the Soviet Union's substantial progress toward political democracy and the apparent demise of the cold war.

cial structure that made citizens totally beholden to the state—and then demanded practically nothing of them.

The state provided food, apartments, medicine, education, jobs and old-age pensions. It ran factories, department stores, farms, film studios and excellent orchestras and ballet companies. It may not have provided the weather, but it certainly controlled the news and sports.

This benevolence, such as it was, was practically free of charge. Citizens were judged mainly by political reliability—or political indifference, which also was acceptable.

During Leonid Brezhnev's reign, the situation steadily worsened and hypocrisy was elevated to the level of state policy. Bluster covered the increasing weakness.

Incompetence, sloppiness and corruption at work were widely overlooked. Citizens quickly learned that an effective way to get along was to do the minimum and challenge nothing—neither politics, nor the efficiency of their workplace.

Under such conditions, many believed the Soviet economy—and the country's whole spirit—was heading for a crash long before Mikhail Gorbachev came to power in 1985. He accelerated the slide by public openness about the country's troubles, a tactic that invigorated a few reformers but threw millions into despair and inactivity when facades came down and they realized how far their country had already crumbled.

Gorbachev's attempts at restructuring the economy were too tentative and too late. They could not make up for seven decades of an economic system that did not work, and the intricate structure established to maintain it at all cost.

That cost included stamping out personal initiative and putting political order ahead of everything else—including the welfare of the Soviet people.

Source: Thomas Kent, "Soviets Mixed Achievement, Facade as Dreams Failed," *The Lincoln Star*, Lincoln, Nebraska, December 26, 1991. Reprinted by permission.

The United States and the other market democracies have a great economic stake in the former Soviet Union's transition to democracy and capitalism. If the transition fails, the peace dividend associated with the end of the cold war will not be realized and the possibility of accelerated economic growth through expanded

international trade with a free-market Commonwealth will also be sacrificed. The political benefit is that a democratic Commonwealth will isolate the last strongholds of communism—China, Cuba, North Korea, and Vietnam—and perhaps force their leaders toward political and economic reform.

But there are serious reservations concerning aid to the Soviet Union. One argument is that aid is likely to be ineffectual and wasteful until the transition to market capitalism has been accomplished. Aside from humanitarian aid in the form of foodstuffs and medicine, economic aid is not likely to be of much help under existing institutional arrangements.

A second contention is that the Soviet Union has not yet exploited the opportunity it now has to divert vast amounts of resources from the military to the civilian sector. Cutting military spending in half would release resources three or four times as large as the amount of aid Gorbachev requested.

Third, it is pointed out that the Soviet Union is in fact a gigantic \$2 trillion economy. Even granting the \$30 billion in aid would only amount to about \$100 per year per Soviet citizen.

Finally, there is the hard political fact that foreign aid for a long-time cold war foe may not be popular among the voters of industrialized nations who see their own countries troubled with problems of unemployment, poor education, poverty, and drug abuse.

The United States' position has been that at this time it is appropriate to provide food and other humanitarian aid along with technical and educational assistance to Soviet enterprises and political officials, but to withhold unrestricted aid until substantial reform of the economy has been achieved.

In fact, in the spring of 1992 the United States and its G-7 partners (Germany, Japan, France, Britain, Canada, and Italy) have promised Russia a \$24 billion aid package. This includes \$11 billion in direct aid; \$6 billion to provide a fund for stabilizing the ruble; \$4.5 billion of IMF and World Bank loans; and \$2.5 billion in debt rescheduling (on an estimated \$89 billion debt). The United States' share is approximately \$5 billion.

Private Investment

As the former Soviet Union attempts to move toward a capitalistic system, will it be able to attract foreign investment to shore up its economy? Given the vast potential market provided by some 288 million citizens,

Foreign Aid

In 1991 President Gorbachev appealed to the industrialized countries for \$20 to \$30 billion in aid. The argument for granting this aid was that it would ease the painful transition process when central planning was being dismantled and free enterprise had not yet taken