

In no case, therefore, on the free market can a “monopoly price” be conceptually distinguished from a “competitive price.” *All* prices on the free market are competitive.⁵⁷

4. Labor Unions

A. RESTRICTIONIST PRICING OF LABOR

It might be asserted that labor unions, in exacting higher wage rates on the free market, are achieving identifiable monopoly prices. For here two *identifiable* contrasting situations exist: (a) where individuals sell their labor themselves; and (b) where they are members of labor unions which bargain on their labor for

⁵⁷See Mises:

Prices are a market phenomenon. . . . They are the result-ant of a certain constellation of market data, of actions and reactions of the members of a market society. It is vain to meditate what prices would have been if some of their determinants had been different. . . . It is no less vain to ponder on what prices ought to be. Everybody is pleased if the prices of things he wants to buy drop and the prices of the things he wants to sell rise. . . . Any price determined on a market is the necessary outgrowth of the interplay of the forces operating, that is, demand and supply. Whatever the market situation which generated this price may be, with regard to it the price is always adequate, genuine, and real. It cannot be higher if no bidder ready to offer a higher price turns up, and it cannot be lower if no seller ready to deliver at a lower price turns up. Only the appearance of such people ready to buy or sell can alter prices. Economics . . . does not develop formulas which would enable anybody to compute a “correct” price different from that established on the market by the interaction of buyers and sellers. . . . This refers also to monopoly prices. . . . *No alleged “fact finding” and no arm-chair speculation can discover another price at which demand and supply would become equal.* The failure of all experiments to find a satisfactory solution for the limited-space monopoly of public utilities clearly proves this truth. (Mises, *Human Action*, pp. 392–94; italics added)

them. Furthermore, it is clear that while cartels, to be successful, must be economically more efficient in serving the consumer, no such justification can be found for unions. Since it is always the individual laborer who works, and since efficiency in organization comes from management hired for the task, forming unions *never* improves the productivity of an individual's work.

It is true that a union provides an identifiable situation. However, it is *not* true that a union wage rate could ever be called a monopoly price.⁵⁸ For the characteristic of the monopolist is precisely that he monopolizes a factor or commodity. To obtain a monopoly price, he sells only part of his supply and *withholds* selling the other part, because selling a lower quantity raises the price on an inelastic demand curve. It is the unique characteristic of labor in a free society, however, that it *cannot* be monopolized. Each individual is a self-owner and cannot be owned by another individual or group. Therefore, in the labor field, no one man or group can own the total supply and withhold part of it from the market. Each man owns himself.

Let us call the total supply of a monopolist's product P . When he withholds W units in order to obtain a monopoly for $P - W$, the increased revenue he obtains from $P - W$ must more than compensate him for the loss of revenue he suffers from not selling W . A monopolist's action is always limited by loss of revenue from the withheld supply. But in the case of labor unions,

⁵⁸The first to point out the error in the common talk of "monopoly wage rates" of unions was Professor Mises. See his brilliant discussion in *Human Action*, pp. 373–74. Also see P. Ford, *The Economics of Collective Bargaining* (Oxford: Basil Blackwell, 1958), pp. 35–40. Ford also refutes the thesis advanced by the recent "Chicago School" that unions perform a service as sellers of labor:

But a union does not itself produce or sell the commodity, labour, nor receive payment for it. . . . It could be more fitly described as . . . fixing the wages and other conditions on which its individual members are permitted to sell their services to the individual employers. (*Ibid.*, p. 36)

this limitation does not apply. Since each man owns himself, the “withheld” suppliers are *different people* from the ones getting the increased income. If a union, in one way or another, achieves a higher price than its members could command by individual sales, its action is *not* checked by the loss of revenue suffered by the “withheld” laborers. If a union achieves a higher wage, some laborers are earning a higher price, while others are excluded from the market and lose the revenue they would have obtained. Such a higher price (wage) is called a *restrictionist price*.

A restrictionist price, by any sensible criterion, is “worse” than a “monopoly price.” Since the restrictionist union does not have to worry about the laborers who are excluded and suffers no revenue loss from such exclusion, restrictionist action is not curbed by the elasticity of the demand curve for labor. For unions need only maximize the net income of the *working* members, or, indeed, of the union bureaucracy itself.⁵⁹

How may a union achieve a restrictionist price? Figure 69 will illustrate. The demand curve is the demand curve for a labor factor in an industry. *DD* is the demand curve for the labor in the industry; *SS*, the supply curve. Both curves relate the number of laborers on the horizontal axis and the wage rate on the vertical. At the market equilibrium, the supply of laborers offering their work in the industry will intersect the demand for the labor, at number of laborers *OA* and wage rate *AB*. Now, suppose that a union enters this labor market, and the union decides that its members will insist on a higher wage than *AB*, say *OW*. What unions do, in fact, is to insist upon a certain wage

⁵⁹A restrictionist, rather than a monopoly, price can be achieved because the number of laborers is so important in relation to the possible variation in *hours* of work by an individual laborer that the latter can be ignored here. If, however, the total labor supply is limited originally to a few people, then an imposed higher wage rate will cut down the number of hours purchased from the workers who remain working, perhaps so much as to render a restrictionist price unprofitable to them. In such a case it would be more appropriate to speak of a *monopoly* price.

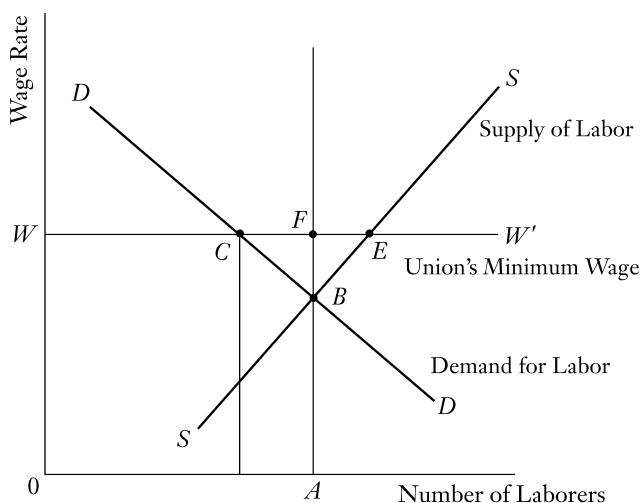


FIGURE 69. FORMATION OF A
RESTRICTIONIST WAGE RATE

rate as a minimum below which they will not work in that industry.

The effect of the union decision is to shift the supply curve of labor available to the industry to a horizontal one at the wage rate WW' , rising after it joins the SS curve at E . The minimum reserve price of labor for this industry has risen, and has risen for all laborers, so that there are no longer laborers with lower reserve prices who would be willing to work for less. With a supply curve changing to WE , the new equilibrium point will be C instead of B . The number of workers hired will be WC , and the wage rate OW .

The union has thus achieved a restrictionist wage rate. It can be achieved regardless of the shape of the demand curve, granting only that it is falling. The demand curve falls because of the diminishing DMVP of a factor and the diminishing marginal utility of the product. But a sacrifice has been made—specifically, there are now fewer workers hired, by an amount CF . *What happens to them?* These discharged workers are the main

losers in this procedure. Since the union represents the remaining workers, it does not have to concern itself, as the monopolist would, with the fate of these workers. At best, they must shift (being a nonspecific factor, they can do so) to some other—nonunionized—industry. The trouble is, however, that the workers are less suited to the new industry. Their having been in the now unionized industry implies that their DMVP in that industry was higher than in the industry to which they must shift; consequently, their wage rate is now lower. Moreover, their entry into the other industry depresses the wage rates of the workers already there.

Consequently, at best, a union can achieve a higher, restrictionist wage rate for its members only at the expense of lowering the wage rates of all other workers in the economy. Production efforts in the economy are also distorted. But, in addition, the wider the scope of union activity and restrictionism in the economy, the more difficult it will be for workers to shift their locations and occupations to find nonunionized havens in which to work. And more and more the tendency will be for the displaced workers to remain permanently or quasi-permanently unemployed, eager to work but unable to find nonrestricted opportunities for employment. The greater the scope of unionism, the more a permanent mass of unemployment will tend to develop.

Unions try as hard as they can to plug all the “loop-holes” of nonunionism, to close all the escape hatches where the dispossessed workmen can find jobs. This is termed “ending the unfair competition of nonunion, low-wage labor.” A universal union control and restrictionism would mean permanent mass unemployment, growing ever greater in proportion to the degree that the union exacted its restrictions.

It is a common myth that only the old-style “craft” unions, which deliberately restrict their occupational group to highly skilled trades with relatively few numbers, can restrict the supply of labor. They often maintain stringent standards of membership and numerous devices to cut down the supply of labor entering the trade. This direct restriction of supply doubtless

makes it easier to obtain higher wage rates for the remaining workers. But it is highly misleading to believe that the newer-style “industrial” unions do not restrict supply. The fact that they welcome as many members in an industry as possible cloaks their restrictionist policy. The crucial point is that the unions insist on a minimum wage rate higher than what would be achieved for the given labor factor without the union. By doing so, as we saw in Figure 69, they necessarily cut the number of men whom the employer can hire. *Ergo*, the consequence of their policy is to restrict the supply of labor, while at the same time they can piously maintain that they are inclusive and democratic, in contrast to the snobbish “aristocrats” of craft unionism.

In fact, the consequences of industrial unionism are more devastating than those of craft unionism. For the craft unions, being small in scope, displace and lower the wages of only a few workers. The industrial unions, larger and more inclusive, depress wages and displace workers on a large scale and, what is even more important, can cause permanent mass unemployment.⁶⁰

There is another reason why an openly restrictionist union will cause less unemployment than a more liberal one. For the union which restricts its membership serves open warning on workers hoping to enter the industry that they are barred from joining the union. As a result, they will swiftly look elsewhere, where jobs can be found. Suppose the union is democratic, however, and open to all. Then, its activities can be described by the above figure; it has achieved a higher wage rate OW for its working members. But such a wage rate, as can be seen on the SS curve, attracts more workers into the industry. In other words, while OA workers were hired by the industry at the previous (nonunion) wage AB , now the union has won a wage OW . At this wage, only WC workers can be employed in the industry. But this wage also *attracts more* workers than before, namely WE . As a result, instead of only CF workers becoming unemployed from

⁶⁰Cf. Mises, *Human Action*, p. 764.

the union's restrictionist wage rate, more—*CE*—will be unemployed in the industry.

Thus, an open union does not have the one virtue of the closed union—rapid repulsion of the displaced workers from the unionized industry. Instead, it attracts even more workers *into* the industry, thus aggravating and swelling the amount of unemployment. With market signals distorted, it will take a much longer time for workers to realize that no jobs are available in the industry. The larger the scope of open unions in the economy, and the greater the differential between their restrictionist wage rates and the market wage rates, the more dangerous will the unemployment problem become.

The unemployment and the misemployment of labor, caused by restrictionist wage rates need not always be directly visible. Thus, an industry might be particularly profitable and prosperous, either as a result of a rise in consumer demand for the product or from a cost-lowering innovation in the productive process. In the absence of unions, the industry would expand and hire more workers in response to the new market conditions. But if a union imposes a restrictionist wage rate, it may not cause the unemployment of any current workers in the industry; it may, instead, simply prevent the industry from expanding in response to the requirements of consumer demand and the conditions of the market. Here, in short, the union destroys *potential* jobs in the making and imposes a misallocation of production by preventing expansion. It is true that, without the union, the industry will bid up wage rates *in the process* of expansion; but if unions impose a higher wage rate at the beginning, the expansion will not occur.⁶¹

⁶¹See Charles E. Lindblom, *Unions and Capitalism* (New Haven: Yale University Press, 1949), pp. 78 ff., 92–97, 108, 121, 131–32, 150–52, 155. Also see Henry C. Simons, “Some Reflections on Syndicalism” in *Economic Policy for a Free Society* (Chicago: University of Chicago Press, 1948), pp. 131 f., 139 ff.; Martin Bronfenbrenner, “The Incidence of Collective Bargaining,” *American Economic Review, Papers and Proceedings*,

Some opponents of unionism go to the extreme of maintaining that unions can *never* be free-market phenomena and are always “monopolistic” or coercive institutions. Although this might be true in actual practice, it is not *necessarily* true. It is very possible that labor unions might arise on the free market and even gain restrictionist wage rates.

How can unions achieve restrictionist wage rates on the free market? The answer can be found by considering the displaced workers. The key problem is: Why do the workers *let themselves* be displaced by the union’s *WW* minimum? Since they were willing to work for less before, why do they now meekly agree to being fired and looking for a poorer-paying job? Why do some remain content to continue in a quasi-permanent pocket of unemployment in an industry, waiting to be hired at the excessively high rate? The only answer, in the absence of coercion, is that they have adopted on a commandingly high place on their value scales the goal of *not undercutting union wage rates*. Unions, naturally, are most anxious to persuade workers, both union and nonunion, as well as the general public, to believe strongly in the sinfulness of undercutting union wage rates. This is shown most clearly in those situations where union members refuse to continue working for a firm at a wage rate below a certain minimum (or on other terms of employment). This situation is known as a *strike*. The most curious thing about a strike is that the unions have been able to spread the belief throughout society that the striking members are still “really” working for the company even when they are deliberately and proudly *refusing* to do so. The natural answer of the employer, of course, is to turn somewhere else and to hire laborers who *are* willing to work on the terms offered. Yet

May, 1954, pp. 301–02; Fritz Machlup, “Monopolistic Wage Determination as a Part of the General Problem of Monopoly” in *Wage Determination and the Economics of Liberalism* (Washington, D.C.: Chamber of Commerce of the United States, 1947), pp. 64–65.

unions have been remarkably successful in spreading the idea through society that anyone who accepts such an offer—the “strikebreaker”—is the lowest form of human life.

To the extent, then, that nonunion workers feel ashamed or guilty about “strike-breaking” or other forms of undercutting union-proclaimed wage scales, the displaced or unemployed workers agree to their own fate. These workers, in effect, are being displaced to poorer and less satisfying jobs voluntarily and remain unemployed for long stretches of time *voluntarily*. It is voluntary because that is the consequence of their voluntary acceptance of the *mystique* of “not crossing the picket line” or of not being a strikebreaker.

The economist *qua* economist can have no quarrel with a man who voluntarily comes to the conclusion that it is more important to preserve union solidarity than to have a good job. But there is one thing an economist *can* do: he can point out to the worker the consequences of his voluntary decision. There are undoubtedly countless numbers of workers who do not realize that their refusal to cross a picket line, their “sticking to the union,” may result in their losing their jobs and remaining unemployed. They do not realize this because to do so requires knowledge of a chain of praxeological reasoning (such as we have been following here). The consumer who purchases directly enjoyable services does not have to be enlightened by economists; he needs no lengthy chain of reasoning to know that his clothing or car or food is enjoyable or serviceable. He can see each perform its service before his eyes. Similarly, the capitalist-entrepreneur does not need the economist to tell him what acts will be profitable or unprofitable. He can see and test them by means of his profits or losses. But for a grasp of the consequences of acts of governmental intervention in the market or of union activity, knowledge of praxeology is requisite.⁶²

⁶²See Murray N. Rothbard, “Mises’ *Human Action*: Comment,” *American Economic Review*, March, 1951, pp. 183–84.

Economics cannot itself decide on ethical judgments. But in order for anyone to make ethical judgments rationally, he must know the consequences of his various alternative courses of action. In questions of government intervention or union action, economics supplies the knowledge of these consequences. Knowledge of economics is therefore necessary, though not sufficient, for making a rational ethical judgment in these fields. As for unions, the consequences of their activity, when discovered (e.g., displacement or unemployment for oneself or others), will be considered unfortunate by most people. Therefore, it is certain that when knowledge of these consequences becomes widespread, far fewer people will be “prounion” or hostile to “nonunion” competitors.⁶³

Such conclusions will be reinforced when people learn of another consequence of trade union activity: that a restrictionist wage raises costs of production for the firms in the industry. This means that the marginal firms in the industry—the ones whose entrepreneurs earn only a bare rent—will be driven out of business, for their costs have risen above their most profitable price on the market—the price that had *already* been attained. Their ejection from the market and the general rise of average costs in the industry signify a general fall in productivity and output, and hence a loss to the consumers.⁶⁴ Displacement and unemployment, of course, also impair the general standard of living of the consumers.

Unions have had other important economic consequences. Unions are not *producing* organizations; they do not work for capitalists to improve production.⁶⁵ Rather they attempt to

⁶³The same is true, to an even greater extent, of measures of governmental intervention in the market. See chapter 12 below.

⁶⁴See James Birks, *Trade Unionism in Relation to Wages* (London, 1897), p. 30.

⁶⁵See James Birks, *Trades' Unionism: A Criticism and a Warning* (London, 1894), p. 22.

persuade workers that they can better their lot at the expense of the employer. Consequently, they invariably attempt as much as possible to establish work rules that hinder management's directives. These work rules amount to preventing management from arranging workers and equipment as it sees fit. In other words, instead of agreeing to submit to the work orders of management in exchange for his pay, the worker now sets up not only minimum wages, but also work rules without which he refuses to work. The effect of these rules is to *lower the marginal productivity of all union workers*. The lowering of marginal value-product schedules has a twofold result: (1) it itself establishes a restrictionist wage scale with its various consequences, for the marginal value product has fallen while the union insists that the wage rate remain the same; (2) consumers lose by a general lowering of productivity and living standards. Restrictive work rules therefore also lower output. All this is perfectly consistent with a society of individual sovereignty, however, provided always that no force is employed by the union.

To advocate coercive abolition of these work rules would imply literal enslavement of the workers to the dictates of catallactic consumers. But, once again, it is certain that knowledge of these various consequences of union activity would greatly weaken the voluntary adherence of many workers and others to the *mystique* of unionism.⁶⁶

⁶⁶We can deal here only with the directly catallactic consequences of labor unionism. Unionism also has other consequences which many might consider even more deplorable. Prominent is the fusing of the able and the incompetent into one group. Seniority rules, for example, are invariable favorites of unions. They set restrictively high wages for less able workers and also lower the productivity of all. But they also *reduce* the wages of the more able workers—those who must be chained to the stultifying march of seniority for their jobs and promotions. Seniority also decreases the mobility of workers and creates a kind of industrial serfdom by establishing vested rights in jobs according to the length of time the employees have worked. Cf. David McCord Wright, "Regulating Unions" in Bradley, *Public Stake in Union Power*, pp. 113–21.

Unions, therefore, are theoretically compatible with the existence of a purely free market. In actual fact, however, it is evident to any competent observer that unions acquire almost all their power through the wielding of force, specifically force against strikebreakers and against the property of employers. An implicit license to unions to commit violence against strikebreakers is practically universal. Police commonly either remain “neutral” when strikebreakers are molested or else blame the strikebreakers for “provoking” the attacks upon them. Certainly, few pretend that the institution of mass picketing by unions is simply a method of advertising the fact of a strike to anyone passing by. These matters, however, are empirical rather than theoretical questions. Theoretically, we may say that it is possible to have unions on a free market, although empirically we may question how great their scope would be.

Analytically, we can also say that when unions are permitted to resort to violence, the state or other enforcing agency has implicitly delegated this power to the unions. The unions, then, have become “private states.”⁶⁷

We have, in this section, investigated the consequences of unions’ achieving restrictionist prices. This is not to imply, however, that unions *always* achieve such prices in collective bargaining. Indeed, because unions do not own workers and therefore do not sell their labor, the collective bargaining of unions is an artificial replacement for the smooth workings of “individual bargaining” on the labor market. Whereas wage rates on the nonunion labor market will always tend toward equilibrium in a smooth and harmonious manner, its replacement by collective bargaining leaves the negotiators with little or no rudder, with little guidance on what the proper wage rates

⁶⁷Students of labor unions have almost universally ignored the systematic use of violence by unions. For a welcome exception, see Sylvester Petro, *Power Unlimited* (New York: Ronald Press, 1959). Also cf. F.A. Hayek, “Unions, Inflation, and Profits,” p. 47.

would be. Even with both sides trying to *find* the market rate, neither of the parties to the bargain could be sure that a given wage agreement is too high, too low, or approximately correct. Almost invariably, furthermore, the union is not *trying* to discover the market rate, but to impose various arbitrary “principles” of wage determination, such as “keeping up with the cost of living,” a “living wage,” the “going rate” for comparable labor in other firms or industries, an annual average “productivity” increase, “fair differentials,” etc.⁶⁸

B. SOME ARGUMENTS FOR UNIONS: A CRITIQUE

(1) *Indeterminacy*⁶⁹

A favorite reply of union advocates to the above analysis is this: “Oh, that is all very well, but you are overlooking the indeterminacy of wage rates. Wage rates are determined by marginal productivity in a *zone* rather than at a point; and within that zone unions have an opportunity to bargain collectively for increased wages without the admittedly unpleasant effects of unemployment or displacement of workers to poorer jobs.” It is curious that many writers move smoothly through rigorous price analysis until they come to wage rates, when suddenly they lay heavy stress on indeterminacy, the huge zones within which the price makes no difference, etc.

In the first place, the scope of indeterminacy is very small in the modern world. We have seen above that, in a two-person barter situation, there is likely to be a large zone of indeterminacy between the buyer’s maximum demand price and the seller’s minimum supply price for a quantity of a good. Within this zone, we can only leave the determination of the price to

⁶⁸On the nature and consequences of these various criteria of wage determination, see Ford, *Economics of Collective Bargaining*, pp. 85–110.

⁶⁹See the excellent critique by Hutt, *Theory of Collective Bargaining*, *passim*.

bargaining. However, it is precisely the characteristic of an advanced monetary economy that these zones are ever and ever narrowed and lose their importance. The zone is only between the “marginal pairs” of buyers and sellers, *and this zone is constantly dwindling as the number of people and alternatives in the market increase*. Growing civilization, therefore, is always narrowing the importance of indeterminacies.

Secondly, there is no reason whatever why a zone of indeterminacy should be more important for the labor market than for the market for the price of any other good.

Thirdly, suppose that there *is* a zone of indeterminacy for a labor market, and let us assume that no union is present. This means that there is a certain zone, the length of which can be said to equal a zone of the discounted marginal value product of the factor. This, parenthetically, is far less likely than the existence of a zone for a consumers’ good, since in the former case there is a specific amount, a DMVP, to be estimated. But the *maximum* of the supposed zone is the highest point at which the wage equals the DMVP. Now, competition among employers will tend to raise factor prices to precisely that height at which profits will be wiped out. In other words, wages will tend to be raised to the *maximum* of any zone of the DMVP.

Rather than wages being habitually at the bottom of a zone, presenting unions with a golden opportunity to raise wages to the top, the truth is quite the reverse. Assuming the highly unlikely case that any zone exists at all, wages will tend to be at the *top*, so that the only remaining indeterminacy is downward. Unions would have no room for increasing wages within that zone.

(2) *Monopsony and Oligopsony*

It is often alleged that the buyers of labor—the employers—have some sort of monopoly and earn a monopoly gain, and that therefore there is room for unions to raise wage rates without injuring other laborers. However, such a “monopsony” for the purchase of labor would have to encompass all the entrepreneurs

in the society. If it did not, then labor, a nonspecific factor, could move into other firms and other industries. And we have seen that one big cartel cannot exist on the market. Therefore, a “monopsony” cannot exist.

The “problem” of “oligopsony”—a “few” buyers of labor—is a pseudo problem. As long as there is no monopsony, competing employers will tend to drive up wage rates until they equal their DMVPs. The *number* of competitors is irrelevant; this depends on the concrete data of the market. Below, we shall see the fallacy of the idea of “monopolistic” or “imperfect” competition, of which this is an example. Briefly, the case of “oligopsony” rests on a distinction between the case of “pure” or “perfect” competition, in which there is an allegedly horizontal—infinately elastic—supply curve of labor, and the supposedly less elastic supply curve of the “imperfect” oligopsony. Actually, since people do not move *en masse* and all at once, the supply curve is never infinitely elastic, and the distinction has no relevance. There is only free competition, and no other dichotomies, such as between pure competition and oligopsony, can be established. The shape of the supply curve, furthermore, makes no difference to the truth that labor or any other factor tends to get its DMVP on the market.

(3) *Greater Efficiency and the “Ricardo Effect”*

One common prounion argument is that unions benefit the economy through forcing higher wages on the employers. At these higher wages the workers will become more efficient, and their marginal productivity will rise as a result. If this were true, however, no unions would be needed. Employers, ever eager for greater profits, would see this and pay higher wages now to reap the benefits of the allegedly higher productivity in the future. As a matter of fact, employers often train workers, paying higher wages than their *present* marginal product justifies, in order to reap the benefits of their increased productivity in later years.

A more sophisticated variant of this thesis was advanced by Ricardo and has been revived by Hayek. This doctrine holds that union-induced higher wage rates encourage employers to substitute machinery for labor. This added machinery increases the capital per worker and raises the marginal productivity of labor, thereby paying for the higher wage rates. The fallacy here is that only increased saving can make more capital available. Capital investment is limited by saving. Union wage increases do not increase the total supply of capital available. Therefore, there can be no general rise in labor productivity. Instead, the potential supply of capital is *shifted* (not increased) from other industries to those industries with higher wage rates. And it is shifted to industries where it would have been less profitable under nonunion conditions. The fact that an induced higher wage rate shifts capital to the industry does not indicate economic progress, but rather an attempt, never fully successful, to offset an economic retrogression—a higher cost in the manufacture of the product. Hence, the shift is “uneconomic.”

A related thesis is that higher wage rates will spur employers to invent new technological methods to make labor more efficient. Here again, however, the supply of capital goods is limited by the savings available, and there is almost always a sheaf of technological opportunities awaiting more capital anyway. Furthermore, the spur of competition and the desire of the producer to keep and increase his custom is enough of an incentive to increase productivity in his firm, without the added burden of unionism.⁷⁰

⁷⁰On the Ricardo effect, see Mises, *Human Action*, pp. 767–70. Also see the detailed critique by Ford, *Economics of Collective Bargaining*, pp. 56–66, who also points to the union record of hindering mechanization by imposing restrictive work rules and by moving quickly to absorb any possible gain from the new equipment.

5. The Theory of Monopolistic or Imperfect Competition

A. MONOPOLISTIC COMPETITIVE PRICE

The theory of monopoly price has been generally superseded in the literature by the theories of “monopolistic” or “imperfect” competition.⁷¹ As against the older theory, the latter have the advantage of setting up *identifiable* criteria for their categories—such as a perfectly elastic demand curve for pure competition. Unfortunately, these criteria turn out to be completely fallacious.

Essentially, the chief characteristic of the imperfect-competition theories is that they uphold as their “ideal” the state of “pure competition” rather than “competition” or “free competition.” *Pure competition* is defined as that state in which the demand curve for each firm in the economy is *perfectly elastic*, i.e., the demand curve as presented to the firm is completely horizontal. In this supposedly pristine state of affairs, no one firm can, through its actions, possibly have any influence over the price of its product. Its price is then “set” for it by the market. Any amount it produces can and will be sold at this ruling price. In general, it is this state of affairs, or else this state without uncertainty (“perfect competition”), that has received most of the elaborate analysis in recent years. This is true both for those who believe that pure competition fairly well represents the real economy and for their opponents, who consider it only an ideal with which to contrast the actual “monopolistic” state of affairs. Both camps, however, join in upholding pure competition as the ideal system for the general welfare, in contrast to various vague “monopoloid” states that occur when there is departure from the purely competitive world.

⁷¹In particular, see Edward H. Chamberlin, *Theory of Monopolistic Competition*, and Mrs. Joan Robinson, *Economics of Imperfect Competition*. For a lucid discussion and comparison of the two works, see Robert Triffin, *Monopolistic Competition and General Equilibrium Theory* (Cambridge: Harvard University Press, 1940). The differences between the “monopolistic” and the “imperfect” formulations are not important here.

The pure-competition theory, however, is an utterly fallacious one. It envisages an absurd state of affairs, never realizable in practice, and far from idyllic if it were. In the first place, there can be no such thing as a firm *without* influence on its price. The monopolistic-competition theorist contrasts this ideal firm with those firms that have some influence on the determination of price and are therefore in some degree “monopolistic.” Yet it is obvious that the demand curve to a firm *cannot* be perfectly elastic throughout. At some points, it must dip downward, since the increase in supply will tend to lower market price. As a matter of fact, it is clear from our construction of the demand curve that there can be *no* stretch of the demand curve, however small, that is horizontal, although there can be small vertical stretches. In aggregating the market demand curve, we saw that for each hypothetical price, the consumers will decide to purchase a certain amount. If the producers attempt to sell a larger amount, they will have to conclude their sale at a lower price in order to attract an increased demand. Even a very small increase in supply will lead to a perhaps very small lowering of price. The individual firm, no matter how small, always has a perceptible influence on the total supply. In an industry of small wheat farms (the implicit model for “pure competition”), each small farm contributes a part of the total supply, and there can be no total without a contribution from each farm. Therefore, each farm has a perceptible, even if very small, influence. No perfectly elastic demand curve can, then, be postulated even in such a case. The error in believing in “perfect elasticity” stems from the use of such mathematical concepts as “second order of smalls,” by which infinite negligibility of steps can be assumed. But economics analyzes real human action, and such real action must always be concerned with discrete, perceptible steps, and never with “infinitely small” steps.

Of course, the demand curve for each small wheat farm is likely to be very highly, *almost* perfectly, elastic. And yet the fact that it is *not* “perfect” destroys the entire concept of pure competition. For how does this situation differ from, say, the

Hershey Chocolate Company if the demand curve for the latter firm is also elastic? Once it is conceded that all demand curves to firms must be falling, the monopolistic-competition theorist can make no further analytic distinctions.

We cannot compare or classify the curves on the basis of *degrees of elasticity*, since there is nothing in the Chamberlin-Robinson monopolistic-competition analysis, or in any part of praxeology for that matter, that permits us to do so, once the case of pure competition is rejected. For praxeology cannot establish *quantitative* laws, only *qualitative* ones. Indeed, the only recourse of monopolistic-competition theorists would be to fall back on the concepts of “inelastic” vs. “elastic” demand curves, and this would precisely plunge them right back into the old monopoly-price vs. competitive-price dichotomy. They would have to say, with the old monopoly-price theorists, that if the demand curve for the firm is more than unitarily elastic at the equilibrium point, the firm will remain at the “competitive” price; that if the curve is inelastic, it will rise to a monopoly-price position. But, as we have already seen in detail, the monopoly-competitive price dichotomy is untenable.

According to the monopolistic-competition theorists, the two influences sabotaging the possible existence of pure competition are “differentiation of product” and “oligopoly,” or fewness of firms, where one firm influences the actions of others. As to the former, the producers are accused of creating an artificial differentiation among products in the mind of the public, thus carving out for themselves a portion of monopoly. And Chamberlin originally attempted to distinguish “groups” of producers selling “slightly” differentiated products from old-fashioned “industries” of firms making identical products. Neither of these attempts has any validity. If a producer is making a product different from that of another producer, then he is a unique “industry”; there is no rational basis for any grouping of varied producers, particularly in aggregating their demand curves. Furthermore, the consuming public decides on the differentiation of products on its value scales. There is nothing

“artificial” about the differentiation, and indeed this differentiation serves to cater more closely to the multifarious wants of the consumers.⁷² It is clear, of course, that Ford has a monopoly on the sale of Ford cars; but this is a full “monopoly” rather than a “monopolistic” tendency. Also, it is difficult to see what difference can come from the number of firms that are producing the same product, particularly once we discard the myth of pure competition and perfect elasticity. Much ado indeed has been made about strategies, “warfare,” etc., between oligopolists, but there is little point to such discussions. Either the firms are independent and therefore competing, or they are acting jointly and therefore cartelizing. There is no third alternative.

Once the perfect-elasticity myth has been discarded, it becomes clear that all the tedious discussion about the number and size of firms and groups and differentiation, etc., becomes irrelevant. It becomes relevant only for economic history, and not for economic analysis.

It might be objected that there is a substantial problem of oligopoly: that, under oligopoly, each firm has to take into account the reactions of competing firms, whereas under pure competition or differentiated products without oligopoly, each

⁷²Recently, Professor Chamberlin has conceded this point and has, in a series of remarkable articles, astounded his followers by repudiating the concept of pure competition as a welfare ideal. Chamberlin now declares: “The welfare ideal itself . . . is correctly described as one of monopolistic competition. . . . [This] seems to follow very directly from the recognition that human beings are individual, diverse in their tastes and desires, and moreover, widely dispersed spatially.” Chamberlin, *Towards a More General Theory of Value*, pp. 93–94; also *ibid.*, pp. 70–83; E.H. Chamberlin and J.M. Clark, “Discussion,” *American Economic Review, Papers and Proceedings*, May, 1950, pp. 102–04; Hunter, “Product Differentiation and Welfare Economics,” pp. 533–52; Hayek, “The Meaning of Competition” in *Individualism and the Economic Order*, p. 99; and Marshall I. Goldman, “Product Differentiation and Advertising: Some Lessons from Soviet Experience,” *Journal of Political Economy*, August, 1960, pp. 346–57. See also note 28 above.

firm can operate in the blissful awareness that no competitor will take account of its actions or change its actions accordingly. Hiram Jones, the small wheat farmer, can set his production policy without wondering what Ezra Smith will do when he discovers what Jones' policy is. Ford, on the other hand, must consider General Motors' reactions, and *vice versa*. Many writers, in fact, have gone so far as to maintain that economics can simply not be applied to these "oligopoly" situations, that these are indeterminate situations where "anything may happen." They define the buyers' demand curve that presents itself to the firm as *assuming no reaction* by competing firms. Then, since "few firms" exist and each firm takes account of the reactions of others, they proceed to the conclusion that in the real world all is chaos, incomprehensible to economic analysis.

These alleged difficulties are nonexistent, however. There is no reason why the demand curve to a firm cannot *include* expected reactions by other firms.⁷³ The demand curve to a firm is the set of a firm's expectations, at any time, of how many units of its product consumers will buy at an alternative series of prices. What interests the producer is the hypothetical set of consumer demands at each price. He is not interested in what consumer demand will be in various sets of nonexistent situations. His expectations will be based on his judgment of what would actually happen should he charge various alternative prices. If his rivals will react in a certain way to his charging a higher or a lower price, then it is each firm's business to *forecast and take account of this reaction* in so far as it will affect buyers' demand for its particular product. There would be little sense

⁷³This definition of the demand curve to the firm was Mrs. Robinson's outstanding contribution, unfortunately repudiated by her recently. Triffin castigated Mrs. Robinson for evading the problem of "oligopolistic indeterminacy," whereas actually she had neatly solved this pseudo problem. See Robinson, *Economics of Imperfect Competition*, p. 21. For other aspects of oligopoly, see Willard D. Arant, "Competition of the Few Among the Many," *Quarterly Journal of Economics*, August, 1956, pp. 327-45.

in ignoring such reactions if they were relevant to the demand for its product or in including them if they were not. A firm's estimated demand curve, therefore, *already includes* any expected reactions of rivals.

The relevant consideration is not the fewness of the firms or the state of hostility or friendship existing among firms. Those writers who discuss oligopoly in terms applicable to games of poker or to military warfare are entirely in error. The fundamental business of production is service to the consumers for monetary gain, and not some sort of "game" or "warfare" or any other sort of struggle *between* producers. In "oligopoly," where several firms are producing an identical product, there cannot persist any situation in which one firm charges a higher price than another, since there is always a tendency toward the formation of a uniform price for each uniform product. Whenever firm A attempts to sell its product higher or lower than the previously ruling market price, it is attempting to "discover the market," to find out what the equilibrium market price is, in accordance with the present state of consumer demand. If, at a certain price for the product, consumer demand is in excess of supply, the firms will tend to raise the price, and *vice versa* if the produced stock is not being sold. In this familiar pathway to equilibrium, all the stock that the firms wish to sell "clears the market" at the highest price that can be obtained. The jockeying and raising and lowering of prices that takes place in "oligopolistic" industries is not some mysterious form of warfare, but the visible process of attempting to find market equilibrium—that price at which the quantity supplied and the quantity demanded will be equal. The same process, indeed, takes place in any market, such as the "nonoligopolistic" wheat or strawberry markets. In the latter markets the process seems to the viewer more "impersonal," because the actions of any one individual or firm are not as important or as strikingly visible as in the more "oligopolistic" industries. But the process is essentially the same, and we must not be led to think differently by such often inapt metaphors as the "automatic mechanisms of

the market” or the “soulless, impersonal forces on the market.” All action on the market is necessarily *personal*; machines may move, but they do not purposefully *act*. And, in oligopoly situations, the rivalries, the feelings of one producer toward his competitors, may be historically dramatic, but they are unimportant for economic analysis.

To those who are still tempted to make the number of producers in any field the test of competitive merit, we might ask (setting aside the problem of proving homogeneity): How can the market create sufficient numbers? If Crusoe exchanges fish for Friday’s lumber on their desert island, are they both benefiting, or are they “bilateral monopolists” exploiting each other and charging each other monopoly prices? But if the State is not justified in marching in to arrest Crusoe and/or Friday, how can it be justified in coercing a market where there are obviously many *more* competitors?

Economic analysis, in conclusion, fails to establish any criterion for separating any elements of the free-market price for a product. Such questions as the number of firms in an industry, the sizes of the firms, the type of product each firm makes, the personalities or motives of the entrepreneurs, the location of plants, etc., are entirely determined by the concrete conditions and data of the particular case. Economic analysis can have nothing to say about them.⁷⁴

B. THE PARADOX OF EXCESS CAPACITY

Perhaps the most important conclusion of the theory of monopolistic or imperfect competition is that the real world of

⁷⁴For an acute criticism of monopolistic-competition theory, see L.M. Lachmann, “Some Notes on Economic Thought, 1933–53,” *South African Journal of Economics*, March, 1954, pp. 26 ff., especially pp. 30–31. Lachmann points out that economists generally treat types of “perfect” or “monopolistic” competition as static market forms, whereas competition is actually a dynamic process.

monopolistic competition (where the demand curve to each firm is necessarily falling) is inferior to the ideal world of pure competition (where no firm can affect its price). This conclusion was expressed simply and effectively by comparing two final equilibrium states: under conditions of pure and monopolistic competition (Figure 70).

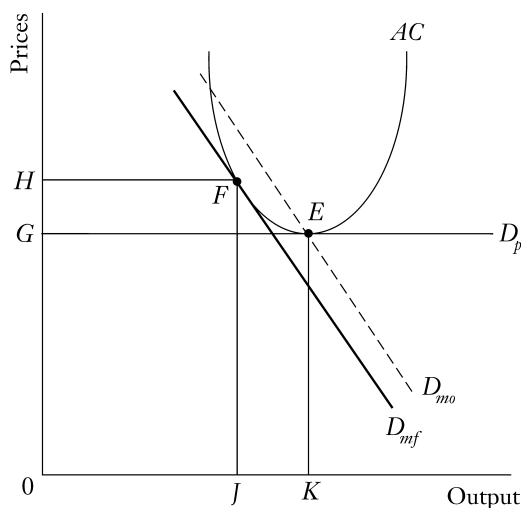


FIGURE 70. FINAL EQUILIBRIUM STATES UNDER CONDITIONS OF PURE AND MONOPOLISTIC COMPETITION

AC is a firm's average total-cost curve—its alternative dollar costs per unit—with output on the horizontal axis and prices (including costs) on the vertical axis. The only assumption we need in drawing the average-cost curve is that, for any plant in any branch of production, there will be *some optimum* point of production, i.e., some level of output at which average unit cost is at a minimum. All levels of production lower or higher than the optimum have a higher average cost. In pure competition, where the demand curve for any firm is perfectly elastic, D_p , each firm will eventually adjust so that its AC curve will be

tangent to D_p , in equilibrium; in this case, at point E . For if average revenue (price) is greater than average cost, then competition will draw in other firms, until the curves are tangent; if the cost curve is irretrievably higher than demand, the firm will go out of business. Tangency is at point E , price at OG , and output at OK . As in any definition of final equilibrium, total costs equal total revenues for each firm, and profits are zero.

Now contrast this picture with that of monopolistic competition. Since the demand curve (D_{mf}) is now sloping downward to the right, it must, given the same AC curve, be tangent at some point (F), where the price is higher (JF) and the production lower (OJ) than under pure competition. In short, monopolistic competition yields higher prices and less production—i.e., a lower standard of living—than pure competition. Furthermore, output will not take place at the point of minimum average cost—clearly a social “optimum,” and each plant will produce at a lower than optimum level, i.e., it will have “excess capacity.” This was the “welfare” case of the monopolistic-competition theorists.

By a process of revision in recent years, some of it by the originators of the doctrine themselves, this theory has been effectively riddled beyond repair. As we have seen, Chamberlin and others have shown that this analysis does not apply if we are to take consumer desire for diversity as a good to be satisfied.⁷⁵ Many other effective and sound attacks have been made from different directions. One basic argument is that the situations of pure and of monopolistic competition cannot be compared

⁷⁵And the product differentiation associated with the falling demand curve may well lower costs of distribution and of inspection (as well as improve consumer knowledge) to more than offset the supposed rise in production costs. In short, the AC curve above is *really* a production-cost, rather than a total-cost, curve, neglecting distribution costs. Cf. Goldman, “Product Differentiation and Advertising.” Furthermore, a genuine total-cost curve would then not be independent of the firm’s demand curve, thus vitiating the usual “cost-curve” analysis. See Dewey, *Monopoly in Economics and Law*, p. 87. Also see section C below.

because the *AC* curves *would not*, in fact, be the same. Chamberlin has pursued his revisionism in this realm also, declaring that the comparisons are wholly illegitimate, that to apply the concept of pure competition to existing firms would mean, for example, assuming a very large number of similar firms producing the identical product. If this were done, say, with General Motors, it would mean that *either* GM must conceptually be divided up into numerous fragments, or else that it be multiplied. If divided, then unit costs would undoubtedly be higher, and then the “competitive firm” would suffer higher costs and have to subsist on higher prices. This would clearly injure consumers and the standard of living; thus, Chamberlin follows Schumpeter’s criticism that the “monopolistic” firm may well have and probably will have lower costs than its “purely competitive” counterpart. If, on the other hand, we conceive of the multiplication of a very large number of General Motors corporations at existing size, we cannot possibly relate it to the present world, and the whole comparison becomes absurd.⁷⁶

In addition, Schumpeter has stressed the superiority of the “monopolistic” firm for innovation and progress, and Clark has shown the inapplicability, in various ways, of this static theory to the dynamic real world. He has recently shown its fallacious asymmetry of argument with respect to price and quality. Hayek and Lachmann have also pointed out the distortion of dynamic reality, as we have indicated above.⁷⁷

⁷⁶See Chamberlin, “Measuring the Degree of Monopoly and Competition” and “Monopolistic Competition Revisited” in *Towards a More General Theory of Value*, pp. 45–83.

⁷⁷See J.M. Clark, “Competition and the Objectives of Government Policy” in E.H. Chamberlin, ed., *Monopoly and Competition and Their Regulation* (London: Macmillan & Co., 1954), pp. 317–27; Clark, “Toward a Concept of Workable Competition” in *Readings in the Social Control of Industry* (Philadelphia: Blakiston, 1942), pp. 452–76; Clark, “Discussion”; Abbott, *Quality and Competition*, *passim*; Joseph A. Schumpeter, *Capitalism, Socialism and Democracy* (New York: Harper & Bros., 1942); Hayek, “Meaning of Competition”; Lachmann, “Some Notes on Economic Thought, 1933–53.”

A second major line of attack has shown that the comparisons are much less important than they seem from conventional diagrams, because cost curves are empirically much flatter than they appear in the textbooks. Clark has emphasized that firms deal in *long-run* considerations, and that long-run cost and demand curves are both more elastic than short-run; hence the differences between *E* and *F* points will be negligible and may be nonexistent. Clark and others have stressed the vital importance of *potential* competition to any would-be reaper of monopoly price, from firms both within and without the industry, and also the competition of substitutes between industries. A further argument has been that the cost curves, empirically, are flat within the relevant range, even aside from the long- vs. short-run problems.⁷⁸

All these arguments, added to our own analysis given above, have effectively demolished the theory of monopolistic competition, and yet more remains to be said. There is something very peculiar about the entire construction, even on its own terms, aside from the fallacious “cost-curve” approach, and practically no one has pointed out these other grave defects in the theory. In an economy that is almost altogether “monopolistically competitive,” how can *every* firm produce too little and charge too much? What happens to the surplus factors? What are they doing? The failure to raise this question stems from the

⁷⁸See the above citations by Clark; and Richard B. Heflebower, “Toward a Theory of Industrial Markets and Prices” in R.B. Heflebower and G.W. Stocking, eds., *Readings on Industrial Organization and Public Policy* (Homewood, Ill.: Richard D. Irwin, 1958), pp. 297–315. A more dubious argument—the flatness of the firm’s demand curve in the relevant range—has been stressed by other economists, notably A.J. Nichol, “The Influence of Marginal Buyers on Monopolistic Competition,” *Quarterly Journal of Economics*, November, 1934, pp. 121–34; Alfred Nicols, “The Rehabilitation of Pure Competition,” *Quarterly Journal of Economics*, November, 1947, pp. 31–63; and Nutter, “Plateau Demand Curve and Utility Theory.”

modern neglect of Austrian general analysis and from undue concentration on an isolated firm or industry.⁷⁹ The excess factors must go somewhere, and in that case must they not go to other monopolistically competitive firms? In which case, the thesis breaks down as self-contradictory. But the proponents have prepared a way out. They take, first, the case of pure competition, with equilibrium at point *E*. Then, they assume a sudden shift to conditions of monopolistic competition, with the demand curve for the firm now sloping downward. The demand curve now shifts from D_p to D_{mo} . Then the firm restricts production and raises its price accordingly, reaps profits, attracts new firms entering the industry, the new competition reduces the output salable by each firm, and the demand curve shifts downward and to the left until it is tangent to the *AC* curve at point *F*. Hence, say the monopolistic-competition theorists, not only does monopolistic competition suffer from too little production in each firm and excessive costs and prices; it also suffers from *too many firms* in each industry. Here is what has happened to the excess factors: they are trapped in too many uneconomic firms.

This seems plausible, until we realize that the whole example has been constructed as a trick. If we isolate a firm or an industry, as does the example, we may just as well start from a position of monopolistic competition, at point *F*, and then suddenly shift to conditions of pure competition. This is certainly just as legitimate, or rather illegitimate, a base for comparison. What then? As we see in Figure 71, the demand curve for each firm is now shifted from D_{mf} to D_{po} . It will now be profitable for each firm to expand its output, and it will then make profits. New firms will then be attracted into the industry, and the demand curve will fall vertically, until it again reaches tangency with the *AC* curve at point *E*. Are we now “proving” that there are *more* firms in an industry under *pure*

⁷⁹But cf. Abbott, *Quality and Competition*, pp. 180–81.

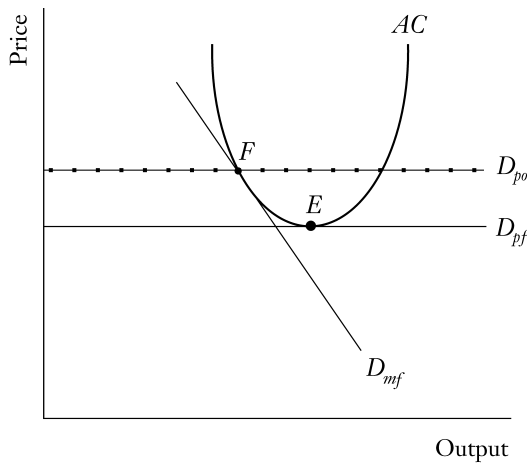


FIGURE 71. CONSEQUENCE OF A SHIFT FROM
MONOPOLISTIC TO PURE COMPETITION

than under monopolistic competition?⁸⁰ The fundamental error here is failure to see that, under the conditions established by the assumptions, *any* change opening up profits will bring new firms into an industry. Yet the theorists are supposed to be comparing two different static equilibria, of pure and of monopolistic competition, and not discussing paths from one to the other. Thus, the monopolistic-competition theorists have by no means solved their problem of surplus factors.

But, aside from this point, there are more difficulties in the theory, and Sir Roy Harrod, himself one of its originators, is the only one to have seized the essence of the remaining central difficulty. As Harrod says:

If the entrepreneur foresees the trend of events, which will in due course limit his profitable output to $x - y$ units, why not plan to have a plant that will produce $x - y$ units most cheaply, rather than encumber

⁸⁰The author first learned this particular piece of analysis from the classroom lectures of Professor Arthur F. Burns, and, to our knowledge, it has never seen print.

himself with excess capacity? To plan a plant for producing x units, while knowing that it will only be possible to maintain an output of $x - y$ units, is surely to suffer from schizophrenia.

And yet, asserts Harrod puzzledly, the “accepted doctrine” apparently deems it “impossible to be an entrepreneur and not suffer from schizophrenia!”⁸¹ In short, the theory assumes that, in the long run, a firm having to produce at F will yet construct a plant with minimum costs at point E . Clearly, here is a patent contradiction with reality. What is wrong? Harrod’s own answer is an excellent and novel discussion of the difference between long-run and short-run demand curves, with the “long run” always being a factor in entrepreneurial planning, but he does not precisely answer this question.

The paradox becomes “curiouser and curiouser” when we fully realize that it all hinges on a mathematical technicality. The reason why a firm can never produce at an optimum cost point is that (a) it must produce at a tangent of demand and average-cost curves in equilibrium, and (b) if the demand curve is falling, it follows that it can be tangent to a U-shaped cost curve only at some point higher than, and to the left of, the trough point. There are two considerations that we may now add. First, there is no reason why the cost “curve” should, in fact, be curved. In an older day, textbook demand curves used to be curves, and now they are often straight lines; there is even more reason for believing that cost curves are a series of angular lines. It is of course (a) more convenient for diagrams, and (b) essential to mathematical representation, for there to be continuous curves, but we must never let reality be falsified in order to fit the niceties of mathematics. In fact, production is a series of discrete alternatives, as all human action is discrete, and cannot be smoothly continuous, i.e., move in infinitely small steps

⁸¹Harrod, *Economic Essays*, p. 149.

from one production level to another. But once we recognize the discrete, angular nature of the cost curve, the “problem” of excess capacity immediately disappears (Figure 72). Thus the falling demand curve to the “monopolistic” firm, D_m , can now be “tangent” to the AC curve at E , the minimum-cost point, and will be so in final equilibrium.

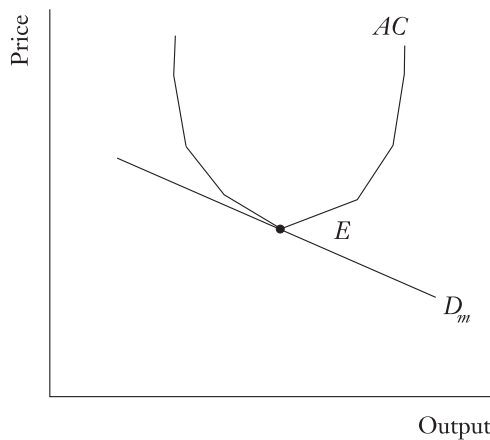


FIGURE 72. DETERMINATION OF PRICE
AT A POINT OF MINIMUM COST

There is another way for this pseudo problem to disappear, and that is to call into question the entire assumption of tangency. The tangency of average cost and demand at equilibrium has appeared to follow from the property of equilibrium: that total costs and total revenues of the firm will be equal, since profits as well as losses will be zero. But a key question has been either overlooked or wrongly handled. Why should the firm produce *anything*, after all, if it earns nothing from doing so? But it will earn something, in equilibrium, and that will be *interest* return. Modern orthodoxy has fallen into this error, for one reason: because it does not realize that entrepreneurs are

also capitalists and that even if, in an evenly rotating economy, the strictly entrepreneurial function were no longer to be required, the capital-advancing function would still be emphatically necessary.

Modern theory also tends to view interest return as a *cost* to the firm. Naturally, if this is done, then the presence of interest does not change matters. But (and here we refer the reader to foregoing chapters) interest is *not* a cost to the firm; it is an earning *by* a firm. The contrary belief rests on a superficial concentration on loan interest and on an unwarranted separation between entrepreneurs and capitalists. Actually, loans are unimportant and are only another legal form of entrepreneurial-capitalist investment. In short, in the evenly rotating economy, the *firm* earns a “natural” interest return, dictated by social time preference. Hence, Figure 72 must be altered to look like the diagram in Figure 73 (setting aside the problem of curves vs. angles). The firm will produce OK , its optimum production level, at minimum average cost, KE . Its demand curve and cost curve will *not* be tangent to each other, but will allow room for

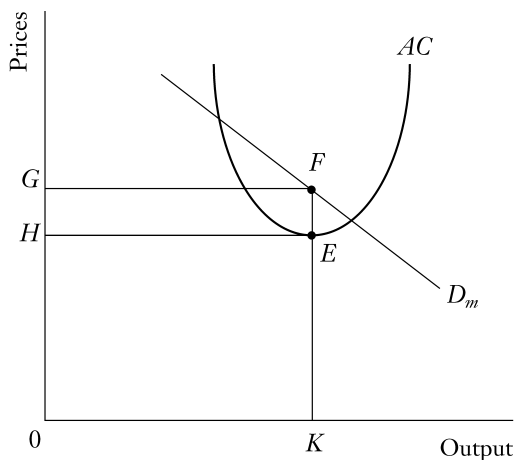


FIGURE 73. EQUILIBRIUM INTEREST RETURN
AS A COMPONENT OF PRICE

equilibrium interest return, represented by the area *EFGH*. (Neither, as some may object, will the price be higher in this corrected version of monopolistic competition; for this *AC* curve is lower all around than the previous ones, which had included interest return in costs. If they did not include interest, and instead assumed that interest would be zero in the ERE, then they were wrong, as we have pointed out above.)⁸² And so the paradox of the monopolistic-competition theory is finally and fully interred.⁸³

C. CHAMBERLIN AND SELLING COST

One of Professor Chamberlin's most important contributions is alleged to have been his sharp distinction between "selling cost" and "production cost."⁸⁴ "Production costs" are supposed to be the legitimate expenses needed to increase supply in order to meet *given* consumer demand schedules. "Selling

⁸²After arriving at this conclusion, the author came across a brilliant but neglected article pointing out that interest is a return and not a cost, and showing the devastating implications of this fact for cost-curve theory. The article does not, however, apply the theory satisfactorily to the problem of monopolistic competition. See Gabor and Pearce, "A New Approach to the Theory of the Firm," and *idem*, "The Place of Money Capital." While there are a few similarities, Professor Dewey's critique of the "excess capacity" doctrine is essentially very different from ours and based on far more "orthodox" considerations. Dewey, *Monopoly and Economics in Law*, pp. 96 ff.

⁸³Since the erroneous but popular theory of "countervailing power," propounded by J.K. Galbraith, falls with the monopolistic-competition theory, it is unnecessary to discuss it here. For a more detailed critique of its numerous fallacies, see Simon N. Whitney, "Errors in the Concept of Countervailing Power," *Journal of Business*, October, 1953, pp. 238-53; George J. Stigler, "The Economist Plays with Blocs," *American Economic Review, Papers and Proceedings*, May, 1954, pp. 8-14; and David McCord Wright, "Discussion," *ibid.*, pp. 26-30.

⁸⁴Chamberlin, *Theory of Monopolistic Competition*, pp. 123 ff. Chamberlin includes in selling costs advertising, sales expenses, and store displays.

costs,” on the other hand, are supposed to be directed toward influencing consumers and *increasing* their demand schedules for the firm’s product.

This distinction is completely spurious.⁸⁵ Why does a businessman invest money and incur *any* costs whatever? To supply a hoped-for demand for his product. Every time he improves his product he is hoping that consumers will respond by increasing their demands. In fact, *all* costs expended on raw materials are incurred in an attempt to increase consumer demand beyond what it would have been in the absence of these costs. *Therefore, every production cost is also a “selling cost.”*

Conversely, selling costs are not the sheer waste or even tyranny that monopolistic-competition theorists have usually assumed. The various expenses designated as “selling costs” perform definite services for the public. Basically, they furnish information to the public about the goods of the seller. We live in a world where there can be no “perfect knowledge” of products by anyone—especially consumers, who are faced with a myriad of available products. Selling costs are therefore important in providing information about the product as well as about the firm. In some cases, e.g., displays, the “selling cost” itself *directly* improves the quality of the product in the mind of the consumer. It must always be remembered that the consumer is not simply buying a *physical product*; he may also be buying “atmosphere,” prestige, service, etc., all of which have tangible reality to him and are valued accordingly.⁸⁶

The view that a selling cost is somehow an artifact of “monopolistic competition” stems only from the peculiar assumptions of “pure competition.” In the “ideal” world of pure

⁸⁵See Mises, *Human Action*, p. 319. Also see Kermit Gordon, “Concepts of Competition and Monopoly—Discussion,” *American Economic Review, Papers and Proceedings*, May, 1955, pp. 486–87.

⁸⁶It is surely highly artificial to call bright ribbons on a packaged good a “production cost,” while labeling bright ribbons decorating the store selling the good as a “selling cost.”

competition, we remember, each firm's demand is *given* to it as infinitely elastic, so that it can sell whatever it wants at the ruling price. Naturally, in such a situation, no selling costs are necessary, because a market for a product is automatically assured. In the real world, however, there is no perfect knowledge, and the demand curves are neither given nor infinitely elastic.⁸⁷ Therefore, firms have to try to increase demands for their products and to carve out market areas for themselves.

Chamberlin falls into another error in implying that selling costs, such as advertising, "create" consumer demands. This is the determinist fallacy. Every man as a self-owner freely decides his own scale of valuations. On the free market no one can force another to choose his product. And no other individual can ever "create" someone's values for him; he must adopt the value himself.⁸⁸

⁸⁷Cf. Alfred Nicols, "The Development of Monopolistic Competition and the Monopoly Problem," *Review of Economics and Statistics*, May, 1949, pp. 118–23.

⁸⁸ See Mises:

The consumer is, according to . . . legend, simply defenseless against "high-pressure" advertising. If this were true, success or failure in business would depend on the mode of advertising only. However, nobody believes that any kind of advertising would have succeeded in making the candlemakers hold the field against the electric bulb, the horsedriers against the motorcars. . . . But this implies that the quality of the commodity advertised is instrumental in bringing about the success of an advertising campaign. . . . The tricks and artifices of advertising are available to the seller of the better product no less than to the seller of the poorer product. But only the former enjoys the advantages derived from the better quality of his product. (Mises, *Human Action*, pp. 317–18)

6. Multiform Prices and Monopoly

Up to this point we have always concluded that the market tends, at any given time, to establish *one uniform market price* for any good, under competitive or monopoly conditions. One phenomenon that sometimes appears, however, is persistent *multiformity* of prices. (We must consider, of course, a good that is really homogeneous; otherwise, there would merely be price differences for different goods.) How, then, can multiformity come about, and does it in some sense violate the workings or the ethics of a free-market society?

We must first separate goods into two kinds: those that are *resalable* and those that are not. Under the latter category come all intangible services, which are either consumed directly or used up in the process of production; in any case, they themselves cannot be resold by the first buyer. Nonresalable services also include the *rental* use of a tangible good, for then the good itself is not being bought, but rather its unit services over a period of time. An example may be the “renting” of space in a freight car.

Let us first take resalable goods. When can there be persistent multiform pricing of such goods? One necessary condition is clearly *ignorance* on the part of some seller or buyer. The market price for a certain kind of steel, for example, may be one gold ounce per ton; but one seller, out of pure ignorance, may persist in selling it for half a gold ounce per ton. What will happen? In the first place, some enterprising person will buy the steel from this laggard and resell it at the market price, thus establishing effective uniformity. Secondly, other buyers will rush to outbid the first buyer for the bargain, thus informing the seller of his underpricing. Finally, the persistently ignorant seller will not long remain in business. (Of course, it may happen that the seller may have a strong desire to sell steel for lower than market price, for “philanthropic” reasons. But if he persists in doing so, then he is simply purchasing the consumers’ good—to him—of philanthropy and paying the price for it in lower revenue. He is here acting as a consumer rather

than as an entrepreneur, just as he would if he hired his ne'er-do-well nephew at the expense of a cut in profits. This, then, would not be a genuine case of multiform pricing, where the good must *always* be homogeneous.)

Nor is the buyer in a different condition. If a buyer were ignorant and continued to buy steel at two gold ounces a ton when the market price was one gold ounce, then some other seller would soon apprise the buyer of his error by offering to sell him the steel for much less. If there is only one seller, then the cheaper buyer can still resell at a profit to the buyer charged a higher price. And a persistently ignorant buyer will also go out of business.

There is only one case where a multiform price could possibly be established for a resalable good: where the good is being sold to consumers—the ultimate buyers. For while entrepreneurial buyers will be alert to price differentials, and a buyer of a good at a lower price can resell to another buyer charged a higher price, ultimate consumers do not usually consider reselling once they buy. A classic case is that of American tourists at a Middle Eastern bazaar.⁸⁹ The tourist has neither the time nor the inclination to make a thorough study of the consumer markets, and therefore each tourist is ignorant of the going price of any good. Hence, the seller can isolate each buyer, charging highest prices to the most eager buyers, less high prices to the next most eager, and much lower prices to the marginal buyers, of the same good. In that way the seller achieves a generally unfulfilled objective of all sellers: the tapping of more of the “consumers’ surplus” of the buyers. Here the two conditions are fulfilled: the consumers are ignorant of the going price and are not in the market to resell.

Does multiform pricing, as has often been charged, distort the structure of production, and is it in some way immoral or exploitative? How is it immoral? The seller aims, as always, to

⁸⁹See Wicksteed, *Common Sense of Political Economy and Selected Papers*, I, 253 ff.

maximize his earnings in voluntary exchange, and he certainly cannot be held responsible for the ignorance of the buyer. If buyers do not take the trouble to inform themselves of the state of the market, they must stand prepared to have some of their psychic surplus tapped by the bargaining of the seller. Neither is this action irrational on the part of the buyer. For we must deduce from the buyer's action that *he prefers to remain in ignorance rather than to make the effort or pay the money to inform himself of market conditions*. To acquire knowledge of any field takes time, effort, and often money, and it is perfectly reasonable for an individual on any given market to prefer to take his chances on the price and use his scarce resources in other directions. This choice is crystal clear in the case of a tourist on holiday, but it is also possible in any other given market. Both the impatient tourist, who prefers to pay a higher price and not spend time and money on learning about the market, and a companion who spends days on an intensive study of the bazaar market are exercising their preferences, and praxeology cannot call one or the other more rational. Furthermore, there is no way to measure the consumer surpluses lost or gained in the case of the two tourists. We must therefore conclude that multiform pricing, in the case of resalable goods, does not at all distort the allocation of productive factors, because, on the contrary, it is consistent with, and in the case of the tourist, the *only* pricing consistent with, the satisfaction of consumer preferences.

It must be emphasized here that no matter how much the seller at the bazaar taps of his customers' psychic surplus, he does not tap it *all*; otherwise the sale would not be made at all. Since the exchange is voluntary, both parties still benefit from making it.

What if the good is *not* resalable? In that case, there is far greater room for multiform pricing, since ignorance is not required. A vendor can sell an intangible service at a higher price to A than to B without fear that B can undercut him by

reselling to A. Hence, most actual cases of multiform pricing take place in the realm of intangible goods.

Suppose now that seller X has managed to establish multiform prices for his customers. He might be a lawyer, for example, who charges higher fees for the same service to a wealthy than to a poor client. Since there is still competition among sellers, why does another lawyer Y not enter the field and undercut X's price to the wealthy clients? In fact, this is what will generally happen, and any attempt to establish "separate markets" among customers will lead to an invasion of the more profitable, higher-price field by other competitors, finally driving the price down, reducing revenues, and re-establishing uniform pricing. If a seller's service is unusual and it is universally recognized that he has no effective competitors, then he might be able to sustain a multiform structure.

There is one simple but very important condition that we have not mentioned which must be fulfilled to establish multiform pricing: the total proceeds from multiformity must be greater than from uniformity. Where one buyer can buy only one unit of a good, this is no problem. If there is and can be only one seller of a nonresalable good, *and* each buyer can buy no more than one unit, then multiform pricing will tend to be established (barring undercutting by competitors), since the total revenue to the seller will always be greater through tapping more of the consumer surpluses of each buyer.⁹⁰ But if a buyer can buy more than one unit, revenue becomes a problem. For then each buyer, confronted with a higher price, will restrict his purchases. This will leave an unsold stock, which the seller will

⁹⁰It is difficult to conceive of a case, in reality, to which such a restriction imposed on buyers (called "perfect price-discrimination") would apply. Mrs. Robinson cites as an example a ransom charged by a kidnaper, but this, of course does not obtain on the free, unhampered market, which precludes kidnapping. Robinson, *Economics of Imperfect Competition*, p. 187 n.

then unload by *lowering* his prices below the hypothetical uniform price in order to tap the demands of hitherto submarginal buyers. Thus, suppose that the uniform price of a good is ten gold grains per unit, at which a hundred units are sold. The seller now decides to isolate each buyer as a separate market and tap more consumer surpluses. Aside from the barely marginal buyers, then, all the others will find their prices raised. They will restrict their purchases, say to an aggregate of eighty-five units, and the other fifteen units will be sold by lowering the price to new, hitherto submarginal buyers.

Multiformity can be established only when total proceeds are greater than uniformity provides. This is by no means always the case, for the supramarginal buyers may restrict their purchases by more than the submarginal buyers can compensate.⁹¹

Multiform pricing has been accorded a curious reception by economists and laymen. In some cases it is deemed vicious exploitation of the consumers; in others (e.g., medicine and education) it is considered praiseworthy and humanitarian. In reality, it is neither. It is certainly not the *rule* in pricing that the most eager *should* pay in proportion to their eagerness (in practice, usually gauged by their wealth), for then everyone would pay in proportion to his wealth for everything, and the entire monetary and economic system would break down; money would no longer function. (See chapter 12 below.) If this is clear in general, it is difficult to see a priori why specific goods should be singled out for this treatment. On the other hand, the consumers are not being “exploited” if there is multiformity. It is clear that the marginal and submarginal buyers are not exploited: the latter obviously gain. What of the *supramarginal* buyers who are receiving less consumer surplus? In some cases, they gain, because without the greater revenues provided by “price discrimination” the good would not be supplied at all. Consider, for example, a country doctor who would leave the

⁹¹See Mises, *Human Action*, pp. 385 ff.

area if he had to subsist on the lower revenues provided by uniformity. And even if the good were still supplied, *the fact* that the supramarginal buyers continue to patronize the seller at all shows that they are content with the seemingly discriminatory arrangement. Otherwise, they would quickly boycott the seller, either individually or in concert, and patronize competitors. They would simply refuse to pay more than the submarginal buyers, and this would quickly induce the seller to lower his prices. The fact that they do not do so shows that they *prefer* multiformity to uniformity in the particular case. An example is private school education, which able but poor youths may often attend on scholarships—a principle that the wealthy parents who pay full tuition demonstrably do not consider unjust. If, however, the sellers have received grants of monopolistic privilege by the government, enabling them to restrict competition in the serving of the supramarginal buyers, then they may establish multiformity without enjoying the demonstrable preference of these buyers: for here governmental coercion has entered to inhibit the free expression of preferences.⁹²

So far we have discussed price discrimination by sellers in consumers' markets, where consumer surpluses are tapped. Can there be such discrimination in producers' markets? Only when the good is not resalable, total proceeds are greater under multiformity, *and* the supramarginal buyers are willing to pay. The latter will happen when these buyers have a higher DMVP for the good in their firms than other buyers have in theirs. In this case, the seller of the good with multiform prices is absorbing a rent formerly earned by the supramarginal buying firm. The most notable case of such pricing has been railroad freight "discrimination against" the firms shipping a cargo

⁹²An example is medicine, where the government helps to restrict the supply and thus to prevent price-cutting. See the illuminating article by Reuben A. Kessel, "Price Discrimination in Medicine," *The Journal of Law and Economics*, October, 1958, pp. 20–53. Also see chapter 12 below on grants of monopoly privilege.

more valuable per unit weight than that of other firms. The gains are not, of course, retained by the railroad in the long run, but absorbed by its own land and labor factors.

Can there be price discrimination by *buyers* when the good is not resalable (and ignorance among sellers is not assumed)? No, there cannot, for the minimum reserve price imposed by, say, a laborer, is determined by the opportunity cost he has foregone elsewhere. In short, if a man earns five gold ounces a week for his labor service in firm A, he will not accept two ounces a week (although he would take two rather than earn nothing at all) since he can earn nearly five ounces somewhere else. And the meaning of price discrimination against sellers is that a buyer would be able to pay less for the same good than the seller can earn elsewhere (cost of moving, etc., omitted). Hence, there can be no price discrimination against sellers. If sellers are ignorant, then, as in the case of the ignorant consumers at a bazaar, we must infer that they prefer the lower income to the cost and trouble of learning more about the market.

7. Patents and Copyrights

Turning now to patents and copyrights, we ask: Which of the two, if either, is consonant with the purely free market, and which is a grant of monopoly privilege by the State? In this part, we have been analyzing the economics of the purely free market, where the individual person and property are not subject to molestation. It is therefore important to decide whether patents or copyrights will obtain in the purely free, noninvasive society, or whether they are a function of government interference.

Almost all writers have bracketed patents and copyrights together. Most have considered both as grants of exclusive monopoly privilege by the State; a few have considered both as part and parcel of property right on the free market. But almost everyone has considered patents and copyrights as equivalent: the one as conferring an exclusive property right in the field of mechanical inventions, the other as conferring an exclusive

right in the field of literary creations.⁹³ Yet this bracketing of patents and copyrights is wholly fallacious; the two are completely different in relation to the free market.

It is true that a patent and a copyright are both exclusive property rights and it is also true that they are both property rights in *innovations*. But there is a crucial difference in their *legal enforcement*. If an author or a composer believes his copyright is being infringed, and he takes legal action, he must “prove that the defendant had ‘access’ to the work allegedly infringed. If the defendant produces something identical with the plaintiff’s work by mere chance, there is no infringement.”⁹⁴ Copyrights, in other words, have their basis in prosecution of implicit theft. The plaintiff must prove that the defendant stole the former’s creation by reproducing it and selling it himself in violation of his or someone else’s contract with the original seller. But if the defendant independently arrives at the same creation, the plaintiff has no copyright privilege that could prevent the defendant from using and selling his product.

Patents, on the other hand, are completely different. Thus:

You have patented your invention and you read in the newspaper one clay that John Doe, who lives in a city 2,000 miles from your town, has invented an identical or similar device, that he has licensed the EZ company to manufacture it. . . . Neither Doe nor the EZ company . . . ever heard of your invention. All believe Doe to be the inventor of a new and original device. They may all be guilty of infringing your patent . . . the fact that their infringement was in ignorance of the true facts and unintentional will not constitute a defense.⁹⁵

⁹³Henry George was a notable exception. See his excellent discussion in *Progress and Poverty* (New York: Modern Library, 1929), p. 411 n.

⁹⁴Richard Wincor, *How to Secure Copyright* (New York: Oceana Publishers, 1950), p. 37.

⁹⁵Irving Mandell, *How to Protect and Patent Your Invention* (New York: Oceana Publishers, 1951), p. 34.

Patent, then, has nothing to do with implicit theft. It confers an exclusive privilege on the first inventor, and if anyone else should, quite independently, invent the same or similar machine or product, the latter would be debarred by violence from using it in production.

We have seen in chapter 2 that the acid test by which we judge whether or not a certain practice or law is or is not consonant with the free market is this: Is the outlawed practice implicit or explicit theft? If it is, then the free market would outlaw it; if not, then its outlawry is itself government interference in the free market. Let us consider copyright. A man writes a book or composes music. When he publishes the book or sheet of music, he imprints on the first page the word "copyright." This indicates that any man who agrees to purchase this product also agrees as part of the exchange *not* to recopy or reproduce this work for sale. In other words, the author does not sell his property outright to the buyer; he sells it *on condition* that the buyer not reproduce it for sale. Since the buyer does not buy the property outright, but only on this condition, any infringement of the contract by him or a subsequent buyer is *implicit theft* and would be treated accordingly on the free market. The copyright is therefore a logical device of property right on the free market.

Part of the patent protection now obtained by an inventor could be achieved on the free market by a type of "copyright" protection. Thus, inventors must now *mark* their machines as being patented. The mark puts the buyers on notice that the invention is patented and that they cannot sell that article. But the same could be done to extend the copyright system, and without patent. In the purely free market, the inventor could mark his machine *copyright*, and then anyone who buys the machine buys it *on the condition* that he will not reproduce and sell such a machine for profit. Any violation of this contract would constitute implicit theft and be prosecuted accordingly on the free market.

The patent is incompatible with the free market *precisely to the extent that it goes beyond the copyright*. The man who has not

bought a machine and who arrives at the same invention independently, will, on the free market, be perfectly able to use and sell his invention. Patents prevent a man from using his invention even though all the property is his and he has not stolen the invention, either explicitly or implicitly, from the first inventor. Patents, therefore, are grants of exclusive monopoly privilege by the State and are *invasive* of property rights on the market.

The crucial distinction between patents and copyrights, then, is not that one is mechanical and the other literary. The fact that they have been applied that way is an historical accident and does not reveal the critical difference between them.⁹⁶ The crucial difference is that copyright is a logical attribute of property right on the free market, while patent is a monopoly invasion of that right.

The application of patents to mechanical inventions and copyrights to literary works is peculiarly inappropriate. It would be more in keeping with the free market to be just the reverse. For literary creations are unique products of the individual; it is almost impossible for them to be independently duplicated by someone else. Therefore, a *patent*, instead of a copyright, for literary productions would make little difference in practice. On the other hand, mechanical inventions are discoveries of natural law rather than individual creations, and hence similar independent inventions occur all the time.⁹⁷ The

⁹⁶This can be seen in the field of *designs*, which can be either copyrighted or patented.

⁹⁷For a legal hint on the proper distinction between copyright and monopoly, see F.E. Skone James, "Copyright" in *Encyclopedia Britannica* (14th ed.; London, 1929), VI, 415–16. For the views of nineteenth-century economists on patents, see Fritz Machlup and Edith T. Penrose, "The Patent Controversy in the Nineteenth Century," *Journal of Economic History*, May, 1950, pp. 1–29. Also see Fritz Machlup, *An Economic Review of the Patent System* (Washington, D.C.: United States Government Printing Office, 1958).

simultaneity of inventions is a familiar historical fact. Hence, if it is desired to maintain a free market, it is particularly important to allow *copyrights*, but not patents, for mechanical inventions.

The common law has often been a good guide to the law consonant with the free market. Hence, it is not surprising that common-law copyright prevails for *unpublished* literary manuscripts, while there is no such thing as a common-law *patent*. At common law, the inventor also has the right to keep his invention unpublicized and safe from theft, i.e., he has the equivalent of the copyright protection for unpublicized inventions.

On the free market, there would therefore be no such thing as patents. There would, however, be copyright for any inventor or creator who made use of it, and this copyright would be *perpetual*, not limited to a certain number of years. Obviously, to be fully the property of an individual, a good has to be permanently and perpetually the property of the man and his heirs and assigns. If the State decrees that a man's property ceases at a certain date, this means that the *State* is the real owner and that it simply grants the man use of the property for a certain period of time.⁹⁸

Some defenders of patents assert that they are not monopoly privileges, but simply property rights in inventions or even in "ideas." But, as we have seen, everyone's property right is defended in libertarian law without a patent. If someone has an idea or plan and constructs an invention, and it is stolen from his house, the stealing is an act of theft illegal under general law. On the other hand, patents actually invade the property rights of those *independent* discoverers of an idea or invention who made the discovery after the patentee. Patents, therefore, *invade* rather than defend property rights. The speciousness of this

⁹⁸Of course, there would be nothing to prevent the creator or his heirs from voluntarily abandoning this property right and throwing it into the "public domain" if they so desired.

argument that patents protect property rights in ideas is demonstrated by the fact that not all, but only certain types of original ideas, certain types of innovations, are considered patentable.

Another common argument for patents is that “society” is simply making a contract with the inventor to purchase his secret, so that “society” will have use of it. In the first place, “society” could pay a straight subsidy, or price, to the inventor; it would not have to prevent all later inventors from marketing *their* inventions in this field. Secondly, there is nothing in the free economy to prevent any individual or group of individuals from purchasing secret inventions from their creators. No monopolistic patent is necessary.

The most popular argument for patents among economists is the utilitarian one that a patent for a certain number of years is necessary to encourage a sufficient amount of research expenditure for inventions and innovations in processes and products.

This is a curious argument, because the question immediately arises: By what standard do you judge that research expenditures are “too much,” “too little,” or just about enough? This is a problem faced by *every* governmental intervention in the market’s production. Resources—the better lands, laborers, capital goods, time—in society are limited, and they may be used for countless alternative ends. By what standard does someone assert that certain uses are “excessive,” that certain uses are “insufficient,” etc.? Someone observes that there is little investment in Arizona, but a great deal in Pennsylvania; he indignantly asserts that Arizona deserves more investment. But what standards can he use to make this claim? The *market does* have a rational standard: the highest money incomes and highest profits, for these can be achieved only through maximum service of consumer desires. This principle of maximum service to consumers and producers alike—i.e., to everybody—governs the seemingly mysterious market allocation of resources: how much to devote to one firm or to another, to one area or another, to present or future, to one good or another, to research

as compared with other forms of investment. But the observer who criticizes this allocation can have no rational standards for decision; he has only his arbitrary whim. This is especially true of criticism of *production*-relations. Someone who chides *consumers* for buying too much cosmetics may have, rightly or wrongly, some rational basis for his criticism. But someone who thinks that more or less of a certain resource should be used in a certain manner or that business firms are “too large” or “too small” or that too much or too little is spent on research or is invested in a new machine, can have no rational basis for his criticism. Businesses, in short, are producing for a market, guided by the ultimate valuations of consumers on that market. Outside observers may criticize ultimate valuations of consumers if they choose—although if they interfere with consumption based on these valuations they impose a loss of utility upon consumers—but they cannot legitimately criticize the *means*: the production relations, the allocations of factors, etc., by which these ends are served.

Capital funds are limited, and they must be allocated to various uses, one of which is research expenditures. On the market, rational decisions are made in setting research expenditures, in accordance with the best entrepreneurial expectations of an uncertain future. Coercively to encourage research expenditures would distort and hamper the satisfaction of consumers and producers on the market.

Many advocates of patents believe that the ordinary competitive conditions of the market do not sufficiently encourage the adoption of new processes and that therefore innovations must be coercively promoted by the government. But the market decides on the rate of introduction of new processes just as it decides on the rate of industrialization of a new geographic area. In fact, this argument for patents is very similar to the infant-industry argument for tariffs—that market processes are not sufficient to permit the introduction of worthwhile new processes. And the answer to both these arguments is the same: that people must balance the superior productivity of

the new processes against the cost of installing them, i.e., against the advantage possessed by the old process in being already built and in existence. Coercively privileging innovation would needlessly scrap valuable plants already in existence and impose an excessive burden upon consumers. For consumers' desires would not be satisfied in the most economic manner.

It is by no means self-evident that patents encourage an increased absolute quantity of research expenditures. But certainly patents distort the *type* of research expenditure being conducted. For while it is true that the *first* discoverer benefits from the privilege, it is also true that his competitors are excluded from production in the area of the patent for many years. And since one patent can build upon a related one in the same field, competitors can often be indefinitely discouraged from *further* research expenditures in the general area covered by the patent. Moreover, the patentee is himself discouraged from engaging in further research in this field, for the privilege permits him to rest on his laurels for the entire period of the patent, with the assurance that no competitor can trespass on his domain. The competitive spur for further research is eliminated. Research expenditures are therefore *overstimulated* in the early stages before anyone has a patent, and they are *unduly restricted* in the period after the patent is received. In addition, some inventions are considered patentable, while others are not. The patent system then has the further effect of artificially stimulating research expenditures in the *patentable* areas, while artificially restricting research in the *nonpatentable* areas.

Manufacturers have by no means unanimously favored patents. R.A. Macfie, leader of England's flourishing patent-abolition movement during the nineteenth century, was president of the Liverpool Chamber of Commerce.⁹⁹ Manufacturer I.K. Brunel, before a committee of the House of Lords,

⁹⁹See the illuminating article by Machlup and Penrose, "Patent Controversy in the Nineteenth Century," pp. 1-29.

deplored the effect of patents in stimulating wasteful expenditure of resources on searching for untried patentable inventions, resources that could have been better used in production. And Austin Robinson has pointed out that many industries get along without patents:

In practice the enforcement of patent monopolies is often so difficult . . . that competing manufacturers have in some industries preferred to pool patents; and to look for sufficient reward for technical invention in the . . . advantage of priority that earlier experimentation usually gives and in the subsequent good-will that may arise from it.¹⁰⁰

As Arnold Plant summed up the problem of competitive research expenditures and innovations:

Neither can it be assumed that inventors would cease to be employed if entrepreneurs lost the monopoly over the use of their inventions. Businesses employ them today for the production of nonpatentable inventions, and they do not do so merely for the profit which priority secures. In active competition . . . no business can afford to lag behind its competitors. The reputation of a firm depends upon its ability to keep ahead, to be first in the market with new improvements in its products and new reductions in their prices.¹⁰¹

Finally, of course, the market itself provides an easy and effective course for those who feel that there are not enough expenditures being made in certain directions. *They can make these expenditures themselves.* Those who would like to see more inventions made and exploited, therefore, are at liberty to join

¹⁰⁰Cited in Edith Penrose, *Economics of the International Patent System* (Baltimore: Johns Hopkins Press, 1951), p. 36; see also *ibid.*, pp. 19–41.

¹⁰¹Arnold Plant, "The Economic Theory concerning Patents for Inventions," *Economica*, February, 1934, p. 44.

together and subsidize such effort in any way they think best. In that way, they would, as consumers, add resources to the research and invention business. And they would not then be forcing other consumers to lose utility by conferring monopoly grants and distorting the market's allocations. Their voluntary expenditures would become *part of the market* and express ultimate consumer valuations. Furthermore, later inventors would not be restricted. The friends of invention could accomplish their aim without calling in the State and imposing losses on a large number of people.

MONEY AND ITS PURCHASING POWER

1. Introduction

MONEY HAS ENTERED INTO ALMOST all our discussion so far. In chapter 3 we saw how the economy evolved from barter to indirect exchange. We saw the patterns of indirect exchange and the types of allocations of income and expenditure that are made in a monetary economy. In chapter 4 we discussed money prices and their formation, analyzed the marginal utility of money, and demonstrated how monetary theory can be subsumed under utility theory by means of the money regression theorem. In chapter 6 we saw how monetary calculation in markets is essential to a complex, developed economy, and we analyzed the structure of post-income and pre-income demands for and supplies of money on the time market. And from chapter 2 on, all our discussion has dealt with a monetary-exchange economy.

The time has come to draw the threads of our analysis of the market together by completing our study of money and of the effects of changes in monetary relations on the economic system. In this chapter we shall continue to conduct the analysis within the framework of the free-market economy.

2. *The Money Relation: The Demand for and the Supply of Money*

Money is a commodity that serves as a general medium of exchange; its exchanges therefore permeate the economic system. Like all commodities, it has a market demand and a market supply, although its special situation lends it many unique features. We saw in chapter 4 that its “price” has no unique expression on the market. Other commodities are all expressible in terms of units of money and therefore have uniquely identifiable prices. The money commodity, however, can be expressed only by an array of all the other commodities, i.e., all the goods and services that money can buy on the market. This array has no uniquely expressible unit, and, as we shall see, changes in the array cannot be measured. Yet the concept of the “price” or the “value” of money, or the “purchasing power of the monetary unit,” is no less real and important for all that. It simply must be borne in mind that, as we saw in chapter 4, there is no single “price level” or measurable unit by which the value-array of money can be expressed. This exchange-value of money also takes on peculiar importance because, unlike other commodities, the prime purpose of the money commodity is to be exchanged, now or in the future, for directly consumable or productive commodities.

The *total demand for money* on the market consists of two parts: the *exchange demand for money* (by sellers of all other goods that wish to purchase money) and the *reservation demand for money* (the demand for money to hold by those who already hold it). Because money is a commodity that permeates the market and is continually being supplied and demanded by everyone, and because the proportion which the existing stock of money bears to new production is high, it will be convenient to analyze the supply of and the demand for money in terms of the *total demand-stock* analysis set forth in chapter 2.¹

¹Cf. Edwin Cannan, “The Application of the Theoretical Analysis of Supply and Demand to Units of Currency” in F.A. Lutz and L.W. Mints,

In contrast to other commodities, everyone on the market has both an exchange demand and a reservation demand for money. The exchange demand is his *pre-income demand* (see chapter 6, above). As a seller of labor, land, capital goods, or consumers' goods, he must supply these goods and demand money in exchange to obtain a money income. Aside from speculative considerations, the seller of ready-made goods will tend, as we have seen, to have a perfectly inelastic (vertical) supply curve, since he has no reservation uses for the good. But the supply curve of a good for money is equivalent to a (partial) demand curve for money in terms of the good to be supplied. Therefore, the (exchange) demand curves for money in terms of land, capital goods, and consumers' goods will tend to be perfectly inelastic.

For labor services, the situation is more complicated. Labor, as we have seen, does have a reserved use—satisfying leisure. We have seen that the general supply curve of a labor factor can be either “forward-sloping” or “backward-sloping,” depending upon the individuals' marginal utility of money and marginal disutility of leisure forgone. In determining labor's demand curve for money, however, we can be far more certain. To understand why, let us take a hypothetical example of a supply curve of a labor factor (in general use). At a wage rate of five gold grains an hour, 40 hours per week of labor service will be sold. Now suppose that the wage rate is raised to eight gold grains an hour. Some people might work a greater number of hours because they have a greater monetary inducement to sacrifice leisure for labor. They might work 50 hours per week. Others may decide that the increased income permits them to sacrifice some money and take some of the increased earnings in greater leisure. They might work 30 hours. The first would

eds., *Readings in Monetary Theory* (Philadelphia: Blakiston, 1951), pp. 3–12, and Cannan, *Money* (6th ed.; London: Staples Press, 1929), pp. 10–19, 65–78.

represent a “forward-sloping,” the latter a “backward-sloping,” supply curve of labor in this price range. But both would have one thing in common. Let us multiply hours by wage rate in each case, to arrive at the total money income of the laborers in the various situations. In the original case, a laborer earned 40 times 5 or 200 gold grains per week. The man with a backward-sloping supply curve will earn 30 times 8 or 240 gold grains a week. The one with a forward-sloping supply curve will earn 50 times 8 or 400 gold grains per week. In both cases, *the man earns more money at the higher wage rate.*

This will always be true. In the first case, it is obvious, for the higher wage rate induces the man to sell more labor. But it is true in the latter case as well. For the higher money income permits a man to gratify his desires for more leisure as well, precisely *because* he is getting an *increased money income*. Therefore, a man’s backward-sloping supply curve will never be “backward” enough to make him earn *less money* at higher wage rates.

Thus, a man will always earn more money at a higher wage rate, less money at a lower. But what is *earning* money but another name for *buying* money? And that is precisely what is done. People *buy* money by selling goods and services that they possess or can create. We are now attempting to arrive at the demand schedule for money in relation to various alternative purchasing powers or “exchange-values” of money. A lower exchange-value of money is equivalent to higher goods-prices in terms of money. Conversely, a higher exchange-value of money is equivalent to lower prices of goods. In the labor market, a higher exchange-value of money is translated into lower wage rates, and a lower exchange-value of money into higher wage rates.

Hence, on the labor market, our law may be translated into the following terms: *The higher the exchange-value of money, the lower the quantity of money demanded; the lower the exchange-value of money, the higher the quantity of money demanded* (i.e., the lower the wage rate, the less money earned; the higher the wage rate,

the more money earned). Therefore, on the labor market, the demand-for-money schedule is *not* vertical, but falling, when the exchange-value of money increases, as in the case of any demand curve.

Adding the vertical demand curves for money in the other exchange markets to the falling demand curve in the labor market, we arrive at a falling exchange-demand curve for money.

More important, because more volatile, in the total demand for money on the market is the *reservation* demand to hold money. This is everyone's *post-income* demand. After everyone has acquired his income, he must decide, as we have seen, between the allocation of his money assets in three directions: consumption spending, investment spending, and addition to his cash balance ("net hoarding"). Furthermore, he has the additional choice of subtraction from his cash balance ("net dishoarding"). How much he decides to retain in his cash balance is uniquely determined by the marginal utility of money in his cash balance on his value scale. Until now we have discussed at length the sources of the utilities and demands for consumers' goods and for producers' goods. We have now to look at the remaining good: *money in the cash balance*, its utility and demand.

Before discussing the sources of the demand for a cash balance, however, we may determine the shape of the reservation (or "cash balance") demand curve for money. Let us suppose that a man's marginal utilities are such that he wishes to have 10 ounces of money held in his cash balance over a certain period. Suppose now that the exchange-value of money, i.e., the purchasing power of a monetary unit, increases, other things being equal. This means that his 10 gold ounces accomplish more work than they did before the change in the PPM (purchasing power of the monetary unit). As a consequence, he will tend to remove part of the 10 ounces from his cash balance and spend it on goods, the prices of which have now fallen. Therefore, *the higher the PPM (the exchange-value of money), the*

lower the quantity of money demanded in the cash balance. Conversely, a lower PPM will mean that the previous cash balance is worth less in real terms than it was before, while the higher prices of goods discourage their purchase. As a result, the lower the PPM, the higher the quantity of money demanded in the cash balance.

As a result, *the reservation demand curve for money in the cash balance falls as the exchange-value of money increases.* This falling demand curve, added to the falling exchange-demand curve for money, yields the market's *total demand curve for money*—also falling in the familiar fashion for every commodity.

There is a third demand curve for the money commodity that deserves mention. This is the demand for *nonmonetary uses* of the monetary metal. This will be relatively unimportant in the advanced monetary economy, but it will exist nevertheless. In the case of gold, this will mean either uses in consumption, as for ornaments, or productive uses, as for industrial purposes. At any rate, this demand curve *also* falls as the PPM increases. As the “price” of money (PPM) increases, more goods can be obtained through expenditure of a unit of money; as a result, the opportunity-cost in using gold for nonmonetary purposes increases, and less is demanded for that purpose. Conversely, as the PPM falls, there is more incentive to use gold for its direct use. This demand curve is added to the total demand curve for money, to obtain the total demand curve for the money commodity.²

At any one time there is a *given total stock* of the money commodity. This stock will, at any time, be *owned by someone*. It is therefore dangerously misleading to adopt the custom of American economists since Irving Fisher's day of treating money as somehow “circulating,” or worse still, as divided into “circulating money” and “idle money.”³ This concept conjures up the image

²From this point on, this nonmonetary demand is included, for convenience, in the “total demand for money.”

³Cf. Irving Fisher, *The Purchasing Power of Money* (2nd ed.; New York: Macmillan & Co., 1913).

of the former as moving somewhere at all times, while the latter sits idly in “hoards.” This is a grave error. There is, actually, no such thing as “circulation,” and there is no mysterious arena where money “moves.” At any one time all the money is owned by someone, i.e., rests in someone’s cash balance. Whatever the stock of money, therefore, people’s actions must bring it into accord with the total demand for money to hold, i.e., the total demand for money that we have just discussed. For even pre-income money acquired in exchange must be held at least momentarily in one’s cash balance before being transferred to someone else’s balance. All total demand is therefore to hold, and this is in accord with our analysis of total demand in chapter 2.

Total stock must therefore be brought into agreement, on the market, with the total quantity of money demanded. The diagram of this situation is shown in Figure 74.

On the vertical axis is the PPM, increasing upward. On the horizontal axis is the quantity of money, increasing rightwards. D_e is the aggregate exchange-demand curve for money, falling

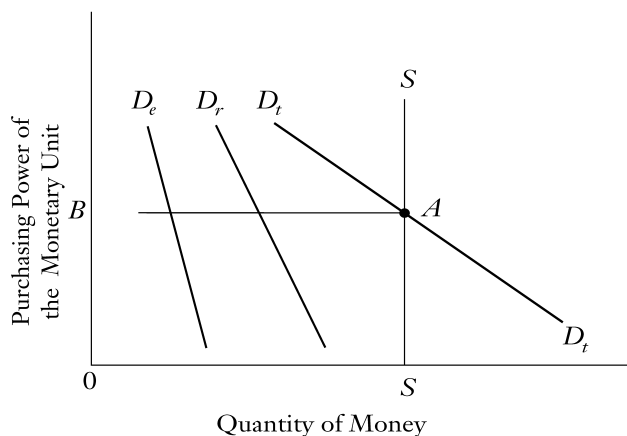


FIGURE 74. DETERMINATION OF THE EQUILIBRIUM POINT FOR THE EXCHANGE-VALUE OF MONEY

and inelastic. D_r is the reservation or cash-balance demand for money. D_t is the total demand for money to hold (the demand for nonmonetary gold being omitted for purposes of convenience). Somewhere intersecting the D_t curve is the SS vertical line—the total stock of money in the community—given at quantity OS.

The intersection of the latter two curves determines the equilibrium point, A , for the exchange-value of money in the community. The exchange-value, or PPM, will be set at OB .

Suppose now that the PPM is slightly higher than OB . The demand for money at that point will be less than the stock. People will become unwilling to hold money at that exchange-value and will be anxious to sell it for other goods. These sales will raise the prices of goods and lower the PPM, until the equilibrium point is reached. On the other hand, suppose that the PPM is lower than OB . In that case, more people will demand money, in exchange or in reservation, than there is money stock available. The consequent excess of demand over supply will raise the PPM again to OB .

3. *Changes in the Money Relation*

The purchasing power of money is therefore determined by two factors: *the total demand schedule for money to hold* and *the stock of money in existence*. It is easy to see on a diagram what happens when either of these determining elements changes. Thus, suppose that the schedule of total demand increases (shifts to the right). Then (see Figure 75) the total-demand-for-money curve has shifted from $D_t D_t$ to $D'_t D'_t$. At the previous equilibrium PPM point, A , the demand for money now exceeds the stock available by AE . The bids push the PPM upwards until it reaches the equilibrium point C . The converse will be true for a shift of the total demand curve leftward—a decline in the total demand schedule. Then, the PPM will fall accordingly.

The effect of a change in the total stock, the demand curve remaining constant, is shown in Figure 76. Total quantity of