The Yale Law Journal
Volume 76, No. 2, December 1966

Legal Restrictions on Exploitation of the Patent Monopoly: An Economic Analysis

*William F. Baxter**

I. A Frame of Reference

The patent laws confer on a patentee power to exclude all others from making, using or selling his invention.[1] In furtherance of a constitutionally recognized goal—"To promote the Progress of Science and the useful Arts"—Congress has thus adopted a constitutionally authorized means—"securing . . . to Inventors the exclusive Right to their respective . . . Discoveries."[2] The constitutional clause is remarkable in several respects. Its recognition of the possibility that invention might require encouragement implies not only that technological innovation is desirable but also that, but for legal subsidization, the quantity of innovation forthcoming would or might be less than optimum. This recognition, coming on the morn of an era during which the tendency of a free market to achieve optimality in all activities was greatly and religiously overestimated,[3] prompts brief inquiry into the soundness of the supposition.

Several considerations support the view that the market would yield less than optimum innovative activity.[4] Invention consists primarily of knowledge; and he whose research, experimentation and insight has

* Professor of Law, Stanford University. A.B. 1951, LL.B. 1956, Stanford University.

1. 35 U.S.C. § 154 (1964).
2. U.S. CONST. art. I, § 8.
3. *Cf.* JAFFE, ADMINISTRATIVE LAW, CASES AND MATERIALS 3-8 (1954).
4. See, *e.g.*, ARROW, *Economic Welfare and the Allocation of Resources For Invention*, in THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 609 (1962); Machlup, *An Economic Review of the Patent System*, Study #15 of the Subcommittee on Patents, Trademarks and Copyrights of the Committee on the Judiciary of the U.S. Senate, 85th Cong., 2d Sess. 58-62 (1958) [hereinafter cited as Machlup, Study #15].

created it has no obvious means of appropriating it to his own reward. It is fruitless unless used, but any extensive use virtually assures disclosure and, in time, equal availability to all. As a form of economic activity, invention is attended by much risk even apart from the problem of appropriation, and difficulty of appropriation greatly increases the economic risk. A man debating whether to commit his resources to manufacturing shoes can estimate with tolerable accuracy both how many shoes he can produce in a month and their worth upon completion. If, alternatively, he considers committing those resources to the process of innovation, he can have little confidence in either the quantity of his informational output or its worth when and if produced. A production process characterized by highly uncertain outputs as a function of given inputs will generally attract less than optimum inputs.[5] For these reasons innovative activity is subsidized in some way and to some extent in almost all advanced economies.[6]

But to conclude that some subsidization is appropriate is a very short step down the road to firm conclusions about "how" and "how much." This paper is addressed principally to issues raised by the question: "how"; the very little that can be said on the question: "how much," is best said immediately.

In the abstract, the criterion for how much subsidy is appropriate is simply stated. Innovative activity should be subsidized as much and no more than is necessary to attract to that activity those inputs which, if invested in any other activity, would yield a product of lesser social value.[7] Market forces can generally be relied upon in areas where no substantial disparity exists between private and social value of either inputs and outputs. Private investment will be attracted to activities in which the private value of output most exceeds private cost of inputs; and if social and private valuations are comparable, activities with outputs of high social value will not be neglected.

The input costs of innovation, both privately and socially valued, are often very large and so too is the social value of its outputs. But the private value of innovative output, if left to be determined by market forces, would usually be small because of unappropriability. Hence private investment in innovation would stop at the point where the next private dollar invested would yield greater private gains if

5. Arrow, *supra* note 4, at 610-14.

6. White & Ravenscroft, Patents Throughout the World (2d ed. 1944).

7. This is the classic economic criterion for optimal allocation. See, *e.g.*, Boulding, Economic Analysis 168-72 (3d ed. 1955).

invested elsewhere. Yet the social value of additional innovative output, being greater than its private value, would exceed the social value of the output to which investment had been diverted. Corrective diversion of investment from other activities to innovation is desirable because the social value of output gained thereby exceeds that of the outputs lost. That corrective diversion can be accomplished by subsidy.

Though subsidy is thus shown to be desirable, the question of how much subsidy is desirable could only be answered by a controlled experiment that would permit measurement of the social value of innovative output and of unsubsidized alternative outputs at each of a series of subsidy levels. Any such experiment is well beyond the reach of present techniques.[8]

Implicit in the assertion that there is an optimum amount of subsidization and that it cannot be determined with any precision is the further assertion that the present amount probably is somewhat too large or too small. Consequently the resources presently being devoted to innovation probably are somewhat too large or too small. The fact that we may already be devoting too many resources to innovation seems worth emphasis because so many commentators accept as absolute truth the contrary assumption that additional encouragement of research, or at least the present amount of encouragement, is desirable.[9]

That too many resources could be, and perhaps are being, devoted to innovation becomes obvious once it is noted that total resources available are not infinite and that there are competing activities of great social utility to which resources would move if the subsidy for innovative activity were decreased. If fewer hours of highly skilled manpower and lesser quantities of expensive equipment were devoted to innovation, as a society we could have more medical care, education, housing and police protection, just to give a few examples. This would be true even if the output of innovative activity were directly proportionate to input at all levels of input; but presumably the law of diminishing returns applies with as much validity to innovative activity as to other activities. The now frequent example of simultaneous

8. Machlup, Study #15, at 60-62.

9. This is true not only of pro-patent writers from whom the assumption might be expected, see, *e.g.*, Frost, *The Patent System and the Modern Economy*, Study #2 of the Subcommittee on Patents, Trademarks and Copyrights of the Committee on the Judiciary of the U.S. Senate, 84th Cong., 2d Sess. 1-20 (1957), but also from anti-patent writers who are forced by the assumption into the position that the patent system does not significantly encourage innovative activity and does have undesirable side effects. See, *e.g.*, Kahn, *Deficiencies of American Patent Law*, 30 AM. ECON. REV. 475 (1940).

independent discovery of some new process or product is an illustration of diminishing returns.[10]

As is implied in the foregoing paragraphs, subsidization by direct payments from tax revenues and subsidization by conferral of a legal monopoly are analogous from the standpoint of their capacity to increase, and perhaps to increase too much, the flow of resources to innovative activity. One might argue that over-investment in innovation could not be induced by the monopoly approach. That argument would run as follows.

Conferral on the inventor of a monopoly over his invention is a direct attack on the problem of unappropriability and thus tends to equalize the private and social values of the invention. Resource allocation to innovation approximating the optimum would result if inventors were granted a total and permanent monopoly over their inventions. The inventor could then garner from all who wished to use it the value of his invention in each context in which it was useful, and its private value to the inventor would equal its utility to society. Since no prospective user could be induced to pay more than the value of the invention to that user, private value to the inventor could never exceed social value; and over-investment in innovation would not occur.

The argument is unsound because it fails to take into account indirect costs not borne by the inventor that the monopoly device imposes on society. These costs diminish the net social value of the invention and upset the private value—social value equation. One of these indirect costs is the following: The monopoly over the invention decreases the incentives of innovators other than the initial patentee to engage in further research within the field covered by the monopoly. For any improvement discovered will be subservient to the initial patent and useless in the hands of anyone but the initial patent holder; and in dealing with the initial patent holder the subsequent inventor will occupy a weak bargaining position since he faces a monopsonistic buyer. The extent of this dampening effect increases as the legal life of the initial monopoly is made longer. A second important indirect cost is the resource misallocation which results from under-utilization of past inventions for the use of which a monopoly price is charged, a point on which more will be said.

10. For a splendid, more extensive articulation of the thesis of this and the next paragraph on which I have drawn extensively, see Machlup, Study #15, 44-73.

The two social costs described by no means exhaust the list, but reference to them is sufficient to make the point that the monopoly device, like direct subsidy payments, is capable of diverting too many resources to innovation and that the question of how much monopoly-subsidy to confer is a difficult one. Thus one debating whether to make a marginal change in the amount of monopoly-subsidy should draw up a balance sheet on the following order. As assets supporting an increase, one would list: (a) the social value of having those inventions which *never* would have been made but for the increase; (b) the social value of having *sooner rather than later* those inventions which would not have been made until later but for the increase. As liabilities one would list: (a) the social value of the medical care, education, etc., not achieved because of resource diversion; (b) the greater cost of administering the larger subsidy—patent office personnel, patent solicitors, lawyers, judges, their secretaries, clerks and the buildings in which they work; (c) the greater amount of equipment discarded and labor temporarily displaced as a result of more rapid obsolescence of existing production processes; and, perhaps most importantly, (d) the social cost of conferring more monopoly and hence more output restriction, not only with respect to the marginal inventions listed as assets because they were generated by the subsidy increase, but with *respect to all the inventions that would have been made without the increase.* The same balance sheet with the "asset" and "liability" column—heads switched and with minor changes in wording would indicate whether the present amount of monopoly-subsidy should be decreased.[11]

To speak of balance sheets is misleading in its suggestion of precision; for the point remains that, given the existing state of the economic art, we cannot say how much subsidy is optimal. The foregoing discussion of optimality is not wholly futile however, for the total intractability of the question seems to me significant in several important respects. As a practical matter the question, despite its intractability, must be answered either explicitly or implicitly by some authoritative rule-maker if a society is to take action on the knowledge that analysis does yield—namely that some subsidy is appropriate. The answer will necessarily be a rough judgment and almost certainly will be "wrong" in the sense that a somewhat different answer would have come closer to achieving optimality; but if the answer is given with awareness of the nature of the problem and with a modicum of intelligence, the resulting situation is very likely to be an improvement

11. See note 10, *supra.*

over that which market determination of innovative inputs would produce.

For the United States the judgment has been made by Congress and made in a way, by conferral of monopoly, which inextricably intertwines the issues of "how" and "how much." In view of the intractability of the question "how much," it seems peculiarly appropriate that the answer should have been given by a political branch, and the same consideration reinforces traditional doctrine in dictating that the judiciary should accept that judgment not grudgingly but with full acquiescence.

It is also true that the Congressional judgment of "how much" was made with no explicit attention to the nature of the problem as I have described it. In part the amount of subsidy is implicit in the monopoly technique through which subsidy is afforded. The other determinant of amount, one more susceptible to adjustment, is the period for which the monopoly is conferred. The monopoly technique was adopted with little attention to possible alternatives,[12] and a term of 14 years was provided for in the first patent act without any recorded discussion of alternative periods of time.[13] Subsequently, however, the issue of term length has received more direct attention. A proposal to shorten the patent term to 10 years was made and rejected in 1793.[14] In 1836

12. That means of reward is suggested by U.S. CONST. art. I, § 8, but the records of the Convention do not reveal any discussion of why that means was selected. Very probably it was adopted because the same means had been employed in England for many centuries. For a colorful, lucid and brief discussion of the English history, see WALTON HAMILTON, PATENTS AND FREE ENTERPRISE, (TNEC Monograph No. 31, 1941). This technique would also commend itself to a new government because it depends largely on private enforcement and minimizes official administration.

Various types of direct subsidies are reported to have been used in the Colonies. *Colonial Monopolies and Patents*, 18 J. PAT. OFF. SOC'Y 35 (Centennial Number, July, 1936). And in the first Congress allusion was made to the use of direct subsidy; but doubt was expressed whether any but that means expressly mentioned in the Constitution was within Congressional power. 1 ANNALS OF CONG. 170 (1789). At one point the Senate passed an amendment to a bill which would have "invest[ed] the judges of the Supreme Court with a power to determine the compensation which persons shall receive for their inventions. . . ." 2 ANNALS OF CONG. 1519 (1790). But the monopoly technique was incorporated in the first patent statute. 1 Stat. 109 (1790).

13. 1 Stat. 109 (1790). In discussions recorded in the Annals the phrase "for a term of years" occurs repeatedly. See, *e.g.*, 1 ANNALS OF CONG. 170, 950, 951, 955 (1789-90). Several petitions for patents sought exclusive rights for a specific period; *e.g.*, id. at 233 (21 years); *id.* at 266 (21 years), but most of them also referred to "a limited time," *e.g.*, *id.* at 328, 424, or "a term of years," *e.g.*, id. at 335, 342, 642. On the time period, too, it seems probable that the primary causal factor was familiarity with the prior English law which provided for a 14 year period. Statute of Monopolies, 21 Jac. 1, c. 3 (1624). The selection of that time period in England has been attributed to the fact that 14 years corresponded to two successive terms of apprenticeship and thus allowed for training two generations of apprentices in the new art. HAMILTON, *op. cit. supra* note 12 at 16; *Origin and Early History of Patents*, 18 J. PAT. OFF. SOC'Y 33 (Centennnial Number, July, 1936).

14. 4 ANNALS OF CONG. 860 (1793).

a provision was passed authorizing extensions of the term for an additional 7 years upon a showing that the patentee had not yet been adequately rewarded for his contribution;[15] but the extension provision was repealed in 1861, and the present 17 year term was adopted as a compromise.[16] On numerous occasions since 1861 the question of term length has been debated in Congress.[17] Hence the amount of subsidy presently afforded, though initially founded on practices of a totally different culture,[18] must be accepted as a deliberate act of legislative judgment.

As a means of paying the subsidy the monopoly device, while conducive to the end, is not the only means that might have been conceived; and it would be rash to suppose that by authorizing one means, the Constitution by implication forbids others. A monetary reward, scaled to the value of the invention and payable out of the federal treasury, might serve the end as well. Both direct subsidy and monopoly-subsidy have advantages and disadvantages. The primary advantage of monopoly is that it eliminates the need for a governmental institution charged with the task of adopting and applying criteria by which the total amount of subsidy would be determined and by which that amount would be apportioned among individual inventors. It is very unlikely that the history of an institution so charged would be a credit to its parent polity.

In place of such an institution, the monopoly technique substitutes a market mechanism. The extent of the inventor's reward is the amount he can induce others to pay him in exchange for his waiver of the exclusive right conferred upon him. The aggregate subsidy is the aggregate of such payments to all inventors. The monopoly device

15. 5 Stat. 124-25 (1836). Note that this provision required official determination of what constituted an adequate reward and therefore involved all the administrative difficulties that would inhere in direct governmental subsidization; yet it eliminated none of the economic costs that inhere in the monopoly technique of subsidization. Its repeal was inevitable; only the fact that repeal took 25 years is surprising.

16. 12 Stat. 249 (1861).

17. *Hearings on S.3325 and S.3410 Before the Senate Committee on the Judiciary*, 67th Cong., 2d Sess. *passim.* (1922); *Hearings Before the Temporary Economic Committee of the Congress of the United States*, 75th Cong., 3d Sess. 852-55, 860-63, 892, 944-45 (1939); S. Doc. No. 35, 77th Cong., 1st Sess. 37 (1941); *Hearings on S.2303 and S.2491 Before the Senate Committee on Patents*, 77th Cong., 2d Sess. 1748 (1942); H.R. Doc. No. 239, 78th Cong., 1st Sess. 7 (1943); *Hearings on H.R. 323 and H.R. 4054 Before the House Committee on the Judiciary*, 82d Cong., 1st Sess. passim. (1951); H.R. Rep. No. 1297, 84th Cong., 1st Sess. *passim.* (1955); *Hearings on H.R. 2128, 3134, and 4700 Before the House Committee on the Judiciary*, 84th Cong., 1st Sess. *passim.* (1955); *Hearings Before the Subcommittee to Investigate the Administration of the Internal Security Act and Other Internal Security Laws of the Senate Committee on the Judiciary*, 84th Cong., 1st Sess. 99-100, 143, 152 (1955); S. Rep. No. 1464, 84th Cong., 2d Sess. 6, 13 (1956); S. Rep. No. 72, 85th Cong., 1st Sess. 12, 30 (1957).

18. See note 13, *supra.*

does not leave the aggregate subsidy, the primary determinant of resource allocation, wholly to chance; for Congressional control over the term of monopoly affords a control valve at least as precise as the criterion for the valve's adjustment. Hence the system may be viewed as embodying two quite different distributional theories in happy symbiosis: to the inventor in proportion to his contribution; to innovative activity generally in accordance with the needs of society.

But other aspects of the monopoly device are less satisfactory. The first of these returns us to the problem of social value. Here the concern is not, as it was before, social value of innovative activity in the aggregate, but the social value of a particular invention. The principal determinant of the social value of an invention is the extent to which it is useful and used. Resource allocation in an economy is optimum when each type of good and service is being used to that extent at which the social cost of producing the last unit of a quantity of product equals the utility or satisfaction consumers derive from the last unit. Other things being equal, inexpensive products will and should be used more extensively than costly ones. An invention is knowledge; its acquisition may have been very expensive, but once it is acquired it costs no more to use it extensively than sparingly. All the costs represented by the invention are "fixed" or "sunk" costs; and the incremental or marginal cost of an additional use made of it is zero. Optimally, an invention should be used in every context and to whatever extent it has any utility at all; and such use would occur if the inventor's reward were provided by direct government subsidy rather than by monopoly-subsidy. Reliance on monopoly imposes a cost on the user which has no marginal cost counterpart either to the inventor or to society as a whole. Hence use is undesirably curtailed; resource allocation is distorted.[19]

The monopoly-subsidy device also has a drawback in that it distorts thinking. In its economic impact, the device closely resembles a federal excise tax on the use of recent innovations collection of which is followed by appropriation and payment of corresponding amounts to the inventor. The patent monopoly is an officially authorized charge that does not correspond to any cost of incremental use. If the device were transformed into the analogous tax-subsidy, its consequences would be

19. The problem discussed in this paragraph is the familiar one of indivisibility. Its classic exposition is, Hotelling, *The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates*, 6 ECONOMETRICA 242 (1938).

The fact that an incremental use of an invention has a marginal cost of zero gives rise to other problems, principally those of economic or monopolistic discrimination, discussed at pp. 280 *et seq., infra.*

far more apparent. There are objections to excise taxes because of their effects on income distribution that cannot be examined here.[20] The resource allocation effect of excise taxes certainly is difficult to defend when the subject of the tax is use of information.[21] And if the present monopoly-subsidy to innovative activity appeared each year as an item in the federal budget, the erroneous but prevalent notion that encouraging innovation is costless[22] would vanish.

The final, and for the purpose of this paper the most important, objection to the monopoly-subsidy is that it is subject to grave abuse. From the Congressional viewpoint this fact must be taken into account with all the previously described advantages and disadvantages of the monopoly device in deciding whether to retain the device as a means of subsidy. But for the courts the potentiality and frequent actuality of abuse pose problems distinct from the other considerations. So long as Congress retains the patent statutes in essentially their present form, those statutes represent a legislative judgment that, on balance, the monopoly-subsidy is desirable; and the courts have neither responsibility for that judgment nor authority to change it. With respect to the problem of abuse their responsibility is different; for they have been given both the task of enforcing the patent monopoly, which carries the appurtenant responsibility of deciding what conduct is a proper assertion of the monopoly reward and what conduct is an abuse, and the task of enforcing the antitrust laws with which improper use of the monopoly often conflicts.

II. The Patent Monopoly

The patentee's exclusive right to make, use and sell[23] his invention is deceptive in its apparent simplicity. Very rarely would a patentee choose to invoke his right in accordance with the literal meaning of the statutory terms and deny to others all that he might deny. The holder of a process patent who was an integrated manufacturer-seller of an unpatented product yielded by the process might choose to license no one and extract his reward exclusively through monopoly profits in marketing the product. In such a case no one but the patentee would be making, using or selling the invention. But if the patent is on an endproduct, then, though no formal licenses are issued, those

20. See, *e.g.*, SIMONS, FEDERAL TAX REFORM 36-37 (1950); Eldridge, *Distribution of Sales and Excise Tax Burdens*, Excise Tax Compendium, Committee on Ways and Means, 88th Cong., 2d Sess., pt. 1, at 73-87 (1964).

21. For a good general discussion of this and other aspects of excise taxes, see DUE, GOVERNMENT FINANCE ch. 19 (2d ed. 1959).

22. See note 9, *supra*.

23. 35 U.S.C. § 154 (1964).

to whom the product is sold will use the invention and have an implied license to do so. In the vast majority of cases the patentee obtains his reward not by affirmatively asserting his exclusive right but by extracting a consideration for a total or partial waiver of that right.

As the reference to partial waiver underscores, the patentee is not limited to the alternatives of full assertion or total waiver but can, for a price, allow others to engage in some but less than all of the conduct that assertion of his exclusive right would preclude. Thus as T. R. Powell insisted with characteristic lucidity almost a half century ago,[24] analysis is aided if the patentee is viewed as having not a single right but a whole bundle of rights to exclude, all, any one or none of which may be waived. The courts, in giving effect to the statutory phrase, have consistently arrived at results incompatible with any other view.[25]

Just as the use of the word "right" in the statute quite properly did not preclude results consistent only with use of the word "rights," so also no dictate of grammar or metaphysics should be thought to allow waiver of one particular group of patentee rights but to preclude waiver of another, nor to allow waiver of a group of rights in exchange for one type of consideration but preclude it for another type. On these points, however, the courts have had more difficulty. A license to use a patented machine with, but only with, supplies furnished by the patentee was first upheld[26] and later struck down.[27] A license to resell a patented product purchased from the patentee at, but only at, a price set by the patentee was struck down;[28] but a license to manufacture and sell a product at, but only at, a price set by the patentee was upheld.[29] These results may be right or wrong, but sound answers to the problems the cases posed cannot be reached by parsing the highly general language of the statute,[30] by metaphysical assertions that the right to exclude totally necessarily embraces the right to exclude partially[31] or by framing question-begging generalities about what is

24. Powell, *The Nature of a Patent Right*, 17 COLUM. L. REV. 663 (1917).
25. See, *e.g.*, Rubber Co. v. Goodyear, 76 U.S. (9 Wall.) 788 (1869).
26. Henry v. A.B. Dick, 224 U.S. 1 (1912); Heaton-Peninsular Button-Fastener Co. v. Eureka Specialty Co., 77 Fed. 288 (6th Cir. 1896).
27. Motion Pictures Patents Co. v. Universal Film Mfg. Co., 243 U.S. 502 (1917).
28. Bauer & Cie v. O'Donnell, 229 U.S. 1 (1913).
29. United States v. General Electric Co., 272 U.S. 476 (1926); Bement v. National Harrow Co., 186 U.S. 70 (1902).
30. See, *e.g.*, Motion Picture Patents Co. v. Universal Film Mfg. Co., 243 U.S. 502, 509-13 (1917).
31. See, *e.g.*, Motion Picture Patents Co. v. Universal Film Mfg. Co., 243 U.S. 502, 519 (1917) (dissenting opinion, Holmes, J.); Heaton-Peninsular Button-Fastener Co. v. Eureka Specialty Co., 77 Fed. 288, 291, 294-95 (6th Cir. 1896).

"normally and reasonably adapted to secure pecuniary reward."[32] By the first sale of a patented article, the Supreme Court has said, "the article . . . [is] thereby carried outside the monopoly of the patent law and rendered free of every restriction which the vendor may attempt to put upon it."[33] But quite obviously the article is not carried outside the monopoly in the sense that a chair may be carried outside a room; no phenomenon susceptible to empirical validation is involved. The Court's statement was not made because it was true; it is "true" only because the statement was made and will be "true" only until the Court or Congress makes a contrary statement.

No one type of consideration more than any other extracted in exchange for a waiver of all or part of the patentee's rights is, for merely logical reasons, "normally and reasonably adapted to secure" the objectives of the patentee. The disadvantages which attend use of the monopoly-subsidy device are tolerable only because of the offsetting advantages of private bargaining—only because private bargaining simultaneously can yield to both the patentee and the licensee more of what each, given his unique situation, deems of value. A promise by the licensee to murder the patentee's mother-in-law is as much "within the patent monopoly" as is the sum $50.00; and it is not the patent laws which tell us that the former agreement is unenforceable and subjects the parties to criminal sanctions.

The last example, while silly in its extremity, nevertheless illustrates two important points. The value to the patentee of the licensee conduct sought by the patentee may exceed its cost to the licensee. Hence both parties may be better off as a consequence of their exchange. Therein, of course, lies the great value of private arrangements. But therein, too, lies their danger; for conduct of value to the patentee may be highly injurious to third persons. Indeed it may be of value to the patentee precisely because it is injurious to third persons. Courts should be alert to penalize such bargains by imposing on the immediate parties the sanction of nullity and other appropriate sanctions. Only rarely will the policies calling for protection of third persons be found in the patent laws; to search for them there is generally futile, to purport to find them there is generally duplicitous and confusing.[34]

32. United States v. General Electric Co., 272 U.S. 476, 490 (1926).
33. Motion Picture Patents Co. v. Universal Film Mfg. Co., 243 U.S. 502, 516 (1917).
34. In Motion Picture Patents Co. v. Universal Film Mfg. Co., 243 U.S. 502, 512-13 (1917), the Court purported to find a basis for invalidating a tie-in in the patent laws. For a devastating criticism of the court's analysis, see Powell, *The Nature of a Patent Right*, 17 COLUM. L. REV. 663 (1917). Persuasive reasons for the result are available, see pp. 318 *et seq.*, *infra*, but not where the court pretended to find them.

Effect of licensee conduct on third persons may or may not, depending on the circumstances, constitute adequate reason for striking down a license agreement. But only in most extraordinary circumstances will the effect on the licensee justify so doing. As a class patent licensees are unlikely candidates for "ward of the court" status, and cases that rest on the premise of overreaching are likely to be unsound, either in their reasoning[35] or in both reasoning and result.[36]

From this conclusion—that legality of a patent license properly turns on its impact on third persons and on whether the parties are privileged to impose this effect—it follows that too much significance has been accorded in the past[37] to the distinction between condition and covenant. Any of the typical license limitations can be imposed through either device. A license can be broad in its waiver of patentee rights and contain a promise by the licensee that he will not sell the patented product for less than a specified price; or it may license only those sales which are made at prices equal to or exceeding a specified price. A license may authorize generally use of a patented machine and contain a promise by the licensee to buy all supplies used with the machine from the patentee; or it may license use of the machine only in conjunction with supplies purchased from the patentee. It may license manufacture and sale generally and contain a licensee covenant to confine his activities to Hecate County; or it may license the activities only in that county. If in each of these cases the conduct of the licensee will be the same, then the third person impact will be the same and there is no basis for upholding one form and invalidating the other.

In general the distinction between covenant and condition should be disregarded because there is no reason to suppose that licensee conduct will be affected by the form used. In either case he is liable to sanctions if he engages in unauthorized conduct—damages for breach and perhaps an injunction in the case of covenant; an infringement suit for damages and injunction in the case of condition. To the extent that the licensee is dependent upon good relations with the patentee,[38] unauthorized operations are as much deterred by condition as by

35. Brulotte v. Thys Co., 379 U.S. 29 (1964).

36. American Securit Co. v. Shatterproof Glass Corp., 268 F.2d 769 (3d Cir. 1959).

37. See, *e.g.*, STAFF OF SUBCOMM. NO. 5, HOUSE COMM. ON THE JUDICIARY, 84TH CONG., 2D SESS., ANTITRUST PROBLEMS IN THE EXPLOITATION OF PATENTS 8 (Comm. Print 1956); Automatic Radio Mfg. Co. v. Hazeltine Research, Inc., 339 U.S. 827, 836 (1950) (dictum); Atlas Imperial Diesel Engine Co. v. Lanova Corp., 79 F. Supp. 1002 (D. Del. 1948).

38. Dependence on the patentee for "know-how" in the use of the invention is an example.

covenant: infringers no less than promise-breakers are pariahs in the business community.

That the distinction between covenant and condition is generally without substance is so clear that one seeks explanation of how it came to be accorded significance. Explanations of nonsense are necessarily somewhat speculative, but the probable origins of the foible are several. First, availability of a federal forum often turns on the distinction. Suit to enforce a condition usually takes the form of an infringement charge and hence "arises under" the patent laws;[39] but suit for breach of covenant must be brought in the state courts unless diversity is present.[40] Second, there is great confusion in the cases regarding the circumstances under which a defendant in a contract action may assert an anti-trust defense,[41] but where the suit is for infringement a defense of "misuse" generally is allowed.[42] Finally it is more difficult, apparently, for courts to perceive anticompetitive effects of conditions than of covenants.[43] A promise not to sell at prices below the patentee's or outside a set territory falls neatly alongside the "contract in restraint of trade" phraseology of the Sherman Act. But when faced with a comparable condition, the courts tend to reason that *all* competition was forbidden before the license and hence no suppression of competition could be attributable to a license which made *some* competition possible. Even to the extent that this latter analysis is sound—and it sounds much better before careful scrutiny than afterward—the distinction between condition and covenant is not aided; for it is equally true that there would have been no occasion for the promise not to compete by certain acts but for the authorization found in the same document to compete in other ways. Maximum rational significance is accorded the distinction between covenant and condition when it is viewed, in conjunction with other circumstances, as probative of future licensee conduct.

III. Some Basic Economics

In considering more particularly just what the patentee's monopoly does entitle him to do and which techniques of exploitation should

39. See, *e.g.*, Henry v. A. B. Dick Co., 224 U.S. 1 (1912); White v. Rankin, 144 U.S. 628 (1892); Mishkin, *The Federal "Question" in the District Courts*, 53 COLUM. L. REV. 157 (1953).

40. See, *e.g.*, Albright v. Teas, 106 U.S. 613 (1882).

41. Compare Kelly v. Kosuga, 358 U.S. 516 (1959) with United States Gypsum Co. v. National Gypsum Co., 352 U.S. 457 (1957).

42. See Edward Katzinger Co. v. Chicago Metallic Mfg. Co., 329 U.S. 394 (1947); MacGregor v. Westinghouse Elec. & Mfg. Co., 329 U.S. 402 (1947).

43. *Cf.* Automatic Radio Mfg. Co. v. Hazeltine Research, Inc., 339 U.S. 827, 828 (1950); Atlas Imperial Diesel Engine Co. v. Lanova Corp., 79 F. Supp. 1002, 1004 (D. Del. 1948).

be denied to the patentee because of their effect on third persons, frequent reference to economic concepts will be necessary. In an Appendix I have attempted to set forth some basic concepts and definitions. Reference to the Appendix may be helpful to some readers at this point.

IV. Price Discrimination in Patent Licensing

Running through much of the controversy over legal limits on a patentee's freedom to exploit his monopoly is a thread I will call economic, or price discrimination.[44] Territorial and field limits and tying provisions, for example, can and probably are used to achieve the private gains of economic discrimination under appropriate circumstances. Each may be used to facilitate segregation of licensees into categories with differing demand elasticities; and returns can then be maximized by setting a different royalty rate for each category in accordance with the marginal revenue function of that category.

A patentee obviously has a fertile field in which to practice price discrimination. He has a legal monopoly; to the extent it represents an economic monopoly as well, competitors cannot seek out and take away his high paying customers. Assuming the enforceability of the license restrictions and prohibitions against sublicenses that are used to segregate purchasers, retransfer of the patentee's product from favored to disfavored purchasers need not be feared.[45] Thus the legal protection afforded a patentee, if unaccompanied by any legal prohibitions which will block discriminatory pricing, places the patentee in a unique position: no regulatory body reviews his rate structure "in the public interest"; no private party may enter his market and spoil his game; the monopoly source of his power to discriminate is immune from direct attack under the anti-trust laws. His opportunity and incentive to extract maximum consumer surplus from his licensees and, perhaps in the course of so doing, to impair or destroy the ability of one licensee to compete with another is great.[46] Does law have a role to play here? Or must the situation, its apparent potential for harm notwithstanding, be viewed as necessarily implicit in the choice to use monopoly to subsidize inventive activity?

44. I do not intend to suggest by my caution in applying this label that either the label or the concept it usually represents is original with me; obviously they are not. Rather, I apply the label hesitantly because, as will be explained, the phenomenon examined here differs in an important respect from that to which the label is usually attached; and whether the label is appropriate is therefore disputable.

45. The term "favored" purchaser or licensee designates the one with a more elastic demand who therefore is charged a lower price or royalty rate.

46. *Cf.* Grand Caillou Packing Co., TRADE REG. REP. (1963-65 Transfer Binder) ¶ 16,927 (1964).

One's instinctive inclination to curb patentee discrimination encounters the difficulty that effective prohibition presupposes an ability to articulate criteria for what does and does not constitute forbidden discrimination. That task is extraordinarily troublesome. In the broadest sense, price discrimination may be said to exist whenever any two products are being sold in an economy at prices which bear differing ratios to the respective marginal costs of the products. This definition is independent of product definition and of seller identity; it requires only that price-marginal cost ratios be ascertainable. But the definition is too broad to yield a workable prohibition. Elimination of discrimination in that broad sense is a major purpose of the entire competitive market system; and legal attack upon it in that broad sense would involve wholesale substitution of administrative for market mechanisms.

For this reason the term economic discrimination is usually defined in much narrower terms,[47] and legal attack upon it is confined to much narrower channels.[48] A typical definition would be, sales by a single seller of closely related or identical products at prices bearing disparate ratios to their respective marginal costs.[49] Even if one is willing to ignore the substantive arbitrariness inherent in this narrower definition, it is very difficult to build upon the definition a legal prohibition that will prove administrable in the patent licensing context. An attempt to construct such a prohibition will be made, as much for the purpose of exploring the difficulties as for the utility of the result.

Initial problems are posed by the vagueness of the narrower definition. It contains three concepts, each of which is difficult to apply: "a single seller," "closely related products," and "price-marginal cost ratio." Like all concepts, these lack inherent meaning and must be given content in accordance with the purposes sought to be accomplished by their use. To ascertain these purposes it is necessary to focus more carefully on the phenomenon of economic discrimination and its asserted evil consequences.

The nature of production processes is generally such that, in any but the very long run, some part of total costs is fixed—independent of output—and another portion varies with output. Marginal cost is a mathematical function of the variable cost component. The seller's interest is never advanced by selling at prices below marginal cost; but

47. See, *e.g.*, Machlup, *Characteristics and Types of Price Discrimination*, in BUSINESS CONCENTRATION AND PRICE POLICY 398 (1955).

48. See, *e.g.*, Robinson-Patman Act § 2, 49 Stat. 1526 (1936), 15 U.S.C. § 13 (1964); Interstate Commerce Act §§ 2-4, 24 Stat. 379 (1887), as amended, 49 U.S.C. §§ 2-4 (1964).

49. See, *e.g.*, Machlup, *supra* note 47, at 398.

if the demand function he faces is sloped, he will sell above marginal cost. The returns earned by reason of the fact that price exceeds marginal cost may serve to cover his fixed costs or to yield him a monopoly profit or both. For present purposes consider the seller's fixed costs and his monopoly profits as a single category labeled "excess needs."[50] The pricing problem faced by the seller may now be put in this form: "I will sell no unit of product at less than marginal cost. Having made that decision, I have implicitly decided how I will allocate my aggregate variable costs among the units of product I sell—I will allocate an equal portion to each unit. *But how shall I allocate my excess needs among those units?*"

A rule against price discrimination is a normative answer to that question: "You must allocate excess needs so that the same percentage of each unit's price is a contribution to those needs," or, "To set price, always mark up marginal cost by the same percentage."[51] If price discrimination is possible because customers with disparate demand elasticities can be segregated, then the objective to maximize yields a different normative answer: "To maximize, allocate so as to achieve prices scaled to demand elasticity."

The last answer given sets forth explicitly the private objective that dictates it: to maximize. What are the social criteria that dictate the first answers? I can perceive only three, the first two of which are rather crude and emotive rather than utilitarian.

(1) Regarding the relationship between seller and disfavored customers, fairness requires that the seller desist from maximizing revenue at the customers' expense. Revenue maximization to the point permitted by a uniform-price price structure is all community standards of equity will tolerate. Charging disfavored customers a price in excess of the maximizing uniform price cannot be justified by the fact that another group of customers is simultaneously charged a price lower than the maximizing uniform price.

(2) Regarding the relationship between favored and disfavored cus-

50. This and the preceding sentence assume that marginal cost equals or exceeds average variable cost at the level of operation involved, a condition that will be satisfied in the bulk of cases. But the analysis in the text is valid even if the condition is not satisfied; a more complex definition of "excess needs" would be the only necessary change.

51. The substance of the sentence in the text may be stated in several ways that are mathematical equivalents of one another. Let p = price, $m.c.$ = marginal cost and $e.n.$ = excess needs. Then $p = m.c. + e.n.$ The statement in the text imposes the constraint that for each unit $e.n. = (x\%)\,(p)$. This is equivalent to the statement that $e.n./p$ must be the same for all units and to the statement that $p/m.c.$ must be the same for all units. The last mode of statement corresponds to the classic definition of discrimination.

tomers, the community standard requiring equal treatment of persons similarly situated applies; all purchasers are to be regarded as similarly situated notwithstanding those dissimilarities which account for their differing demand elasticities. Hence, you may not charge some more than others.

(3) The third and more utilitarian argument is similar to criterion (2): When the favored and disfavored customers are not ultimate consumers but are themselves members of an industry and each group is in competition with the other, then a persistent pattern of price discrimination will cause the disfavored group to have higher costs and to be handicapped in its competition with the favored group. Entry to and exit from the customer industry can be expected to bring about structural changes in that industry which at best will be artificial in that they are unrelated to real cost savings to society, and at worst will be undesirable in that real economies available to the disfavored customer category will have been overbalanced by the discriminatory pricing and eventually those savings may be lost to society.[52]

In attempting to give content to the concepts of "a single seller," "closely related products" and "price-marginal cost ratios," what guidance can be derived from the foregoing objections to discrimination? The first two "fairness" arguments yield no assistance at all. They make clear that homogeneity or similarity of the discriminatorily priced products is relevant only insofar as it increases the visibility of the asserted inequity. The disfavored customer is being treated no more unfairly, vis-à-vis the seller or the favored customer, than is any customer who has a relatively inelastic demand for a product sold under relatively monopolistic conditions. Both the disfavored customer of a discriminating monopolist and the customer of a single-price monopolist are being deprived, without economic justification, of consumer surplus to a greater extent than are other consumers in the economic system. The unfairness is not caused by, but is made obvious by, the availability to others of a "closely related product" at a lower price.

Similarly, it is irrelevant from the standpoint of equity whether the disparately priced products are being sold by the same or by different

52. This argument, obviously, is the economic premise underlying the basic provisions of the Robinson-Patman Act, 49 Stat. 1526 (1936), 15 U.S.C. § 13 (1964). The administration that Act has received should not be permitted to generate false doubts about the soundness of the premise.

The "primary line" Robinson-Patman cases suggest a fourth reason why discrimination may be undesirable; it may be used as a predatory device by a seller against his own competitors. This argument, however, has difficulties of its own and adds nothing helpful for present purposes to the three stated in the text.

sellers. Again it is true that the obviousness of the asserted injustice rather than its existence gives significance to unified control over price.

Even when attention is shifted to the third and most concrete purpose of an anti-discrimination rule, to prevent distortions of structure in the industry which purchases the disparately priced products, no guidance can be found. A disfavored customer will be impaired in his ability to compete if he is subjected to monopolistic exploitation with respect to any important factor input. This will be no less true whether the disfavored customer buys that factor input from the same or a different source of supply than does his rival and whether the factor input is exactly the same as or very different from the functionally analogous factor input used by his rival.[53] Here, too, it is true that the existence of a common seller and of physical or functional similarity between the factor inputs merely enables us to see how and by whom the disfavored purchaser is being prejudiced.

I do not mean to suggest that the law should not strike at ascertainable evils merely because they are indistinguishable in principle from evils not ascertainable. But I do argue that the scope to be given to the prohibition against economic discrimination must be determined not on an economic basis but by reference to problems of proof and other factors relating to ease of administration and enforcement.

The foregoing analysis is applicable to the pricing of the mine-run product; but when the product is the right to practice a patented invention, a further difficulty appears. The marginal cost of any incremental use of a patented invention is zero, and hence it is impossible to make any useful statement about the ratio between price (royalty) and marginal cost. All costs associated with a patent are those past research and development costs which were necessary to bring the patent into existence; and once they have been incurred the patent may be used in one field or ten, in one geographic area or ten, extensively or

53. It is not even necessary that the competing purchasers require functionally analogous factor inputs. Assume the existence of two groups of cotton growers who sell in a single national market but produce in geographically distinct regions. Both require factor inputs 1 through 9 in comparable amounts, but Group *A* needs a specific fertilizer as input 10 which is not needed by Group *B* because of different soil conditions. Group *B*, but not Group *A*, needs a specific insecticide. The marginal cost of a unit of insecticide and of a unit of fertilizer is the same, the unit in each case being that quantity appropriate to use in conjunction with the production of 1 bale of cotton. Thus neither Group has any real cost advantage over the other. A diversified supplier, *S*, sells both the fertilizer and the insecticide; *S* has competitors in fertilizer but not in insecticide sales. *S* will sell fertilizer at a lower markup over marginal cost than the markup he applies to insecticide and Group *B* will be at an artificial cost disadvantage. Is *S* engaging in economic discrimination? Suppose *S* now spins off its insecticide division which becomes independent Company *I* and which continues to price as before. Is *I* engaging in economic discrimination?

not at all, all without any additional cost.[54] Thus any two royalty charges, however disparate they may be from one another, bear the same ratio to the marginal cost of the product: in each case the ratio is infinite.

Another way to put the difficulty is as follows: If rights to use a patented invention were being sold under conditions of perfect competition, price would approach marginal cost of zero. If such rights were being sold by a perfectly discriminating monopolist, the price to his most favored customer would approach the marginal cost of zero. Only the existence of monopoly makes possible the imposition of any price in excess of zero. Any revenue received through successful imposition of a price is a return, either to fixed cost or of monopoly profit—the category previously described as "excess needs"; none is a return to variable cost, for there is no variable cost. Since a rule against economic discrimination is a rule that requires allocation of "excess needs" in proportion to marginal cost, the rule is not susceptible to application in the patent license context because there is no base figure against which to allocate "excess needs."

It may seem at first as though the foregoing argument is a mere mathematical trick, depending perhaps on the impermissible step of dividing by zero, with which "men of affairs" need not be concerned. "Surely," the reader may argue, "we can at least require an *equal* royalty charge in patent licenses of identical scope."

But, subject to a possible exception to be discussed shortly, these reactions are wrong; no mere mathematical trick underlies the difficulty. What does underlie the foregoing position is the arbitrariness inherent in the basic concept of price discrimination. No normative principle but only a very rough and elusive sense of equity dictates that "excess needs" be allocated in proportion to marginal cost. In the generality of situations, a rough sense of equity is enough, however; for any conceivable allocation of "excess needs" among customers is quite as arbitrary as any other, and proportionality to marginal cost, being no worse than any other, needs only to be supported by a featherweight. That featherweight, moreover, receives the support of familiarity; for in competitive markets the existence of rival sellers compels proportionate allocation. And since competition does compel that allocation, other allocations are probative of monopoly and are therefore viewed

54. This statement ignores the cost of administering the license system: the cost of negotiating the license, of collecting the royalties, etc. These costs are roughly analogous to "selling costs" of a more typical product. The existence of these costs does not impair the basic validity of the point made in the text for reasons that will be explained.

with suspicion and hostility. Finally, and of most substance, since in the generality of cases marginal cost is a very large percentage of price, proportionality of "excess needs" to marginal cost is the allocation least likely to induce, in the market structure of customers who compete among themselves, structural features inversely related to efficiency.

In the patent licensing context, however, since marginal cost is zero, some other basis for allocating "excess needs" must be found. And that basis must be sufficient in itself to justify legal enforcement of that allocation, for the reinforcement derived outside the patent area from familiarity and anti-monopoly policies are absent here. If possible, the new basis of allocation must be consistent with, and perhaps shaped entirely by, the objective of avoiding structural distortion in consumer industries, the one traditional purpose of prohibition against discrimination not yet shown to be totally vacuous in this context.

The preceding paragraphs point by inference to the exception suggested earlier. Although a patent license often is unaccompanied by any transfer of a product and consists merely of a grant of legal permission to practice an invention, the contrary is often true. When a patented product is sold or leased, a license to practice the invention at least to the extent of using the product is granted, usually by implication but sometimes expressly.[55] If an undifferentiated price (or rent) is charged for the product in such a case, it obviously comprises several elements of return to the patentee-manufacturer, the cost of the product and his monopoly return as patentee. But the price need not be undifferentiated: the patentee may put a price on the product and charge a separately stated royalty for its use, and to do so is usually the purpose of an express license accompanying such a sale or lease. The question then may arise whether the total charge to each purchaser (or lessee) must bear the same ratio to marginal cost of the product. A workable rule which imposes an affirmative answer to that question could be adopted, either a rule limited to those situations where structural distortion of customer industries was a realistic concern or a more pervasive rule. Hence the license in conjunction with the transfer of a product is an exception to the statement that the classical concept of economic discrimination can have no application in the patent license context. Whether such a rule would be desirable is, of course, another question; and an attempt to give a difinitive answer will be deferred

55. See, *e.g.*, Brulotte v. Thys Co., 379 U.S. 29 (1964); Henry v. A. B. Dick Co., 224 U.S. 1 (1912); Grand Caillou Packing Co., TRADE REG. REP. (1963-65 Transfer Binder) ¶ 16,927 (1964).

pending further consideration of royalty discrimination unaccompanied by product transfer.

I return to the situation of a license unaccompanied by product transfer and dispose briefly of two remaining objections that might be raised against the asserted proposition that the classical concept of economic discrimination cannot be applied. One might object that marginal cost is only equal to zero when the situation is viewed as of the point in time when the patent has been acquired; that all costs are variable in the long run and hence that long-run marginal cost necessarily exceeds zero and can be used as a basis for allocation. The objection fails for several reasons. First, in the theoretical sense in which the assertion of "long-run variability" is true, it affords no workable basis for assigning a particular marginal cost to a particular patent license. The assertion presupposes options about the scale at which an activity yielding a more or less homogeneous product will be conducted. It might conceivably have relevance to one establishing a research laboratory intending to produce a series of patentable inventions over a sustained period of time. But it has no relevance to royalty charges under a single patent which, by definition, is unique and may have been substantially more or less costly to develop than the next preceding or following patent.

Second, a suggestion to invoke long-run marginal cost as a royalty criterion necessarily implies parity of royalty charges for licenses under successive patents of a single patentee. Hence it is totally antithetical to the statutory scheme which has the dominant purpose of utilizing the market mechanism to determine, patent by patent, the value of the patentee's contribution and to generate rewards to the patentee commensurate with the value of each individual patent.

The basic feature of statutory purpose might be thought to have even more extreme implications. It plainly contemplates rewards based upon value of the invention to users rather than upon cost to the patentee. One might argue, therefore, that any cost-related limitation on royalties was incompatible with the purposes of the patent laws, that the law calls for royalties commensurate with value to the licensee and that "value to the licensee" is a mere synonym for the licensee's elasticity of demand. This view, though tenable, is surely not inevitable. Rewards "as great as the traffic will bear" can be permitted with respect to each individual patent without going further still and permitting royalties from each licensee under a single patent to be as high as the patentee can bargain for. The patentee can thus be permitted to take into account the elasticity of aggregate demand of all potential users of a particular invention but denied freedom to charge some users of

a patent more than other users of the same patent in accordance with their individual demand elasticities. Hence, while statutory purpose requires that each patent be looked at singly, that purpose would not be thwarted by and thus does not preclude a rule against discrimination among licensees of a single patent.

Compatibility with statutory purpose is a necessary but not a sufficient condition to the adoption of a no-discrimination rule. Since that purpose does require that any such rule be constructed on a patent-by-patent basis, the question remains whether a rule with beneficial economic impact can be articulated and applied.

One might assert that even in the short run patent-by-patent view imposed by statutory purpose, it is not true that marginal cost equals zero. At least some minimal cost to the patentee is involved in the negotiation and administration of each individual patent license. But even assuming that a figure could be derived from such costs, it would be a totally inappropriate basis by which to allocate the patentee's fixed costs and monopoly return. These "selling" costs will usually be an exceedingly small percentage of the royalty charge and they will often vary substantially from one licensee to another. Used as a mandatory basis of allocating all other costs and monopoly return, they would compel royalty structures of wild absurdity that would not comport with but contravene both the equity objectives and the customer-industry structure objectives underlying the classical concept of discrimination.[56]

Once it is conceded that marginal cost is not available as a basis for allocation (price formation) in the patent license context, the question arises whether there is some other basis on which a rule of law might

56. The obviousness of the fact that the fragmentary costs of negotiating and administering individual licenses are an inappropriate basis for allocation (price formation) might seem to suggest that the allocation problem discussed here is very different in principle from those to which the classical discrimination concept is traditionally applied. However, I think the difference is merely one of degree. The fact is that there is no very sound reason why "excess needs" should always be allocated in proportion to marginal cost. No criterion for allocation is "right" except because it conduces to a social objective; and as has been argued, the identifiable objectives of a no-discrimination rule are fairly elusive. The generally observable loyalty among economists to the classical concept probably owes much to the logically impermissible step from a descriptive statement regarding price formation under competitive conditions to a prescriptive statement regarding price formation generally. The assertion "Price should be proportional to marginal cost," is often used as a synonym for the assertion, "Prices should be formed under competitive conditions." In the context of a legal monopoly, the synonym breaks down.

Even outside the context of legal monopoly, however, the adequacy of short-run marginal cost as a criterion for price formation would seem to diminish as marginal cost became an increasingly small percentage of total cost and price and as the homogeneity of the units of product decreased.

be constructed that will minimize inequity and customer-industry structural distortion. Since the objective of profit maximization will lead, without the social costs of litigation and enforcement, to price formation in accordance with customer elasticity of demand, the question is more appropriately cast in this form: Can we identify an ascertainable basis for price formation that will more fully achieve the objectives of prohibition against discrimination than will price formation in accordance with customer elasticity?

A requirement that each licensee must be charged the "same" royalty is delusive in its apparent simplicity. I use a variation of the facts in the *Grand Caillou Packing* case[57] as an illustration. There the patentee had invented an automatic shrimp peeling machine. Prior to the introduction of the machine, a substantial part of the cost of canned, peeled shrimp was attributable to the labor cost of hand peeling. On the Louisiana Gulf, where the shrimp were comparatively large, a shrimp canning industry had developed. No significant industry emerged in the Pacific Northwest despite the abundance of shrimp in those waters; for Pacific shrimp are comparatively small. Since it takes at least as much hand labor to peel a small shrimp as a large one, the hand labor cost per pound of peeled shrimp was prohibitively high in the Pacific Northwest. Let us assume that if the patentee had issued royalty-free licenses to all applicants, a shrimp canning industry on the Pacific, freed of the prior cost disadvantage, would have emerged and thrived in competition with the Louisiana Gulf industry.[58] But the patentee was entitled to and did charge a royalty. To maximize his returns, he should have set a royalty rate on a per-shrimp basis roughly equivalent to each licensee's prior labor cost of hand-peeling a shrimp.[59]

57. Grand Caillou Packing Co., TRADE REG. REP. (1963-65 Transfer Binder) ¶ 16,927 (1964), *aff'd in part, rev'd in part sub nom.* La Peyre v. FTC, ATRR No. 271, p. x-1 (5th Cir. Sept. 13, 1966). In that case the patentee leased the patented machines to licensees for a minimum annual fee plus a royalty per machine revolution. Hence the case involved transfer of a product with a real marginal cost. In my hypothetical variation it is assumed that the patentee authorizes licensees to build and use their own machines.

58. This situation, of course, is the social optimum; for since the real social cost of an incremental use of an existing innovation is zero, any deterrent to use results in a misallocation of resources. But this misallocation, hopefully limited in duration to the life of the patent, is obviously a contemplated cost of the patent system's selection of the monopoly device is a means of encouraging innovation.

It is assumed throughout the discussion that the two groups of shrimp canners are in competition with one another and that their other inputs are supplied under competitive conditions.

59. Under certain circumstances the patentee would make greater returns by setting a royalty enough below prior labor cost to induce a reduction by his licensees in the price they charged for shrimp. This would be true if, after adopting the invention, the aggregate marginal cost curve of the industry, computed without regard for royalties, intersected the marginal revenue curve confronting the industry at an output greater

Actually, the rate to Gulf licensees would have to be slightly lower than Gulf labor costs to induce those canners to shift their production method. And although the rate for the Pacific canners per pound of peeled shrimp would be higher, their rate would have to be an even smaller percentage of their prior labor costs, they having demonstrated in the past that survival was impossible if they had to pay full hand labor costs. I will refer to these maximizing rates as *GL* (Gulf Labor) and *PL-x* (Pacific labor less an amount essential to survival). From the facts of the case we know that *PL* exceeded *GL* and that *x* was sufficiently small so that *PL-x* exceeded *GL*. In summary, the maximizing royalty structure is such that Gulf canners would have continued to use hand labor had the charge to them been any higher and the Pacific canners were just barely earning the opportunity cost of their invested capital at the rate charged to them. Such a royalty structure could be expected to result from the self-interest of the several parties without the delay or any of the private or social costs of official intervention. And as I read the *Caillou* case, just such royalty structure did in fact emerge.

Can a persuasive case be made for the legal imposition of any different royalty structure? Is there inequity or misallocation that the law might correct?[60] I can perceive no inequity between the patentee, on the one hand and, on the other, either group of licensees. The patentee has afforded the Gulf canners that minimal cost saving necessary to induce them to use his invention and has afforded Pacific canners an even greater cost saving. It is true Gulf canners are less well off than they were before. Their prior labor cost advantage has been diminished and they must now share the market for canned shrimp with the Pacific interlopers. If they were engaged in perfect competition before, some Gulf canners may now fail; if not, prior monopoly profits will not again be earned to the same extent. But surely losses such as these do not warrant legal remedy.

The Pacific canners, on the other hand, while better off than before, may be less well off than they will be in the future when patent protection expires. It may be that they have some cost advantage over the Gulf canners which has heretofore been rendered inoperative by the dominating cost-of-labor disadvantage; any such advantage will now be absorbed by the patentee through the royalty charge and will inure to

than that at which the prior marginal cost curve, including labor costs, intersected the demand curve.

I have assumed that this circumstance did not exist. A contrary assumption would not affect the substance of the analysis but would affect its verbalization.

60. There is, of course, the misallocation referred to in note 58 *supra*; but that is inherent in the monopoly-subsidy approach of the patent law.

their benefit only when the patent expires.[61] Does the present denial to Pacific canners of the private benefit that is potential in such a cost advantage give them a basis for complaint of discrimination vis-à-vis the Gulf canners? If so, my equal protection antenna fails to perceive it. If there is a problem here, it seems to me to be one of resource allocation, an issue which remains to be considered, and not one of equity. Arrayed against any equity argument are the following points: (1) Unlike the Gulf canners, the Pacific canners are better off than before if worse off than hereafter; why is hereafter a more persuasive touchstone of fairness than heretofore? (2) The cost advantage of the Pacific canners was rescued from total irrelevance by the innovation of the patentee; the advantage is a measure of the innovation's value and seems as justly captured by the patentee as by the licensee. (3) If the Pacific canners operate under competitive conditions, the private gain potential in the cost advantage will be short-lived in any event because it will signal entry or expansion in the Pacific industry segment and contraction in the Gulf.

The last observation places emphasis where it belongs—not on equity but on resource allocation between the Gulf and Pacific segments of the industry. The optimum allocation is that which would prevail if the peelers were available to each segment royalty-free, but this precise allocation will be precluded by any royalty charge.[62] The question therefore can again be refined: given the fact that any royalty charges will force a departure from optimum resource allocation by the industry as a whole—it will under-produce relative to other industries—is it nevertheless possible to describe a royalty structure

61. If the heretofore inoperative cost advantage of the Pacific canners were sufficiently great, it would be in the interest of the patentee to license only those canners and supply the entire demand for shrimp through them. This circumstance would imply that the former labor cost disadvantage was very substantial—it necessarily exceeded any present cost advantage—and hence that the invention is of very substantial value to the Pacific canners. But even under this circumstance the patentee will absorb much of the now operative cost advantage and allow Pacific canners only the return on capital essential to induce production.

62. If both segments of the industry were engaged in production at the time of patenting, then, in the short run, a royalty charge would only distort resource allocation from the post-patent optimum if it affected short-run marginal cost. Traditional royalty structures, based on units of output or licensee sales revenue do affect those costs. Lump sum royalty charges, although they do not affect short-run marginal cost, do not offer a practical alternative however. First, except in the short-run they are equally distortive. Second, if as in the shrimp-peeler case, one segment of the industry is more a potential entrant than a present competitor, the distorting effect will occur even in the short-run. Third, if it were required that the lump-sum demanded of the several segments be equal in dollar magnitudes, an artificial economy of scale would be created which would distort allocation and which would be politically unacceptable; and if the several sums need not be equal then the entire problem of discrimination will be reintroduced into the cost structure of the licensee industry.

which will not make matters worse by distorting allocation between the several segments? To speak of distortion presupposes a norm, and it should be explicit: production at minimum average cost per unit of output is desired. If either industry segment can produce all that society wants at a cost per unit below the unit cost at which the other segment can produce any units, then the first industry segment should produce all units. If the two segments jointly can produce the desired aggregate output more cheaply than either segment alone, then each should be producing that part at which the aggregate costs of both are minimized.

One's first impression may be that the task of dictating a royalty structure which will not distort the division of production between segments is quite simple: it is merely necessary, it might seem, to require royalty charges to be equal; then any other cost advantage will make itself felt in the process of production allocation through competition. Equality in such a context, however, is a very elusive criterion. Should each segment pay an equal number of dollars during the 17-year patent life? Regardless of its output? Regardless of the number of years it was engaged in production during the 17-year period? Equal dollars, then, per shrimp? Per pound of raw shrimp? Per peeling machine? Per machine revolution? In the *Grand Caillou* context, and in many patent licensing contexts, each of these apparently equal royalty requirements would have had very different economic effects, and no one of them has any normative justification. Each is equal only in its impact on the unsophisticated ear, and totally arbitrary in the only important sense of having the capacity to distort allocation of production between the segments. Deprived, as we are in this context, of the criterion of marginal cost to the seller, alternative criteria for price formation must be examined with considerable skepticism.

I return to the question not satisfactorily answered by the fruitless pursuit of audio-equality: given that the output of the licensee-industry will be restricted below optimum by any royalty charge, is it possible to describe in generally applicable terms the characteristics of a royalty structure which will achieve a second-best solution to the allocation problem? If the phrase "second-best" is taken in its technical sense, the answer is, no, not except by a linear programming solution based in each case upon the cost, demand and cross-elasticity functions for both segments of the industry. Such a solution requires that these functions be known in their entirety; and such knowledge will never be available as a practical matter.

It is possible to say, on the other hand, that a royalty structure which induces to the same percentage allocation of market between the segments of the industry as would prevail if royalties were zero is preferable to a royalty structure dictated by patentee maximization. In more cases than not the market allocation will approximate the technical second-best solution more closely under the former royalty structure than under the latter. Given that royalty structure, the absolute output of both segments will continue to be smaller than under the royalty-free optimum. One cannot generalize whether utilization of the invention would increase or decrease.[63] But since cost savings in one segment of the industry would be preserved, at least in part, a tendency toward greater aggregate licensee endproduct output would exist under such royalty structures that does not exist under structures dictated by patentee maximization.

Thus, since the second-best allocation is theoretically determinable only in each specific case and not generally, and since, as a practical matter, it cannot be determined even case by case, the best normative rule regarding royalty structure that I have been able to identify is one which would achieve the same percentage market allocation between the segments as would prevail under royalty-free conditions.

Can even so limited a rule be articulated in an administrable form? A royalty charge to each segment of an equal percentage of that segment's total variable costs would not give precisely the desired result. The resulting change in output will be a function of the slopes of the marginal cost curve and of the marginal revenue curve, and the respective slopes for the two industry segments need not be identical. Similarly, a royalty charge of an equal percentage of the gross sales of each segment may not effect precisely the same percentage of output reduction in each industry segment; again the change in output will be a function of the slopes of both marginal cost and marginal revenue. But since cross-elasticity between the demands in the two segments is implicit in the problem, the slopes of the several demand curves are unlikely to be markedly different; and therefore, unless the technology in the two segments is markedly different so as to produce marginal cost curves of very different slopes, a royalty structure based on either a sales percentage or a variable cost percentage might serve as a workable approximation.

Emphasis is warranted on the fact that it is not the absolute level or amount of cost but the slope, or rate of change, of the marginal cost

63. See Appendix, *infra*, at note 5.

curve that is one of the determinants of output restriction as compared with the royalty-free optimum output. Very substantial differences between cost magnitudes of the two segments may exist, and it is the central objective of a rule governing royalty structure to assure that such differences *do* affect the market allocation that emerges between the two segments of the industry. If such differences in cost magnitude do exist, their impact on inter-segment allocation will not be substantially deflected by such differences in marginal cost slope as are likely to exist.[64]

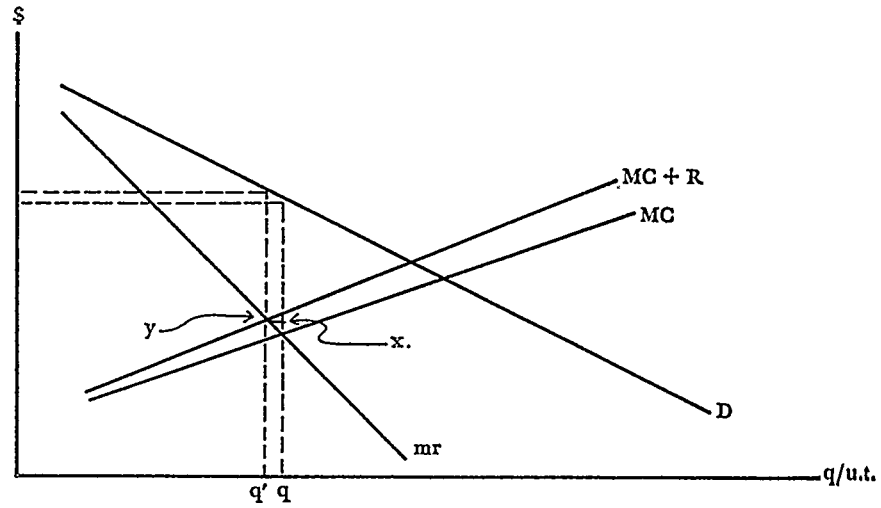

FIG. a.

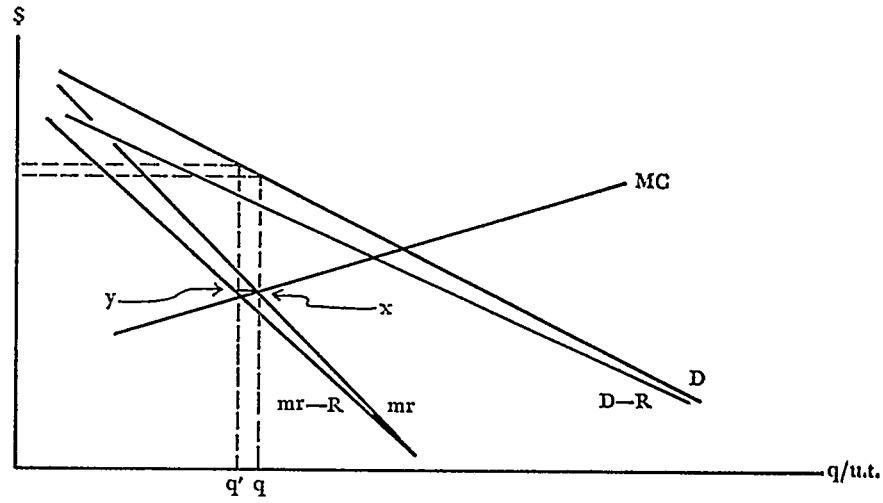

FIG. b.

64. These diagrams illustrate some of the relationships discussed in the preceding paragraphs. In figure (a) a royalty of 10% of variable cost raises the position

Tentatively, then, one might conclude that desirable economic objectives would be achieved in the generality of applications by a rule forbidding royalty structures which imposed on different segments of an industry charges that were disproportionate to the sales or to the variable costs of the respective segments. Since the figures to which the royalty percentage is to be applied are those that prevail after the industry is using the invention and since sales figures are more easily determinable and less likely to be regarded as confidential by the licensee than are variable cost figures, a royalty percentage based on dollar sales would constitute the more desirable of these two approaches.

However, I will show in the next section of this article that royalty structures based on endproduct sales generally will have a more restrictive impact on licensee output than a royalty structure based on an input selected to measure use of the invention. The argument of the next section applies equally to a royalty structure based on variable cost of all inputs other than the invention. The restrictive effect of endproduct royalty structures would introduce a tendency in opposition to that toward greater output introduced by the preservation of cost advantages in one of the industry segments. In my opinion the restrictive effect of endproduct royalty structures would fully offset any increase in output that might result from preserving the cost advantage. Certainly it would nullify to a substantial extent any tendency toward increased output. For this reason, I conclude that any attempt to forbid royalty structures based on licensee elasticity of demand in cases not involving transfer of a product having substantial marginal cost would involve costs of administration and enforcement far in excess of any gains that might be derived. Hence no such attempt should be made.

In fact, however, the *Grand Caillou* case involved transfer of a product, a lease of the patented peeling machine. And although FTC counsel urged that patentee be required to file a royalty structure based on endproduct sales, the Commission rejected the request, at least tentatively, and ordered only that patentee submit a new royalty structure

of the marginal cost curve by 10% of its absolute magnitude at every point and results in an output restriction of q-q'. In figure (b) a royalty of 10% of sales lowers the demand, or average revenue, curve and the corresponding marginal revenue curve by 10% of its absolute value at every point and results in an output restriction of q-q'. In each case it can be seen that the change in output is the line xy, the altitude of an approximate triangle two sides and the apex of which are determined by the slope of the marginal cost and marginal revenue curves near the point of their intersection.

In both cases it can be seen that dramatic alterations in curve *slope* would be necessary to cause substantially different changes in the percentage of output restriction.

free of the objectionable "discrimination."[65] What type structure would be deemed by the FTC to comply with that order is therefore open to conjecture.

In a case such as *Grand Caillou* which involves transfer of a product, the patentee does have a marginal cost upon which the law could base a requirement that "excess needs" be allocated proportionately. That fact does not answer the question whether such a requirement should be imposed in those cases.

Surely the answer to that question should not depend on whether or not the patentee states the purchase or rental price for the machine separately from the royalty for its use subsequent to transfer. Nor should it depend upon whether the transfer is a sale or a lease. The first distinction is wholly verbal, and the second may or may not have significance for other purposes but is surely irrelevant in the present context. To apply the act to sales but not leases would merely encourage sellers to distort their distribution arrangements to achieve insulation. In my view, the language of Robinson-Patman is broad enough

65. Grand Caillou Packing Co., TRADE REG. REP. (1963-65 Transfer Binder) ¶ 16,927, at 21,982 (1964). The patentee sought judicial review of the FTC order and apparently has not filed a new royalty structure with the Commission pending review. 3 TRADE REG. REP. 25,901.

Apparently patentee has in fact renegotiated its royalty structure, however. See Laitram Corp. v. King Crab, Inc., 245 F. Supp. 1019 (D. Alaska 1966). The new structure is not described in the published opinions.

The Court of Appeals for the Fifth Circuit affirmed that part of the FTC order which relates to the problem discussed in this paper. La Peyre v. FTC, ATRR No. 271, p. x-1 (5th Cir. Sept. 13, 1966).

The FTC order is described differently in the Court of Appeals opinion than in the report of the FTC opinion. In the court opinion, *supra*, at x-3, it is stated that the Commission's order defined discrimination as ". . . rental terms which result in any lessee paying a higher rate than the rate charged to any other lessee for use . . . for the same period of time or through the same number of mechanical revolutions . . . ."

Read literally, this definition suggests that the rate must be equal when computed against a machine revolution rate base *and* when computed against a time—presumably calendar time—basis. Unless all licensees use the machines an equal number of revolutions per unit of calendar time, compliance with the FTC order, thus interpreted, would seem to be impossible. Therefore the order probably should be interpreted to require rate equality when computed against at least one of those rate bases.

The proper requirement is that rates be equal when computed against a base that corresponds closely to the useful life of the machines. In theory, any differences in delivery or maintenance costs borne by the patentee-manufacturer should also be reflected; but as a practical matter of administration, unless such differences are large, they might well be ignored.

In assessing useful life, technological obsolescence as well as physical exhaustion should be taken into account. Hence both the calendar-time base and the machine-revolution base have *prima facie* validity. Whether one of these bases has substantially greater validity than the other, or whether some third basis would be better than either, would depend on issues of fact and prediction not easily resolved.

Minimization of both private and public costs of administration dictates that a royalty base selected by the licensing parties be accepted unless it obviously correlates less well with useful life than does some other practicable alternative. Once the FTC has intervened in accordance with this standard, as in *Grand Caillou*, FTC selection of an appropriate royalty base should be accorded similar deference on review.

to cover all the situations if coverage is desired;[66] but if that language is thought confining, then resort should be made to the spaciousness of § 5 of the Trade Commission Act to achieve parity among these several situations.[67]

To require prices proportionate to marginal cost where a product transfer is involved is, as a matter of practical administration, to require prices proportionate to physical units of product transferred subject to a cost-saving defense by a patentee who asserts that different units of product had different costs. To impose that requirement is, in short, to assert the desirability of applying to these situations the Robinson-Patman Act, either directly or via the assimilation of § 5 of the Trade Commission Act. Surely our experience with that baleful statute and its past administration affords little reason to believe that administration in the present context would be highly satisfactory.

Although the question is a close one, on balance I think allocation of "excess needs" in proportion to cost should be required. First, in a case involving a sale of a patented product, it would be very difficult to construe Robinson-Patman as being inapplicable. Both the language of the act and its purpose, however ill-conceived that purpose may be, embrace the situation supposed. One might argue that the monopoly creating the potential for discrimination here is legally conferred and hence the situation is not of the type at which the act was aimed, but the argument is not persuasive. The dominant purpose of Robinson-Patman, as opposed to old section 2, was to protect the structural integrity of the customer industry; and legality of the seller's monopoly is irrelevant to that purpose. Moreover, while the patented product in the type of case now under discussion will usually be a durable good not intended for resale, the logic of the "legal monopoly" argument extends equally to sales for resale of nondurables—patented drug products for example. Certainly no persuasive argument could be made for exempting sales of the latter type.

Second, application of Robinson-Patman to transfers of patented durables is less likely than most applications of the statute to be attended by the undesirable side effect of rigidifying price structures and impairing competitive processes. The very short-run price fluctuations and geographically spotty price competition which do and should occur

66. 49 Stat. 1526 (1936), 15 U.S.C. § 13 (1964).
67. Federal Trade Commission Act § 5, 38 Stat. 719, 15 U.S.C. § 45 (1964).
In the *Grand Caillou* case the FTC based its order on § 5, apparently because the Robinson-Patman Act was thought inapplicable. Grand Caillou Packing Co., TRADE REG. REP. (1963-65 Transfer Binder) ¶ 16,927, at 21,992 (1964) (concurring opinion).

in the distribution of nondurable consumer goods are unlikely to occur in the case of patented durable goods; and Robinson-Patman enforcement thus is less likely to impair such competition as does exist in the markets for such goods.

Third, when discrimination does occur with respect to patented durables, it is likely to be of the enduring type which may affect significantly the structure of customer industries. A requirement of proportionate allocation of excess needs will diminish the patentee's ability to absorb cost advantages extraneous to the invention that may be possessed by one segment of the customer industry. For all these reasons the argument for applying Robinson-Patman standards to these cases is probably stronger than for applying them in the typical Robinson-Patman case.

Application of Robinson-Patman standards, even assuming flawless administration of the process, is in no sense an ideal solution: it is merely better than any other practicable alternative that occurs to me. It is better than allowing patentee maximization through royalty charges scaled to demand elasticity, for it will diminish the patentee's ability to absorb a potential cost advantage available to a segment of the industry. It thus will permit a partial shift of resources toward the more efficient segment, a shift in the direction of that allocation between segments that presumably will prevail after patent expiration. This shift should achieve, in the generality of cases though not necessarily in any particular case, a better resource allocation. Second, it is better than allowing the patentee to base royalties on the unpatented endproduct. Although an endproduct royalty structure would induce an even closer percentage approximation of post-patent resource allocation between the industry segments, that royalty structure maximizes the restraint imposed on the endproduct market for reasons explained in the next section.

Since *Grand Caillou* did involve product transfer, the result reached by the FTC seems justifiable although the majority opinion furnishes no support either for the result or for the traditional *raison d'être* of administrative agencies.[68] Rejection of the endproduct royalty structure suggested by its staff was sound. Whether the Commission will now

68. The Commission explicitly rejected direct testimony that the royalty structure was adopted by reference to the saving in labor cost that the invention afforded. This rational and highly probable explanation was held "not worthy of belief." *Id.* at 21,976. The Commission found, instead, that the structure was intended "to protect . . . (patentee's) own interests as shrimp canners." *Ibid.* That explanation is not only wholly irrational but the evidence cited to support it is unconvincing. *Id.* at 21,977. Commissioner Elman, in his concurring opinion, noted this error in the majority's analysis, but to no avail. *Id.* at 21,991-92.

reduce the entire proceedings to absurdity by accepting a new royalty structure that has only audio-equality to recommend it remains to be seen.

V. Substitutability of Factor Inputs

Like the problem of discrimination, a second economic phenomenon that has received inadequate attention by legal scholars is involved in much patent license litigation; I will refer to it as input substitutability.[69] A producer will use various goods and services, or inputs, in his production process including land, products produced by others, capital and labor; and at any given point in time he will be using them in a certain proportion to one another—in accordance with his current recipe. For simplicity assume the producer uses only two inputs, *a* and *b*. Even if his recipe remains unchanged, his consumption of each *a* and *b* will increase or decrease as his own output fluctuates; and in all but most extraordinary cases he also has the possibility of varying his recipe—using more *a* and less *b* to produce any given output. Obviously there are limits on the variability of production recipes: a given amount of corn can be raised by using more fertilizer and less land, but only up to a point. The law of diminishing returns states that increases in one input while other inputs are held constant will yield greater total outputs up to some ratio and then outputs will start to decline. The economic term for that change in output that results from an incremental change in the quantity of a single input, the amount of other inputs being held constant, is the *marginal physical product* of the varied input. Input *a* will be used in a proportion to other inputs never exceeding the ratio at which marginal physical product of *a* becomes negative (a further increase in input *a* results in reduction in output), and never less than the ratio at which the marginal physical product of other inputs becomes negative (a reduction in other inputs

69. The role of this concept in legal control of economic activity has been noted in several sources of which I am aware. In *Tying Arrangements and the Leverage Problem*, 67 YALE L.J. 19, 25-27 & nn. 19-20 (1957), Professor Bowman alludes to the concept; but his examination is exclusively from the seller's viewpoint in a tie-in context and fails to make explicit the implications for resource allocation at the level of the buyer and the buyer's customers.

Professor Merton Burstein discusses the general problem at length in his splendid article, *A Theory of Full-Line Forcing*, 55 NW. U. L. REV. 62 (1960). He also treats the problem in a shorter, more technical piece, *The Economics of Tie-In Sales*, 42 REV. OF ECON. & STATISTICS 68 (1960).

In *Tying Arrangements in Law and Economics*, 8 ANTITRUST BULL. 743 (1963) Professors Baldwin and McFarland discuss the problem in the tie-in context. Like Bowman, *supra*, they neglect the more general application of the concept and, by dealing with tie-ins of consumer products, leave obscure the significance of the concept for resource allocation.

results in increased output). Within this range, the producer obtains output at minimum cost per unit of output by a recipe such that the ratio of each input's marginal physical product to its cost is the same as that for every other input.[70] It follows that any change in price of one input will dictate a change in the production recipe so as to re-establish the parity of ratios between marginal physical product and price for all inputs. In the simple two input case, assume the producer is using a recipe of 3 units of *a* to 1 unit of *b*, the unit cost of *a* being $1.00 and that of *b* $2.00, and their respective marginal physical products being 5 and 10 at that ratio. If the unit cost of *a* now increases to $1.25, he will use more than 1 unit of *b* for each 3 units of *a*: by increasing the absolute amount of *b* used or decreasing the absolute amount of *a*, he will increase his proportionate use of *b* until its marginal physical product falls and that of *a* rises to values proportionate to prices, at which point parity is re-established.

The foregoing paragraphs may first appear to be no more than a restatement, in unnecessarily complex form, of the rudimentary proposition that use of a product will vary inversely with its price, *ceteris paribus*. It is true that the two propositions are closely related, but the foregoing statements about factor inputs permit a further conclusion. Assume that patented factor input *a* has no use other than as an input of unpatented endproduct *x*. The quantity of *x* that can be sold will vary inversely with the price of *x* with some degree of elasticity. Since the consumption of *a* is derived from the demand for *x*, demand for *a* will vary in direct proportion to that for *x* if the production recipe of *x* remains constant. But if the cause of a price increase (and quantity decrease) of *x* is a production cost increase attributable to input *a*, the percentage change in the demand for *a* will necessarily be at least as great and in most cases will be greater than the percentage change in demand for *x*. The quantity of *a* consumed will diminish proportionately apart from any recipe change and will diminish further to the extent that input substitution occurs. Conversely, a reduction in the price of input *a* will stimulate an increased use of *a* more than proportionate to any resultant increase in *x* output. The situation can be summarized by the statement that, given input substitutability, the elasticity of demand for a factor input is always greater than the re-

70. This recipe is a theoretical ideal and, at best, can be achieved only in the long run. For it implies freedom to vary all inputs, some of which will be fixed in the short run. And it assumes that all inputs can be varied by appropriately small increments, whereas in practice problems of indivisibility will be encountered. Neither of these qualifications, however, substantially impairs the basic point made in the text.

sponsiveness of endproduct output to a price change of that same factor input.

In this context one must be very wary of the terms "complement" and "substitute." In general usage two goods are called complements if an increase in the price of one leads to a decrease of consumption of the other—typically because they are used extensively in conjunction with one another, like nails and lumber. And two goods are called substitutes if an increase in the price of one leads to an increase in the consumption of the other—typically because they are widely substitutable for one another, like copper and aluminum. Goods which are complements to one another in this general sense nevertheless will be technical substitutes for one another, within limits, in most of their joint uses; and the foregoing analysis will apply to them.

Substitutability of factor inputs has obvious implications for one who has a monopoly on a factor input. If he can devise a way to block, or to lessen the incentive of his customers to engage in, factor substitution, he can lessen the elasticity of demand for his patented product and increase his monopoly profits. In the case of my two-input endproduct x, the theoretical limit on the monopoly profits of a patentee of input a (a having no other uses) is the consumer surplus area under the demand curve for unpatented x. That area necessarily equals or, given substitutability, exceeds the consumer surplus area under the demand curve for a.

There are various ways the monopolist of a can garner larger monopoly profit. One is by vertical integration: he buys input b himself, uses a and b in proportions indicated by his costs for each to produce x and sells x at the maximizing price indicated by the intersection of his marginal cost curve and the marginal revenue curve for endproduct x.

Precisely the same result can be achieved by devices not requiring entry into the process of producing and selling unpatented x. A second method is to buy b on the competitive market and tie sales of a to re-sales of b, either in a fixed proportion of a to b as dictated by the ratios of their marginal physical product to his costs, or at prices which induce the x producer to use them in that same ratio. The tie-in device, thus used, will call for a lower nominal price for a and a higher nominal price for b than would prevail had the monopolist been limited to earning his profit by exploiting the demand for a. Consequently more a and less b will be used in producing x. Since less b is used and more is nominally paid for it, one can permissibly

say that the monopolist has "extended" his *a* monopoly to *b*.[71] Where the tied and tying products are the only inputs as in my simplified example, no possibility of input substitution remains; and by correct pricing of inputs the *a* monopolist will bring about the same higher price and lower output of *x* and will garner the same profits for himself as would a monopolist of unpatented *x*. If there are other inputs, the monopolist of *a* will have succeeded in that objective to the extent that he has selected a tied product *b* against which inputs *c*, *d*, . . . *n* cannot be substituted to a substantial extent in the *x* production process. Hence, it may be more helpful to characterize the situation as one in which the monopolist of the tying product has extended his monopoly to the unpatented endproduct *x* rather than to input *b*.

The tying device has great potential usefulness to the patentee. Since incremental uses of his invention are costless to him, they should be used—by him if he is integrated forward to the endproduct market or, if not, by his customers if he is to obtain maximum revenues—up to the point where the marginal physical product of the invention becomes zero. In short, the patentee should make incremental use by his licensee costless and should derive his revenue by levying a tax on some other basis. To tie a different factor input, one not extensively subject to input substitution, is a simple solution. However, the emergence of a very rigid legal prohibition against tie-ins has placed this solution beyond the patentee's reach.

There is still a third solution for the patentee, however, and anomalously it is the most time-honored and legally blessed of royalty structures: a royalty based on a percentage of licensee sales of an unpatented endproduct. Plainly the licensee cannot substitute against his own output other than by restricting it and raising his prices just as a monopolist would. Under an endproduct royalty structure, a licensee pays nothing for incremental use of the invention in the production of any given output. And through royalties the patentee may drain off the full monopoly potential inherent not in the invention

71. See Bowman, *supra* note 69, at 25-27. The tying cases repeatedly use the characterization, "extending the monopoly" to the tied product, but they do so with equal glibness in contexts which plainly do not involve the phenomenon under discussion as they do in contexts which may. Hence it is fair to conclude that in present judicial use the phrase is gibberish—a confused and confusing way of stating the conclusion of illegality. Mercoid v. Mid-Continent Inv. Co., 320 U.S. 661, 664-65 (1944); Ethyl Gasoline Corp. v. United States, 309 U.S. 436, 456 (1940); Carbide Corp. v. American Patents Dev. Corp., 283 U.S. 27, 31 (1931).

There are, of course, other undesirable potentials of tie-in sales, principally economic discrimination among purchasers and establishment of entrenched positions in the market for the tied product which may be longer-lived than the tying-product monopoly. The foregoing criticism of case law is not intended to imply general disagreement with the judicial prohibition.

but in the unpatented endproduct of which the invention is only one of many inputs.

There is a limited sense in which the consequences of sales percentage royalties are desirable: resource allocation btween factor inputs in the given production process are not distorted. Each production process does have a socially optimum recipe. It is that mixture of inputs at which the marginal physical product to price ratio of each input is the same, and each input price equals the real social cost of producing the last or marginal unit of that input. In a process involving an existing invention a and other inputs $b, c, \ldots n$, economic waste attends the substitution of costly units of b, c or $\ldots n$ for costless incremental units of a. Assuming that $b, c, \ldots n$ are produced under competitive conditions, the optimum recipe results from a sales percentage royalty.

A choice of evils is thus presented. If the patentee bases his royalty charge on incremental uses of input a, the consequence is partial monopolization of the endproduct x and an inefficient production recipe which under-utilizes a and consumes $b, c, \ldots n$ in proportions, though not necessarily in absolute amounts, in excess of optimality. If through vertical integration, or tie-in, or a sales percentage royalty, the patentee blocks the substitution effect, the consequence is that x is monopolized to an even greater degree, and all inputs are underutilized although in an efficient production recipe. Since one or the other of these evils necessarily attends the patent monopoly, is there a rational basis for a general choice of one over the other? I believe that a general choice in favor of greater x output is appropriate.

I can offer no formal proof of my answer to this question, but I believe that something more than a purely intuitive answer can be given. In all situations of the type under consideration, the greater impairment of social welfare that attends the more extensive monopolization of endproduct x argues for limiting the patentee's monopolization strictly to incremental uses of the inventive input a. Unless some offsetting social disutility attends that choice of royalty structure, that choice should be preferred. Prima facie it may appear that the less efficient production process and the more extensive consumption of inputs $b, c, \ldots n$ resulting from that choice is an offsetting disutility. The magnitude of the disutility will be considered both from the standpoint of x production and from that of the rest of the economy.

From the standpoint of the x production process the answer is implicit in the fact that output is greater and more nearly optimum if only input a is monopolized. Under either type of royalty structure,

the aggregate private cost of output of x is greater than the aggregate social cost, the difference between the two in either case being royalty payments for a which has no social cost. Hence under either royalty structure, output is suboptimal and the social utility of additional x output would exceed its social cost. A shift from an endproduct royalty structure to a structure based on input a will increase x output by some percentage and result in a greater percentage increase in consumption of $b, c, \ldots n$. The social utility of $b, c, \ldots n$ in their alternative uses, from which they are now drawn away, will be reflected in their prices; and their prices will be a primary determinant of the extent to which they are now substituted for a. Their value in x production is diminished by the inefficiency of the production recipe, and they will be drawn away from alternative uses only to the extent that their value in x production, thus diminished by inefficiency, nevertheless exceeds their value in alternative uses. From the standpoint of x production, then, the apparent disutility of the inefficient recipe has already been taken into account; the social gain attributable to greater x production is net gain and need only be weighed against disutility elsewhere in the economy imposed by the incremental consumption of $b, c, \ldots n$.

Whether there is disutility elsewhere in the economy depends on the slope of the supply functions for $b, c, \ldots n$. Will the prevailing price charged to other users of $b, c, \ldots n$ rise, fall, or remain constant as a result of the increased consumption of those inputs in the production of x? Assuming the inputs are produced under competitive conditions, their prices will follow the incremental social costs involved in supplying the additional units consumed in x production. If these incremental units of $b, c, \ldots n$ can be produced as cheaply as the next preceding units, there will be no disutility to set off against the social gain of greater x output. If they can be produced more cheaply—if, for example, there are returns to large scale still to be realized in their production—there will be no disutilities but rather gains elsewhere in the economy which add to the social gain of greater x production. But if the incremental units have greater costs than preceding units, then the prices of $b, c, \ldots n$ will rise to all their users and there will be disutilities which may be equal to or greater or less than the utility gained by increases x production.

In any particular situation, then, a shift from one royalty structure to the other might have desirable or undesirable consequences; and in the particular case one could ascertain the consequences only if he had knowledge far beyond that likely to be available. In the

totality of situations covered by a legal rule forbidding one royalty structure or the other, confident prediction is less difficult. If patterns of rising, falling and constant marginal cost of input factors appeared with equal frequency and degree, the royalty structure yielding greater x output obviously would be preferable; for there would be in each case the utility gain to x users and the consequences elsewhere would, in the generality of cases, offset one another. But rising supply functions are probably more common than falling functions; and since that state of events is least hospitable to my analysis, I assume it to be true. Given that assumption, less than all the gains of greater x output are likely to be realized by such a rule.

Since one is forced to concede that some indefinite fraction of the social gains of greater x output will be lost through rising supply functions elsewhere in the economy, one might question the assertion of a confident conclusion that the fraction lost will be less than and not greater than one. In a more technical context that question would call for a general examination of the developing literature on the theory of second best,[72] but such a treatment would be out of place here. A simpler answer, one that proceeds by analogy, is offered.

Why has the law set itself against monopoly at all? The effect of monopoly is to lessen output, raise prices, increase returns to producers and diminish social utility in the particular market monopolized; hence it is generally condemned. But monopolistic restrictions on output in a particular market reduce consumption of factor inputs, and elimination of monopoly increases their consumption. Most considerations of monopoly satisfy themselves with condemnation of the utility loss in the restricted market and ignore the external consequences attributable to the supply functions of inputs. But if rising input supply functions are assumed to predominate, as realism compels, then welfare gains in the monopolized market resulting from elimination of monopoly are not inevitably net gains. Yet, for good reason no one doubts that net welfare gains generally are realized. In part this confidence is based on the not wholly persuasive factor that the gains in the monopolized market are immediate and measurable whereas the external losses are remote and indefinite. More soundly, the confidence stems from conceptions of resource allocation. Monopolization of x causes under-utilization of endproduct x and under-utilization of factor inputs in x production. Rising supply functions of those inputs there-

72. See, *e.g.*, Lipsey & Lancaster, *The General Theory of Second Best*, 24 REV. OF ECON. STUDIES 211 (1957); Mishan, *Second Thoughts on Second Best*, 14 OXFORD ECONOMIC PAPERS 205 (1962).

fore lead to utilization of those inputs elsewhere in excess of optimality. Hence, it is soundly assumed that welfare gains realized in the monopolized market upon destruction of monopoly will exceed diminutions of utility elsewhere which result from reductions in output from their present excessive rates.

Of course, in the world of pure theory, in which all segments of the economy except the malignant monopoly sector under attention are subject to perfect competition, there can be absolute confidence in the desirabilty of destroying monopoly; for then the system proceeds forthwith to Pareto optimality and perfect resource allocation. But no sensible economist or economically literate lawyer supposes that is the justification for attacks on monopoly in the real world. Defects remained elsewhere in the economy after Alcoa's aluminum monopoly was broken and Pareto's conditions were not met. Yet surely allocations were shifted in the right direction and the overwhelming probability is that net welfare gains were realized.

The patent problem is different only in that the remaining defect—the continued patent monopoly—is more apparent, having been a stipulated part of the problem from the outset, and in that the relevance of supply functions of other factor inputs is more apparent—the problem being one of input substitutability.

VI. Royalty Structure

The foregoing analyses of discrimination and input substitutability obviously have implications for royalty structure. As I observed earlier, the principal implication of each phenomenon conflicts with that of the other. In my discussion of discrimination I concluded that, since incremental uses of an existing invention—essentially knowledge—had no marginal cost, the classical test of discrimination, proportionality of unit price to marginal cost, could have no application. I concluded further, however, that one of the major evils of economic discrimination, the potential misallocation of resources between segments of an industry, could stem from patent royalty structures and could be avoided by requiring that the units upon which the royalty percentage was based be units of endproduct as measured by sales.

In the discussion of input substitutability I concluded that exploitation of the patent monopoly occasioned less damage to the economy if the monopoly charge was attached to incremental utilization of the patented input than if attached to incremental units of output. The apparent inconsistency between these two conclusions calls for examination.

Each conclusion is sound for a category of situations, and the apparent conflict between the conclusions disappears if the category to which each pertains is defined more precisely. In my discussion of input substitutability, I examined the effect of the two royalty structures on the output, prices and input recipes of "a producer of x." This might be a single producer, or it might be a large number of producers provided that their production technologies were very similar. Under these circumstances the feasible range of input substitutability for each would be similar, and all would respond to input price changes by making similar recipe changes. The input substitutability analysis is valid, in short, when the demand elasticities of the licensees are roughly equal and the potential for profitable discrimination between licensees is therefore lacking.

In the discrimination analysis I examined the effect of different royalty structures on the prices, outputs and inputs of "two segments of an industry" in one of which the invention had significantly greater potential for cost savings than in the other. Under these circumsances, where production technologies differ substantially in a relevant respect and hence give rise to different demand elasticities for the invention on the part of different industry segments, the discrimination analysis is valid.

Two further points of clarification are necessary before turning from these economic analyses to the legal rules that might be founded on them. The discussion of input substitutability assumed that it was a practical possibility for the patentee to base his royalty charge on incremental uses of the invention. Essentially an invention is always knowledge, and measuring the extent to which it is being used is elusive. But generally the invention, in application, has some physical embodiment—a machine whose revolutions can be counted, a chemical whose quantity can be expressed in grams or gallons. But this is not invariably so: process and combination patents pose difficult problems in this regard,[73] and often force the patentee to base his royalty on some physical input less directly related to the invention itself than was the peeling machine in the *Grand Caillou* case. In any event, it will be the physical unit on which the royalty is based, often but not necessarily a patented product, whose use will be restricted by the monopoly charge and for which substitutions will be made in the production recipe. Whenever there is a choice of physical units to which

73. This difficulty was the practical problem that underlay the legal problem in the *Mercoid* cases. Mercoid Corp. v. Mid-Continent Inv. Co., 320 U.S. 661 (1944); Mercoid Corp. v. Minneapolis-Honeywell Regulator Co., 320 U.S. 680 (1944).

the royalty charge might be applied, the patentee will attempt to select that unit least susceptible to substitution; and to the extent he succeeds he will have achieved the effects of imposing a royalty on endproduct.

Secondly, the physical input adopted as proxy for the invention will invariably have some real social cost, if only its transportation to the point of use. In my input substitutability analysis I treated incremental use of the invention as being socially costless; and the invention as opposed to its physical proxy is costless. This factor causes no complication so long as the patentee does not supply the physical proxy. Then, presumably, it will be supplied by another under competitive conditions, and the costs of the proxy and of using the invention will be separately stated. But if the patentee supplies the proxy, then a difficulty arises of separating the real social cost of the proxy from the monopoly charge for practicing the invention.

The foregoing considerations suggest the possible desirability of adopting either or both of two rules limiting patentee freedom to shape his royalty structure: (1) When granting licenses to each of a group of licensees who are in competition with one another, if the technology of one subgroup differs significantly from that of another subgroup such that their demand elasticities for the invention differ, the patentee must base his royalty percentage on the endproduct sales of his licensees. (2) When granting licenses in circumstances other than those described in rule (1), a patentee must base his royalty charge on the licensee's consumption of a particular input or group of inputs closely associated with the practice of the invention.

Adoption of both rules would be desirable in some ideal society where addressees of rules always complied in good faith and without significant cost to themselves in ascertaining what rule applied to their immediate situation. It is very doubtful that adoption of both rules is desirable under more realistic conditions. The two rules call for precisely opposite courses of conduct—one for an input-based royalty and the other for an endproduct royalty. The conduct called for in each situation is opposed to the patentee's financial interest: in each type of situation the forbidden royalty structure would maximize royalty income. And the dividing line between the two categories—substantial technological differences which result in demand elasticity differences—will be difficult to apply in many situations. The person against whom sanctions must be applied, the patentee, will have less ready access to the facts determinative of which rule applies than will his bargaining opponents, the licensees. Finally, the prohibition of endproduct royalties would forbid, when it applied, the most familiar

and easily administered of all royalty systems, that based on a percentage of licensee sales. All these factors suggest that the adoption and effective enforcement of both rules would be attended by substantial administrative difficulty, both on the part of patentees and on the part of the enforcement authority. Certainly it would be unwise to adopt both rules as self-applying rules, violation of which could be punished after the fact through judicial proceedings.

It would be equally unwise to adopt neither rule and continue our present course of ignoring the implications of both economic phenomena. Patentees, within the limits of their ability to characterize the situations of their licensees, would pursue their self-interest and adopt in every case that royalty structure that causes greater economic harm.

Of the two phenomena, that involving economic discrimination is, in my opinion, the less important. Resource distortion and incremental restrictions on output through discriminatory royalties can occur only in a narrowly limited set of circumstances. There must be two groups of licensees who are in competition with one another and who have substantially different production technologies. One group must have a potential cost advantage over the other that is independent of the invention; yet the cost advantage must not be so great that it benefits the patentee to license only that one group and supply the entire market for the endproduct through that group. Even when all these conditions are met, it is doubtful that a substantially greater output of the endproduct would be achieved if an endproduct royalty structure was adopted. A tendency toward greater output will be created since the potential cost advantage of the one group will not be wholly absorbed. But the tendency will be countered by the more restrictive effect on output of endproduct-based royalty structure.

The significance of the input substitution phenomenon, on the other hand, is very broad. It applies in every case except those in which the patent applies to an ultimate consumer good and those very few, if any, cases in which there is no possibility of varying the licensee's production recipe. Surely if one of the two economic phenomena is to be disregarded by the law, the discrimination phenomenon should be disregarded. Unpatented endproduct royalty structures should be prohibited.

A legal pattern which takes both phenomena into account may be practicable, however, if the rule against discrimination is not made self-applying. The legal pattern that holds the greatest promise is, I think, the following. First, adopt a rule forbidding unpatented end-

product royalty structures. The imposition by a patentee of such a royalty charge would constitute patent misuse and an agreement in restraint of trade within the meaning of § 1 of the Sherman Act.

Second, provide an officially administered procedure by which an existing licensee can apply for an order directing the patentee to convert his license and that of all other licensees with whom he is in direct endproduct competition from an input-based to an endproduct royalty structure. To entitle himself to such an order the petitioner would have to prove: (1) that he is in, or proposes to enter into, direct endproduct competition with other present licensees; (2) that the endproduct market is quantitatively substantial; (3) that petitioner has a real cost advantage over existing licensees who hold a significant share of the endproduct market; (4) that the existing royalty structure imposes on him royalty charges that represent a certain percentage of his sales; (5) that the royalty charge currently imposed on the technologically different competing licensees, expressed as a percentage of their sales, is two or more percentage points less than his royalty charge so expressed; (6) and that, but for the differences in royalty percentage, he would probably be able to increase significantly the fraction of endproduct sales he is presently able to make.

If petitioner establishes those propositions, the official authority will then enter an order directing the patentee to tender to all licensees who are in competition in the endproduct market new licenses, substantially identical in their terms, in which royalties are expressed as a percentage of sales. The patentee will be free to impose any percentage not exceeding the presently effective percentage to the petitioner, and each prospective licensee will be free to take or reject the new license. On a date set in the order all old licenses become invalid and all who have not accepted new licenses are enjoined from practicing the invention and excused from making any royalty payments.

Under this legal system, input-based royalty structures would be the norm and endproduct royalties would be used only when sought by a competitor from whose activities resources were being diverted by a structure that nullified real cost advantages. If administered with reasonable intelligence and perception of the goals being sought, the system might work quite well.

A final observation is appropriate regarding the relationship between cases which do and cases which do not involve transfer of a product by the patentee to his licensees. As I have indicated, I would apply Robinson-Patman standards to product transfer cases. Cases involving product

transfers can be viewed as a subcategory of input-based royalty structure cases in which the product transferred is a mandatory royalty base. In cases not involving product transfer, the patentee is initially required to adopt an input royalty structure and, necessarily, the input or inputs he selects as a base will be measurable units of goods or services not supplied by the patentee. In this latter type of case a licensee could resort to the official proceedings just described and, upon proper proof, would be entitled to an order directing license renegotiation which would shift the royalty structure to an output-based structure. The question arises whether similar resort to official proceedings to obtain an endproduct royalty would be available to licensees in cases involving product transfer. The answer is no.

Proportionality of royalty charges to some particular input or inputs is no guarantee against patentee absorption of a potential cost advantage of one group of licensees. In the *Grand Caillou* case, for example, royalties were roughly proportionate to the number of raw shrimp processed, an input which measured with considerable accuracy the prior labor cost disadvantage of the Pacific canners and thus nullified any raw material or other cost advantage those canners may have had. The susceptibilty of input bases of royalty calculation to be thus used is the fact that argues for the establishment of the official procedure described.

But selection of an input which will measure elasticity of demand and thus maximize royalties and nullify a potentially significant cost advantage of one group of licensees requires considerable discretion in the selection among various possible inputs. In cases not involving product transfer, the patentee has a range of discretion in selecting an input: the only restraint that can practicably be placed upon him is that he select one or a small group of inputs closely related to practicing his invention. That restraint leaves sufficient discretion to create the risk insurance against which is afforded by the official machinery described.

But in product transfer cases the royalty base is required to be the product transferred. While in theory the patentee has discretion to decide what precise product he will produce and transfer, as a practical matter his range of discretion is very narrowly limited. He must select a product over which he can maintain a legal monopoly since he must transfer it at a monopoly price to earn his royalty. Therefore, the product must be one covered by the patent or one which "constitut[es] a material part of the invention . . . [is] especially adapted for use in

. . . [practicing the invention] and [is] not a staple article or commodity . . . suitable for substantial . . . [alternative] use. . . ."[74]

While it would be presumptuous to assert that a case can never arise in which a product transferred subject to Robinson-Patman standards will both meet the foregoing requirements and also serve as an accurate measure of licensee elasticity, such cases will be rare. Their rarity would not justify excluding them from the official machinery if they could be handled by that machinery with the same ease as cases not involving product transfer; but they could not be so handled. In a product transfer case licensee payments are attributable to two economic factors: the actual cost to the patentee of producing and delivering the transferred product and the patentee's monopoly return. To achieve sound results in shifting to an endproduct royalty structure, it would be necessary to separate these two components, to require the licensees to continue paying the cost component in proportion to physical units of product transferred, and to convert the latter component into endproduct royalties. Separation into the two components would require a cost analysis similar to public utility cost-of-service ratemaking. The social costs of administration are not warranted by the magnitude of the goal being pursued.

VII. Restrictions on Licensee Conduct

The basic thesis of this paper is that validity of patent license agreements is determined by the nature of the conduct the agreement is likely to induce. The patent law explicitly authorizes the extraction of monopoly profits by restricting utilization of and raising the price for using the invention. The antitrust laws explicitly forbid interference with the processes of competition to the end that prices will be reduced, ouput will be increased and resource allocation will be more nearly optimal. It is convenient to describe the patentee's freedom to monopolize as an "exception" to the antitrust laws' general mandate of competitive behavior; but I do not suggest that because the freedom is an "exception" to the general mandate it should be narrowly construed. It should be construed neither narrowly nor broadly but rather to achieve its obvious purpose—the subsidization of innovative activity. Yet, because the patentee's authority is an island of permission in a sea of prohibition, there is no area at the edge of permission toward which the law is indifferent: what is not authorized is forbidden. Every case which presents the issue of patentee freedom to monopolize neces-

74. See 35 U.S.C. § 271 (1964).

sarily presents a reciprocal issue of prohibited interference with competitive processes.

If the patentee has engaged in, or induced by agreement, conduct which restricts output of his invention, it would seem that he stands safely on his isle of permission. If he has engaged in or induced conduct that restricts output of any other good or service, it would seem that he has stepped into prohibited waters. But the metaphor is too simple, and fails. A restriction on use of the invention will in almost every case have the consequence of restricting output of other goods and services. If, for example, the invention has a physical embodiment, restriction on the invention will necessarily restrict the unpatented goods and services which are production inputs of the physical embodiment.

One is tempted, therefore, to fall back either to language of patentee "purpose" or to language of "direct" and "indirect" restraints. But in this context as in others those phrases tend more to obscure than to clarify. Articulated as tests, the concepts inevitably degenerate to inarticulate statements of legal conclusion. I suggest a formulation less crisp to the ear but better calculated to focus the mind on relevant differences: a patentee is entitled to extract monopoly income by restricting utilization of his invention, notwithstanding that utilization of other goods and services are consequently restricted, provided that in each case he confines the restriction to his invention as narrowly and specifically as the technology of his situation and the practicalities of administration permit. This formulation, I believe, gives appropriate scope to both antitrust and patent policy, is sufficiently specific to enable courts to reach reasonable results through reasoned applications, and, in the process of application, will be suggestive of even more specific subsidiary rules.

The standard must be read with emphasis on its first phrase: The patentee is entitled "*to extract monopoly income* by imposing restrictions." Generation of returns to patentees is the immediate objective of the patent laws and the only justification for tolerating the many baleful results of those laws. The only justification for using monopoly as opposed to direct government subsidy to induce innovation is to utilize the competitive processes of the private economy to assess and reward proportionately the value of each particular invention. This process of competitive assessment assumes that licensees will resist patentee demands and that out of the bargaining process between them will emerge a stream of benefits to the patentee and a stream of detriments to the licensee roughly comparable to the ultimate value of the invention to the licensee and to society. If the licensee conduct de-

manded by the patentee is only the payment of royalties, the hoped for comparability of monopoly income with invention utility will usually be achieved. But if other types of licensee conduct are demanded, no such comparability is necessarily to be expected. The value to the patentee of licensee conduct may far exceed its detriment to the licensee; indeed, the conduct may be as beneficial to the licensee as to the patentee, in which case the licensee has no incentive to resist the demands, and any expectation of comparability is foolish. In the absence of comparability, the social loss caused by restricting the use of the invention and otherwise permissible incidental restrictions on other goods and services may exceed, and often far exceed, the value of patentee contribution to knowledge.

There are then two critical points that underlie analysis of any particular situation: first, minimizing incidental restriction to the extent technologically and administratively feasible; and second, forbidding patentee demands for conduct of a nature that destroys assurance of comparability between restriction and invention utility. From these two points of vantage, I turn to specific problems of application.

A. *Royalty Structure: A Second Look*

1. *Input vs. Output Based Structures.* The relationship between the general standard and my previous analysis and conclusions regarding royalty structure is clear. A royalty based on licensee output is necessarily at least as restrictive and usually is more restrictive of licensee utilization of goods and services other than the invention as is a royalty based on invention-related inputs. Therefore, royalty structures based on unpatented endproducts, including the traditional percentage of licensee sales, should generally be forbidden.

The exception to that prohibition, if any is to be recognized, involves accomodation of the "discrimination" phenomenon through the officially administered procedure previously described.

2. *Royalties that Vary with Output.* Another royalty structure that should be forbidden is one which involves two or more licensees in competition with one another and which, when expressed as a percentage of licensee sales of an endproduct, increases as licensee output increases and therefore has the actual or potential consequence of limiting a licensee's share of the market. This prohibition should apply whether or not the endproduct is patented. Although such a structure calls explicitly for no conduct other than payment of money, it destroys any reasonable assurance of comparability. Where a licensee is in competition for endproduct sales with another licensee, this royalty structure establishes a division of the market between them

corresponding to the set of outputs at which their royalty costs per unit of output are roughly equal. Any new efficiency that one licensee may achieve that would otherwise enable him to reduce endproduct price and capture a larger market share is wholly or partially nullified by the rising royalty rate that attends increased output. Comparability is destroyed because licensees may affirmatively desire and be willing to pay for such a market division without regard for the actual value of the invention to them. They might be in a position to show the invalidity of the patent or to design around it and nevertheless be willing to take licenses and pay a modest royalty to achieve market division under a patent license umbrella. The licensees will restrict their output, at least in accordance with the royalties they actually pay, and this restriction will exceed that that would be dictated by royalties corresponding to the value of the invention. And if their industry structure is oligopolistic they are likely to restrict output and raise prices even more sharply than is dictated by actual royalty payments so as to exploit the monopoly position the market division affords them. When licenses take this form, the patentee is not selling the right to use the invention but rather a package which includes that right and a sanctuary from interlicensee competition.[75]

It would be appropriate to except from this prohibition royalty structures which impose a constant charge per unit on the licensee's utilization of an input related to the invention. Royalties under such a structure will not normally vary disproportionately with output, although a small variance would occur if the optimum production recipe varied with output. Rarely would the variance be substantial enough to produce the division of markets against which the prohibition is intended to guard. Exception of such structures from the prohibition would eliminate the possibility of a type of litigation the costs of which might exceed any benefits yielded.

One might argue this prohibition against royalties that increase more than proportionately to output should be extended to license arrangements not involving potential interlicensee competition but

75. This was one of the many outrageous abuses that led to a conclusion of antitrust violation in United States v. General Electric Co., 82 F. Supp. 753 (D.N.J. 1949) (lamp case). The licenses expressly provided that each licensee was limited in its sales in any one year to a certain fixed quota of General Electric's total net sales for that year. In the large lamp field, for example, Consolidated Electric was restricted to selling 3.89093% of GE's total sales, Sylvania, 8.124%, and Kenrad, 1.7584%; in the field of miniature lamps, Tungsol was authorized to sell up to 26.71956% of GE's sales, and Chicago Miniature, 2.975%. Royalties were 3⅓% of sales within the set quota. But if a licensee's sales exceeded its quota by more than 5%, an additional royalty of 20% of net sales became due.

involving potential competition between the patentee and his licensee. Quite clearly a patentee can use such a royalty structure to shelter his own share of a market from licensee competition. The argument for extension is unpersuasive, however. First, it is not in the immediate economic interest of the patentee to supply through his own production what a licensee can supply at a lower real cost. Secondly, the licensee receives no sheltered market share since the patentee-competitor has no royalty costs; so the arrangement lacks the kind of licensee advantage that destroys assurance of comparability. The first point has general significance and warrants brief elaboration.

The optimum objective for a patentee during the life of the patent is to cause the entire complex of firms engaged in using the invention to behave as if it were a single firm owned by himself. The maximum profit he can earn is that which such a firm could earn by producing as efficiently as possible and by selling at maximizing prices in all endproduct markets. The monopoly profits latent in those endproduct markets set the outside limit on the patentee's return. Assuming that actual consolidation under patentee ownership of all those firms is not practicable, the patentee's objective is to induce by license arrangement the same behavior that would attend consolidation—the same endproduct prices and outputs, the same production recipes and the same profits flowing to himself, in part perhaps through his direct sales of endproducts and in part through royalties. This is so whether or not the endproduct is patented.

As I argued in regard to input substitution, one way to accomplish this is to base royalties on unpatented endproduct sales, setting them at a rate such that they effectively raise the marginal cost curves of licensee sellers until those curves, aggregated, intersect the industry demand curve at a point directly above the intersection of the industry's real marginal cost curve with the marginal revenue curve, as illustrated in the footnote.[76] Forbidding use of unpatented endproduct royalty structures prevents the patentee from achieving this result to the extent his invention can be substituted against in the production process; and the patentee's maximizing objective becomes dependent on the demand curve for the invention which is, because of input substitution, more elastic than the unpatented endproduct curve from which it is derived. It nevertheless continues to be true that patentee maximization during the life of the patent depends on minimizing aggregate production costs exclusive of royalties; for anything that

76. See illustration at foot of next page.

raises the aggregate marginal cost curve for the industry diminishes the monopoly profit potential. Therefore, a patentee is better off to let a licensee supply any portion of the output, whether patented or not, that he can supply at a lower real cost than the patentee could supply it; royalty income on those sales will exceed the monopoly profits that could be earned by direct patentee sales.[77]

A patentee might substitute his own more expensive production for that of his licensee, thus sacrificing his own economic interest during the life of the patent, in the hope of establishing a long-run position of dominance in the market which will persist after the expiration of the patent. But it is always possible for a producer to expand his present market share by short-run sacrifices. In situations like the one under consideration, there is little reason to believe that a patentee could significantly influence his long-run position by this tactic. The single licensee will become the low cost firm when the patent expires and will take over a larger share of the market. If there is no other

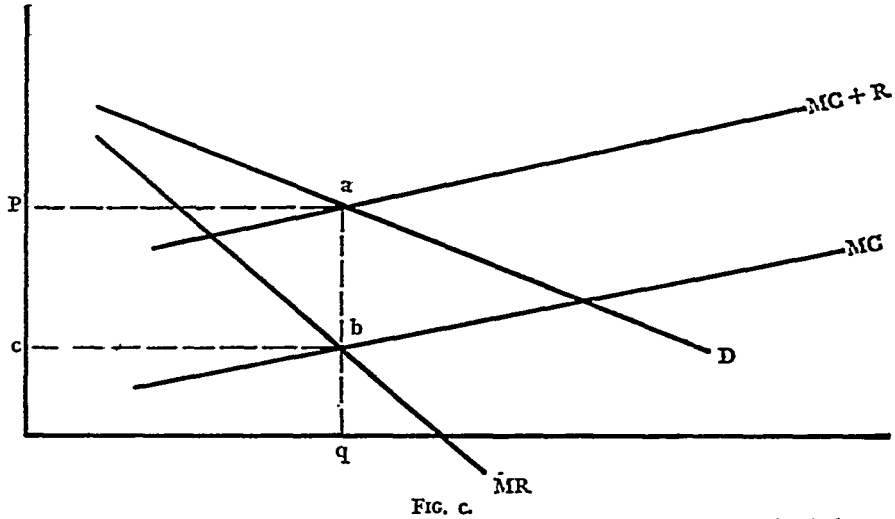

FIG. c.

D and *MR* are the industry demand and marginal revenue curves. *MC* is the industry marginal cost curve exclusive of royalties. *MC* + *R* is the effective marginal cost curve after the imposition of royalties. Monopoly profits (royalties) correspond to the rectangle pabc, and its area is independent of the division of output among individual producers.

77. The import of this analysis is not altered by the fact that the patentee, under the circumstance described, will be using a different and more efficient production recipe than will his licensees; he will not be substituting against his own invention, its use for him being costless. The patentee, to this extent, will be more efficient in real terms than his licensees; but he will have no incentive to substitute his own production for licensee production if, by reason of other cost savings, the licensee is nevertheless more efficient than he is.

The analysis in the text underscores the illogic of the FTC opinion in *Grand Caillou* insofar as it rejected the wholly credible explanation of why the patentee had established the "discriminatory" royalty structure and found as a fact that the motive was to protect the patentee's interest as a shrimp canner. See note 68 *supra*.

entry the industry will then exhibit oligopolistic tendencies; but that will be attributable to the number of firms and not to the earlier limit on the licensee's market share.

For the foregoing reasons, then, there is no sound argument for invalidating a royalty structure that increases more than proportionately to output when the licensee's only competitor is the patentee. The licensee receives no shelter against encroachment on his market share and must be presumed to be accepting the royalty obligation in exchange for a right of comparable value—the right to practice the invention. The patentee's reason for seeking such a structure may be based on administrative convenience of the structure, or on a misconception of his economic interest or on a non-economic consideration; for he cannot gain in economic terms by substituting his output for the more efficient output of his licensee. If the explanation is a misconception of economic interest or a non-economic factor, the royalty structure may do economic harm; but no justification occurs to me for the general subordination of unidentified non-economic objectives to economic goals or for using the antitrust laws to assure that private economic interests are correctly perceived.

B. *Tying Provisions: A Second Look*

I have previously described the way a tying clause can be used to block input substitution. Prohibition of its use when that is its effect is justified for the reasons given above.[78] It is anomalous, however, that the law should have arrived at its present rather extreme hostility to tying clauses and yet never have questioned endproduct royalty structures which achieve the same result far more effectively.

The anomaly is lessened but not eliminated by recognition that tying is subject to two additional objections. The elasticity of demand of a user for a monopolist's product can be measured in some situations with considerable accuracy by the extent of the user's consumption of some other product. By tying purchases of the secondary measuring product to his sales of the monopolized product, the monopolist can engage in economic discrimination between users of the tying product. In the most typical case the monopolist selects a tied product which users with inelastic demands for the tying product consume in greater quantities than do those with elastic demands. He then sells the tying product to all at a lower price than would otherwise maximize his returns and sells the tied product at a price exceeding its marginal cost and, presumably, its market price. In theory, at least, a reverse

78. Pp. 301-02 *supra*.

tie-in is equally effective. The monopolist sells the monopolized product at a price higher than would otherwise be appropriate and simultaneously commits himself to supply the user's requirements for a second product at a price below market value, the second product being one consumed in greater quantities by users with elastic demands for the invention. The effective price of the monopolized product is thus reduced to users with greater demand elasticity.

The significant characteristic for present purposes of discriminatory use of the tie-in is that the tied product must be supplied at a price different from its competitive price in order to perform its crucial surcharging or rebating function. In order for a tie-in to perform an input-substitution blocking function, a non-competitive price is also essential.

The existing prohibition against tie-ins extends, however, to situations where the tied product is supplied at a competitive price.[79] So applied the prohibition must be justified on the third possible objection to tie-ins—the "market foreclosure" effect. By tying to his legal patent monopoly the right to supply users' requirements of the tied product, the monopolist captures a segment of the market for the tied product and "forecloses" to competing sellers of the tied product that part of the market represented by sales of the tied product for use with the tying product. Case analysis generally stops with this descriptive statement as if it adequately portrayed some economic evil. But other than by the use of pejorative language in making the statement, no evil is portrayed. No price increase or output restriction is probable as a consequence of the arrangement. Nor is it likely that output has been shifted from more to less efficient producers; if the patentee were substantially less efficient than others at producing the tied product, the arrangement would seem to be contrary to his economic interest.

Indeed, with regard to the threat of the patentee's acquiring a monopoly position in the general market for the tied product,[80] it is irrelevant whether the price is above, below or equal to the competitive price. The price can only be above market if the patentee is charging less for his invention than the licensee would have been willing to pay absent the tie-in; the balance of the invention price is being paid in this indirect way. The price can be lower than market only if the patentee is ignoring his present economic interest or if he is indirectly rebating to compensate for overpricing the invention. Both these phenomena have implications for economic discrimination and input

79. See, *e.g.*, International Salt Co. v. United States, 332 U.S. 392 (1947).
80. By the general market, I refer to sales of the tied product generally rather than sales for use with the patentee's invention.

substitution, but neither is relevant to the extent or duration of foreclosure in the general market for the tied product.

Intelligent analysis of the foreclosure problem must start by focusing on the distinction between sales of the tied product for use with the tying product and sales for other uses. Let a represent the tying product, b the tied product or its sale for alternative uses and b/a sales of the tied product for use with a. The patentee has a legal monopoly on a—and the patent can be presumed valid since there is no collateral advantage to the licensee in a tie-in sale which destroys the assurance of comparability. Subject to the input substitution and discrimination phenomena, a patentee of a can extract all monopoly profits potential in endproducts which include a. I have shown that a price for b/a other than market price is essential if the patentee is to cope with those phenomena. It is senseless and misleading then to speak of foreclosure in the market b/a; the patentee already has achieved all that a tie-in at market price can give him in that market.

With regard to sales of b for other purposes during the life of the patent, the foreclosure concept is also vacuous. The patentee is guaranteed no sales of b by his control of b/a; he makes any such sales in open competition with other sellers of b. If there are scale economies in the production of b, he may be assisted in achieving the necessary volume by his control of b/a; but short of the point of monopolistic control over the total production and sale of b, achievement of such economies should be encouraged rather than thwarted.

The temporal focus in which the foreclosure concept becomes relevant is a long-run view which extends beyond the life of the a patent. During the life of the patent the patentee may use his monopoly profits from the invention to subsidize his expansion into the general market for the tied product, at least to the extent of b/a. This conduct is economically self-defeating in the short run, being comparable in all respects to selling b/a below market price. But if b/a is very substantial compared with b, the patentee may succeed in buying for himself a substantial share of the general market for b. Upon the expiration of the patent on a, the markets for b and b/a become legally indistinguishable. And at that time—after the patentee stops buying the share b/a with patent-generated profits—if there are barriers to entry he may continue to hold that market share and perhaps to exercise monopolistic power to raise price and restrict output in the general market for b.

While this concept of foreclosure cannot be dismissed as pure nonsense, I think it unlikely that the danger it anticipates is realistic in

very many situations. An investment in the form of underselling carried on over a fairly long time span is contemplated; and uncertainty whether the supposed entry barriers will prove secure after the investment period terminates introduces substantial risk. Hence, if the necessary investment is substantial, as it will be if the patentee is significantly less efficient than other *b* producers, it is unlikely to be a sound investment in any but rare cases; and both the motive underlying the arrangement and its primary economic effects are likely to lie elsewhere than in the foreclosure explanation. On the other hand, if the patentee is about as efficient at producing the tied product as all other producers, the investment will be small and the long-run game may be worth the candle.

The legitimate functions that can be served by tie-in sales, on the other hand, are also unimpressive in most situations. Patentee concern that the tied product is of suitably high quality to be used in conjunction with his invention so as to preserve the invention's reputation for utility can be met in most cases by generic specification. The administrative convenience yielded by the tie-in is generally insubstantial.

On balance the case against tie-in sales at market price is a close one. In theory they should probably be allowed only in the following cases: (1) where the general market for the tied product is much larger than the submarket represented by sales for use with the tying product and the tied product is sold for the prevailing market price; and (2) where, although the conditions of (1) are not met, the tying seller can show that the tie-in serves a legitimate objective which cannot be achieved by other means and is socially more important than any economic harm the tie-in might cause.

Whether the law ought to recognize these exceptions is a doubtful question; one might argue that ease of administration called for a flat prohibition in view of the very small number of situations in which any significant value would be lost because of a flat prohibition.[81] I would recognize the first exception. It is sufficiently precise to avoid administrative difficulty, and I cannot conceive how economic harm could be caused by a tie-in under the conditions the exception specifies. Input substitution cannot be blocked and discrimination cannot be achieved if the tied product is sold at market price, the foreclosure problem can be disregarded if the comparative market sizes are as specified, and the existence of an independent market for the tied product will facilitate ascertainment of market price. While the

81. See Turner, *The Validity of Tying Arrangements under the Antitrust Laws*, 72 HARV. L. REV. 50 (1958).

number of situations that qualify for the exception may be few, I see no reason to deny sellers whatever advantages they may find in the tie-in in those situations.

The second exception, if it can fairly be called an exception, is much more troublesome. Given its very general and subjective terms, it is little more than a reassertion of the rule of reason in this particular context and thus has the capacity to swallow the entire prohibition against tie-ins. I am unable, nevertheless, to give it greater precision nor can I assert with any confidence that no such exception should be recognized.

The case law seems to be in a similar state of indecision regarding the second proposed exception[82] and, with far less justification, apparently has rejected the first exception.[83] For purposes of this article an understanding of the nature of economic harm that can be anticipated by forbidding tie-in sales is sufficient. Like the Supreme Court[84] I will leave to the indefinite future the task of articulating generally the circumstances by which their use may be justified; for the tie-in problem I wish to discuss does not call for a general resolution of that issue.

The problem is the practice of package licensing. A patentee who holds a group of patents offers to issue a license to use the group of inventions for a singly stated royalty, the royalty being the same regardless of how many licenses are taken. Should this practice be viewed as a prohibited tie-in, the license desired by the licensee being viewed as the tying product and those he does not desire being viewed as the tied product?[85]

For reasons previously explained, I would require that all patent royalties be based either on sales of patented products or on quantitative consumption of unpatented inputs closely related to the licensed invention. This restriction itself generally would prohibit the package licensing of technologically unrelated patents; for only in most unusual circumstances would a common royalty base be appropriate for such patents. Although the reasons for this restriction were stated in

82. *Compare* United States v. Jerrold Electronics, 187 F. Supp. 545 (E.D. Pa. 1960), *aff'd*, 365 U.S. 567 (1961), *with* United States v. Loew's, Inc., 371 U.S. 38 (1962), *and* International Salt Co. v. United States, 332 U.S. 392 (1947). See Note, *The Use of Tie-Ins in New Industries*, 70 YALE L.J. 804 (1961).

83. International Salt Co. v. United States, 332 U.S. 392, 396-98 (1947).

84. United States v. Loew's Inc., 371 U.S. 38, 49-50 (1962).

85. See, *e.g.*, Automatic Radio Mfg. Co. v. Hazeltine Research, Inc., 339 U.S. 827 (1950); American Securit Co. v. Shatterproof Glass Corp., 268 F.2d 769 (3d Cir. 1959). In the *Hazeltine* case the Supreme Court raised doubts about the validity of the practice but did not rule on it, holding that the issue was not presented by the record. The practice was held to constitute misuse in the *Shatterproof* case, which relied on *Hazeltine* as well as on the tie-in doctrine generally.

another context, it will become apparent that the reasons for the restriction are closely related to the objections that can be lodged against package licensing. Therefore I discuss here only package licensing of technologically related patents.

In the absence of other objectionable features in the license,[86] there is no reason to prohibit package licensing of patents which are technologically related to such a degree that a common royalty base is permissible. The legal analogy which has cast doubt on package licensing is based on the judicial condemnation of block-booking in the motion picture antitrust cases.[87] No satisfactory analysis of the economic consequences of block-booking has ever been advanced by the courts; their standard incantation against "extending the monopoly" and against "foreclosure" has been relied on to vault to the conclusion of illegality.

The foreclosure analysis is not persuasive in the block-booking context. If the tied motion picture is viewed as unique, a market unto itself, then by virtue of the seller's copyright he already possesses a legal monopoly over that market. If the tied picture is viewed, more realistically, as one of many which the exhibitor might display to provide entertainment, then the total market for films is vastly larger than the submarket of those tied to the tying film; and the utility of the tie-in to achieve a long-run sheltered market position in film publication is minimal. Block-booking arrangements do not preclude the exhibitor from acquiring and exhibiting films from other sources. The tied film is not "used with" the tying film as are staples in a shoebutton stapling machine or salt in a salt dispensing machine. The exhibitor is free to show a film acquired elsewhere the night before and after he shows the tying film. Contrary to what may be one's first intuitive thought, the exhibitor is not foreclosed or deterred from acquiring films elsewhere even if he has paid more for the block of films than he would have paid (under some alternate marketing system) for the individual films in the package. If the least wanted film in the block will not attract an audience adequate to pay the marginal costs of exhibition, it is worthless to the exhibitor and will not be exhibited. A different film that will attract a sufficiently large audience to pay those costs, including

86. In *Hazeltine* the royalty was a percentage of sales of the unpatented endproduct. 339 U.S. at 829. In *Shatterproof*, too, the royalty base apparently was unpatented end product, although the scope of the patents involved is not made clear in the opinion. 268 F.2d at 771. Hence both royalty structures may have been objectionable on this ground.

If the tied license contained a stipulation of patent validity, prohibition of the tie-in would make more sense; but a superior solution would be direct invalidation of stipulation.

87. See, *e.g.*, United States v. Paramount Pictures, Inc., 334 U.S. 131 (1948); United States v. Loew's, Inc., 371 U.S. 38 (1962).

the cost of the new film, will be acquired. The assumed fact that the exhibitor paid more for the block than he would have paid under other arrangements for the films of value to him has relevance only to the size of the exhibitor's profits from exhibition of the desired films. Whether the assumed fact that he paid more is true is a point to which I will return.

Block-booking does not affect input-substitution if the royalty structure is not independently objectionable because based on endproduct sales—receipts from exhibition. The exhibitor is not deprived of economic incentive to shift his recipe so as to use more extensively comfortable seats or better theater location or prime broadcast time and to use film input less extensively in view of its monopolized price.

Finally, block-booking does not maximize the distributor's receipts by extracting from each exhibitor the largest sum he would be willing to pay for each individual film. If the costs of ascertaining each exhibitor's elasticity for each film is ignored, the producer could increase his receipts by abandoning block-booking and bargaining film by film with each exhibitor. Hence the device does not achieve, but rather foregoes, "perfect" economic discrimination in this classical sense.

Assume, however, that the realistic alternative to block-booking is not a process of individual bargaining with exhibitors over individual films, but rather that each individual film would be offered at a flat sum to exhibitors generally throughout the country.[88] As Professor Stigler has explained,[89] a distributor's receipts under that marketing system would probably be lower than under block-booking. They will be lower if one further assumption can be made: namely, that for any given film different exhibitors, attempting to appeal to different audiences in different parts of the country, will be willing to pay quite different prices; and that any given exhibitor will be willing to pay quite different prices for one type film than for another. Given these assumptions, both of which seem realistic, certain marketing consequences follow. In marketing a single film the distributor would be unable to obtain from exhibitors willing to pay a comparatively high price for

88. By a "flat sum" I mean a charge such as $500 per exhibition or per day or per any other base that did not vary among exhibitors inversely with elasticity. If the base selected were one that measured elasticity of demand with a high degree of accuracy, receipts should exceed receipts under block-booking. The fact that producers have resorted to block-booking may suggest either that no elasticity-measuring base is readily available or that they anticipate that discrimination charges would be brought against them if they selected such a base. My prior discussion of the inapplicability of classical discrimination theory to a product with negligible marginal cost would be applicable to such charges of discrimination.

89. Stigler, *United States v. Loew's Inc.: A Note on Block-Booking*, 1963 SUPREME COURT REVIEW 152.

that film a price as high as those exhibitors were willing to pay; for to set the uniform price that high would preclude use of the film by many exhibitors willing to pay only lesser sums. But if films of different types are combined into a package, each exhibitor in his own mind will assign different values to each film in the package; and the set of maximum values that exhibitors will be willing to pay for the package will be more homogeneous than the sets of maxima for individual films in the package. In an extreme example, all exhibitors would be willing to offer the same maximum, $10,000, for a block of two films; but half of them would be "paying" $4,000 for film *a* and $6,000 for film *b*, and the other half would be "paying" $6,000 for *a* and $4,000 for *b*. The distributor therefore is able to take advantage, to some extent of the individual instances of price inelasticity: while all exhibitors are paying the same amount for the package, each is paying a different amount, in accordance with his own demand inelasticity, for each individual film. Greater receipts can be garnered by selling the packages at uniform prices than the aggregate of receipts yielded by selling the individual films at uniform prices.[90] Block-booking, therefore, justifiably can be labeled a price discriminating device by one who thinks the label helpful.

For the reasons set forth in my discussion of price discrimination, I do not think that the potential of block-booking to achieve these results is a sufficient reason to prohibit block-booking. Satisfactory analysis must take into account the economic consequences of alternate marketing systems, in particular distribution costs under those systems, the restrictive effect of endproduct royalty structures and the essential arbitrariness of any royalty base other than marginal cost, a factor which renders the concept of discrimination unuseable where marginal cost approaches zero.

The case against package licensing of technologically related patents is even weaker than that against block-booking of films. No blockage of input substitution occurs if the royalty base is an input closely

90. Professor Stigler presents the mathematical basis for this proposition in his article. *Id.* at 153, n.3. The mathematics involved exceed the capacity of many lawyers, and I will not repeat them here.

A doubtful reader may satisfy himself by constructing a matrix of ten horizontal rows corresponding to Exhibitors 1-10 and nine vertical columns, eight columns corresponding to films *a-h* and the ninth corresponding to the package. For each exhibitor assign a value to each film at random from 1 to 10. The sum of each row will represent the maximum value that distributor would pay for the package of eight films; enter the ten sums in the ninth column. Each of the nine columns (8 films and 1 package) can now be represented as a demand schedule. The maximum receipts allowed by the demand schedule for the package will exceed the sum of the eight individual film maxima.

related to the invention. The foreclosure argument is wholly without foundation in this context: there can be no long-run market position in supplying the tied patent. And the revenue maximizing potential of package licensing, *compared with probable alternative licensing practices,* surely is negligible.

The revenue maximizing potential of block-booking inheres in the fact that the package is of value to a large number of exhibitors who have different relative preferences for the various components in the package; and it increases revenue only as compared with flat-sum charges for individual films imposed on a measuring base not correlated to elasticity. Although the two assumptions—relative preference differentials and uniform single film price as an alternative marketing system—may be realistic in the context of film distribution, neither seems realistic regarding patent licensing.

Where the potential licensees of technologically related patents are producing the same endproduct, as will often be the case, the assumption that they will have markedly different relative preferences for different patents in the group is not likely to be true. Their production processes will usually be similar and some one patent will be the most essential for all, some other the least essential for all. If relative preferences do not differ, package licensing achieves no discriminatory effect. Even if potential licensees are producers of different endproducts, the probability of their having significantly different relative preferences is, I think, less than in the case of motion picture exhibitors; for although different endproducts are involved, the intermediate production stages at which the inventions are used are likely to be similar. The relative values of technologically related patents do not depend on matters as idiosyncratic as tastes in entertainment.

Licenses under one or more of a group of technologically related patents are not likely to be of value to very large numbers of potential licensees. The smaller the number of potential licensees, the less likely is the alternative marketing system to be one of general offers to license each patent at a uniform price. The patentee may resort to individual patent-by-patent bargaining with each licensee or to a royalty structure which may sound uniform but is based on a factor correlated to elasticity. Either of these alternatives will have more discriminatory impact than package licensing in most cases. He "may" resort to these alternatives both in the sense that it may be economically practicable for him to do so and in the sense that he will be legally free to do so. The latter point follows, of course, from my earlier conclusion that the law can not practicably deal with the phenomena of discrimina-

tion in patent licensing other than by insisting generally on end-product royalty bases and that such insistence would sacrifice more important economic objectives.

Since the assumptions essential to the rationality of prohibiting block-booking, however valid they may be in that context, are very dubious in the context of package licensing, the prohibition should not be casually imported into the latter cases from the former. There is little reason to believe package licensing is an effective discriminatory device. When it is used, both the patentee's purpose and the economic effect of the device are more likely centered on cost savings of administration. Prohibition in such cases will force the patentee to more effective discriminatory techniques theretofore eschewed because of greater administrative costs. The patentee's net position may be worse as a consequence of prohibition but only because he is forced to absorb part of an increased real social cost that the prohibition has imposed upon the entire economic system.

One final situation must be distinguished. Assume that a patentee has a group of patents consisting of a valuable basic patent and several trivial improvement patents. The basic patent expires in five years, the improvement patents in ten. Package licenses are issued which run until the expiration of the last patent and call for payments of 2¢ for each unit of an appropriately related input consumed during the ten year period. In the minds of the licensees these payments are warranted by the utility of the basic patent; they would pay very little for the improvements if they could license the basic patent individually.

Can it be said that the patentee has extended the life of the basic patent in any meaningful sense?[91] It is immaterial that royalties continue to be paid after the patent expires: surely there is no economic objection to a license under which a royalty obligation is computed and accrues over five years but which allows the obligation to be paid off over ten years with interest on the accrued and unpaid balance. The effect of this delayed payout arrangement on licensee prices and output is no different than if they had obtained a bank loan to finance their royalty obligations.

In the hypothetical case described, however, the licensee had the option to forego use of the invention for five years and to use it royalty-

91. In *Shatterproof Glass* the court invoked this line of reasoning as an alternative ground for the holding of misuse. American Securit Co. v. Shatterproof Glass Corp., 268 F.2d 769, 777 (3d Cir. 1959)

free thereafter. His payments during the sixth through tenth years, like his payments in earlier years, correspond to the value to him of using the patent during the first five years. If the licensee is willing to pay 2ϕ per unit consumed over a period of ten years, he would have been willing to pay more than 2ϕ per unit consumed over a period of five years, how much more depending on prevailing interest rates among other factors.

It is critical, however, that in the package licensing hypothetical not only the payout period but the computation period is extended. Assume that computation for five years at 4ϕ with a ten year payout would produce a stream of receipts with the same present value as computation and payout for ten years at 2ϕ. Although the private economic interests of the parties may be identically affected by the two arrangements, the output and prices of the licensees will not be; and it is the latter to which the law should direct its attention. Royalties affect marginal cost, and hence output and prices, of licensees only during the computation period; thereafter any remaining obligation is a fixed cost. Output will be more severely restricted and prices higher during the first five years and output higher and prices lower during the last five years under the 4ϕ, five-year computation licenses than under the alternative.[92]

The basic question is whether there is an economic basis which justifies depriving the patentee and licensee of freedom to bargain for a license under which royalty computation, as opposed to royalty payout, extends beyond the life of the patent to which the royalties pertain. Since royalty computation after patent expiration will restrict output after expiration, perhaps it is enough to say that such a license does extend the monopolistic impact of the patent beyond its statutory term and is illegal.[93] That seems to me an inadequate analysis for it ignores the fact that royalty rates generally will be lower and the degree of restriction on output less during the life of the patent under the longer royalty computation provision. Royalties will be lower during patent life because allowing extended computation periods would not significantly increase the patentee's bargaining power vis-à-vis the licensee: the licensee has the alternative of foregoing the invention

92. Compare the majority and dissenting opinions in Brulotte v. Thys Co., 379 U.S. 29 (1964). In a careless opinion, Mr. Justice Douglas held invalid a license under which royalties continued to accrue after the expiration of the patents. He made no attempt to reply to the dissent's plausible but erroneous assertion that the economic consequences of the two types of arrangements were the same. The implications of the case for delayed pay-out arrangements are therefore unclear.

93. Brulotte v. Thys Co., 379 U.S. 29 (1964).

for the patent period and using the invention without charge thereafter. The patentee must give up something to induce submission to the more enduring restraint, for he has only a fixed period of exclusionary power to waive.

But another restraint limits the incentive of the patentee to charge higher royalty rates for the shorter period. He will diminish rather than increase his royalty income if the royalty rate is so high that it restricts licensee output short of the point where his marginal cost (apart from royalties) equals his marginal revenue. And it is in precisely those cases where this restraint would operate during a short computation period that patentees would have maximum incentive to bargain for extended computation periods. Prohibition of the extension is therefore justified, for in the generality of cases the prohibition will lessen the aggregate restraint; the magnitude of increased restriction on output during patent life will be less than the magnitude of restriction that would be imposed but for the prohibition during the subsequent period.

This conclusion does not warrant a flat prohibition on package licensing, however; for a far lesser inroad on bargaining freedom will achieve the objective. A package license should be held terminable at the licensee's option upon the expiration of any patent covered by the license. If the licensee does not choose to terminate upon expiration of a given patent, that fact affords adequate assurance that the royalty rate is comparable to the value of the viable patents licensed. This rule would not prevent the parties from agreeing to computation over the patent life with payments spread over a longer period; exercise of a licensee's power to terminate would leave unaffected any accrued royalty obligation.

C. *Restrictions on Price or Output*

The most notorious of patent license restraints is limitation of the price at which the licensee can sell his endproduct. The validity of such restraints has been litigated repeatedly over most of the life of the antitrust laws.[94] Several variations of this type of restriction must be considered: the endproduct for which the selling price is fixed may itself be patented or it may be an unpatented product to which the invention is an input of greater or lesser essentiality. Each of these two variations may involve a limit either on price or on quantity. Four

94. See, *e.g.*, Bement v. National Harrow Co., 186 U.S. 70 (1902); United States v. Huck Mfg. Co., 382 U.S. 197 (1966).

prototype restrictive clauses are thus identified. Each of the four may occur in a factual setting which does or does not involve two or more competing licensees. The possible variations will be examined in light of the general criterion I have invoked throughout this article: is the restriction of a type that impairs assurance of comparability between the social utility of the invention and the disutility of monopolistic restraint? Here as elsewhere assurance of comparability may be impaired either because the restraint is significantly less detrimental to the licensee than advantageous to the patentee, or because the restrictive impact of the license falls upon unpatented goods and services to a greater extent than is justified by technological and administrative considerations.[95]

A license which restricts the price at which the licensee can sell an unpatented endproduct to which the invention is merely one of various inputs should be prohibited even if only one licensee is involved. Such a clause generates a monopoly return to the patentee only by sheltering his own sales from licensee competition. This is plainly so if no royalty is being paid by the licensee: patentee sales are then the only source of patentee income. Less obviously but no less inevitably, this is the only effect of a price restriction even if the license calls for royalties too. A royalty obligation increases licensee costs and therefore the price at which he must sell if his total costs, including return to his capital and personal labor ("profits"), are to be covered. The amount of royalty obligation either is or is not of sufficient magnitude to force the licensee to sell at the fixed price. If it is of sufficient magnitude, the price clause serves no independent purpose; if it is not, the economic effect of the price clause on patentee income is to shelter his own sales at the fixed price.

But sales at the fixed price generate a monopoly profit that depends on the elasticity of demand for the endproduct; and as I have shown, the percentage change in the quantity of endproduct demanded will generally be less than the percentage changes in the quantity of patented input demanded in response to a change in input price. Hence monopoly profits will be greater and output of the endproduct lower than if the monopolistic restraint were confined to the invention itself. Use of other inputs in the production process is unduly restricted both because of excessive endproduct restraint and because, in the case of licensee production, input substitution occurs less extensively than if the same selling price were compelled by royalty obligations. Thus in

95. See pp. 312-14 *supra*.

any case in which the price restriction serves any function, the clause fails to meet the requirement that the patentee confine as narrowly as practicable the restrictive impact on unpatented goods and services.[96] Such price restrictions are also objectionable on grounds which will be discussed in conjunction with patented endproducts.

A limit on licensee output of an unpatented endproduct has all the adverse consequences of a price restriction. The restriction is inoperative if royalty obligations would force a price as high and an output as low as that imposed by the output quota. Whenever the quota is operative, the licensee will charge the highest price at which he can dispose of his quota; and the effects on patentee income and input substitution are those which would result from a price limit. In one important respect an output limit is even more harmful than a price restriction: the latter eliminates price competition; output quotas render futile all forms of non-price competition as well. The two types of clauses should be subject to the same legal prohibition.

The foregoing analysis of price or output limits on unpatented endproducts applies equally to cases involving only one licensee and to those involving two or more selling the endproduct in competition with one another. Additional adverse effects of such clauses in multiple licensee cases will be discussed hereafter.

Where the price or output restriction applies to sales of a patented endproduct, the restriction is not open to the foregoing objections. Although consumption of unpatented goods and services which are inputs to the endproduct will be restricted, that restriction is inherent in the lawful restriction on use of the invention and essential to the generation of monopoly profits correlative to the utility of the invention. Hence if one could assume that the same degree of restriction would occur in the generality of cases from increased licensee costs attributable to royalty obligations and from direct license restraints on price or output, there would be little reason to prohibit direct restraints. There are, however, persuasive reasons for rejecting that assumption and for accepting the contrary proposition that far more drastic restrictions are likely to follow from price or output restraints than from the cost pressure of royalty obligations.

Although I think patentee purpose should be dismissed as irrelevant for legal purposes because of difficulty of ascertainment, a consider-

96. Cases involving price restrictions on unpatented endproducts have held them invalid. See, *e.g.*, Standard Sanitary Mfg. Co. v. United States, 226 U.S. 20 (1912); United States v. General Electric Co., 82 F. Supp. 753 (D.N.J. 1949).

tion of patentee purpose is suggestive of economic consequences in the present context. The patentee would seem to have a strong incentive to achieve monopolistic restriction on output of his invention by means of royalties rather than by direct restraints on price or output. Assume that the maximizing price for patented widgets is $1.00 each, that industry sales will be ten million units at that price, that marginal and average total costs for all producers at that output are $.85, and that the patentee's direct sales constitute 25% of the market. Potential monopoly profit is $1,500,000,[97] and the patentee can garner all that profit if the industry, apart from the patent, is competitive and the royalty is 15% of gross sales. If, instead, the patentee sets price at $1.00 and imposes a 5% royalty he will receive only $750,000 in monopoly profits, half from his own sales and half from royalties.[98] The other three-quarter million in monopoly profit will be realized by the industry, but it will stay in the pockets of the licensees. In addition to giving up half the potential monopoly profits the patentee will incur significant costs in assuring compliance with the $1.00 price by his licensees, who will be in an economic position to shave prices and further eat into his profits by taking away his market share. This example can be generalized as follows: a price restriction is inoperative except to the extent it dictates a price in excess of that which would be dictated by royalties; and imposition of an effective price restriction forsakes monopoly profit to the extent sales are made by licensees. In the present example the effective price restriction is $.10 per unit,[99] licensee sales are 7.5 million units and $750,000 of profits are forsaken. In view of this inevitability one would suppose that there was no need for prohibiting price restrictions: patentees would not use them.

Two objections to the foregoing analysis are likely to be made, but neither will withstand examination. First, it may be objected that the patentee is entitled to protect himself against licensee price competition and that a royalty does not guarantee that objective.[100] But if the patentee is as efficient as his licensees, a royalty does protect him by giving him a cost advantage to the extent of the royalty. Only if his costs, which include no royalty, exceed the sum of licensee costs plus royalty will the licensees be able to compete with him effectively. Since the patentee can demand royalties of a magnitude correlative

97. ($1.00-.85) 10 million.
98. $.15 (¼) (10,000,000) from his own sales plus $.05 (¾) (10,000,000) from royalties.
99. $1.00 sale price less (.85 average cost + .05 royalty) = $.10.
100. See *Hearings on H.R. 4523 Before the House Committee on Patents*, 74th Cong., 1st Sess., at 1032 (1935) (testimony of Charles Neave).

to the utility of the invention,[101] royalties afford a cost advantage proportionate to patentee contribution. Unless the patentee is substantially less efficient than his licensees, royalties will afford a wholly adequate competitive shelter. And if the patentee is substantially less efficient than his licensees, it is contrary to his economic interest to substitute his inefficient production for that of his licensees.[102] Therefore an economically rational patentee can derive through royalties as much protection from licensee competition as is justified by the value of his contribution.

Second, it may be objected that a price restriction gives the patentee continuing flexibility of control whereas a royalty provision is fixed once the license is executed. It is true that the typical price clause, which requires the licensee to charge at any point in time the same price the patentee is then charging, gives more flexibility than the typical royalty clause which sets a percentage of some base for the entire life of the license. But clauses of those types need not be used: a rigid price clause could be used although the parties are unlikely to find it desirable; and a royalty clause could reserve to the patentee

101. A patentee will not necessarily be able to bargain for royalties fully commensurate with the economic value of his invention; for the market for the invention may be highly imperfect. Because substantially all the patentee's costs are sunk costs, he will license the invention for almost nothing if he has no other alternatives. The full economic value of the invention to a prospective licensee sets a ceiling on what the licensee will pay. The precise point within that range at which agreement will be reached will depend on the bargaining skills and on the alternatives of the parties.

A patentee with capacity to exploit the invention directly in the licensee's field has a substantially stronger bargaining position than does a patentee who lacks that capacity. Similarly, if the potential licensee industry is competitively structured, the patentee will have alternatives; and he will be able to negotiate a royalty approximating the value of his invention.

On the other hand, a patentee who lacks capacity for direct exploitation and whose invention is applicable in an industry which is monopolistically or oligopolistically structured will have very little bargaining strength. The situation of such a patentee is not materially improved by his ability to use a price restriction. By hypothesis he has no substantial direct sales to shelter with the restriction. A potential licensee who already has a monopoly will pay nothing for the service of being subjected to a price restriction. Potential licensees in an oligopolistic industry often would be benefited by a price restriction and this fact will increase the value to them of the *license* as opposed to the invention. This increase in value raises the ceiling but not the floor of the bargaining range. But patentee continues to lack any significant alternative, and he is unlikely to receive a substantially higher royalty.

Any incremental royalty such a patentee does receive is a payment for the service of eliminating inter-licensee competition rather than for use of the invention.

Attempts to justify the use of price restrictions by reference to the plight of such a patentee should be rejected. His unfortunate situation is similar to that of all other suppliers and customers who must deal with an industry so structured. The appropriate legal response is an assault upon the industry structure. It is wholly inappropriate to allow such a patentee to improve his position marginally by selling, not his invention but a restrictive clause that will further suppress competition in the industry and thus worsen the position of all others who must deal with the industry. It is even more inappropriate to make price restrictions available to patentees generally because of the asserted marginal value of the clauses to patentees in the situation described.

102. See pp. 316-17 *supra*.

power to alter the royalty rate, perhaps within limits, from time to time. It is no doubt true that licensees would resist variable royalty clauses more strenuously than they resist variable price clauses; but the difference in resistance to be expected is closely related to the vice of price clauses—namely, that licensees are beneficiaries of such clauses and have little incentive to resist them. There is no reason to believe that a patentee could not bargain successfully for a royalty rate subject to his unilateral power of alteration from time to time up to some fixed maximum rate; and for the reasons stated above, such a clause would serve the patentee's economic interests much more effectively than would a variable price clause. Hence the objection based on flexibility, at least in the naïve form stated in this paragraph, assumes the point it attempts to prove by starting with a rigid royalty clause and a flexible price clause.

A more sophisticated form of the inflexibility argument can be made. It starts with the roughly accurate observation that a patentee's ability to dictate price gives him a flexible tool with which to assure that the industry operates at the price and output that maximizes industry profits.[103] Next it is correctly asserted that even a flexible royalty clause would not assure to the patentee ability to bring about this maximizing of industry price and output, particulary in an oligopolistic industry. In such an industry there may be no such close relationship between firm costs and prices as under either competitive or monopolistic conditions. Therefore a small increase in the flexible royalty rate might not produce the desired price increase; and a small reduction in royalties would be even more likely to fail to bring about the desired price decrease.

All this may be conceded; the deficiency of the argument lies in its irrelevance. What interest has the patentee in whether the industry is operating at optimum price and output? None, except insofar as he is either a direct seller or a royalty recipient or both. In his capacity as a direct seller, he benefits from control over industry price only in accordance with his percentage of total sales; he shares his potential monopoly profit with licensees to the extent they make direct sales. He would be better off with the less precise control device of a flexible royalty rate unless he has a very large fraction of the market.

With respect to the patentee's capacity as a collector of royalties, gen-

103. Even this statement is not wholly accurate since it ignores the possibility that firms may be operating at outputs at which their individual marginal costs may be unequal and therefore the industry is not operating at minimal cost. See LEFTWICH, THE PRICE SYSTEM AND RESOURCE ALLOCATION 216 (3d ed. 1966).

eralization is more difficult. If, as is usually the case, his royalty is a percentage of sales, he has no interest in optimum industry price. Royalty income would be maximized by a lower price where demand elasticity was unitary. Similarly, a lower price would benefit him more than the maximizing price if his royalty was based on a production input. Only if the royalty base was net profit, a base I believe is seldom used, would royalties be increased by industry price maximization.

This second argument regarding the flexibility of price restrictions, then, is plainly invalid unless the patentee's direct sales constitute a large share of the market; and even in that situation he is failing to capture all his profit. Such a patentee should insist on both a price restriction and a flexible royalty rate. The former would allow precise control over industry price and output, and the latter would enable him to extract all the resulting benefits. Arguably a price clause should be permitted if the patentee enjoys a large share of the market and also has power to vary royalties and uses that power to garner all monopoly profit.

But even that argument should be rejected. Such a patentee can achieve his legitimate objectives without the price clause. By setting his own price at the industry optimum he can prevent any licensee from charging significantly higher prices. And by raising royalty rates he can prevent any licensee from charging less. Such a patentee does not need a price restriction; and since restrictions pose grave threats of abuse in other contexts they should be flatly prohibited.

In view of the fact that royalty obligations serve the interest of patentees much better than do price clauses, how does one account for the fact that price clauses are so extensively used? The answer is that price clauses are used precisely because their favorable economic impact inures to licensees as well as patentees, because their adverse economic impact falls not on the parties but upon the customers of the parties, and because the monopoly income they generate may exceed substantially the value of the licensed invention.[104]

In the one situation of the patentee who makes a large percentage of direct sales and also has the power to vary royalties, there would be little difference in industry performance whether or not there was a price clause. A form of low-cost-firm-price-leadership would result, the patentee being the low-cost firm by reason of his control over royalties.

104. The argument in this and immediately following paragraphs is made by Helmut F. Furth in one of the few useful pieces of scholarship dealing with patent licensing problems. Furth, *Price Restrictive Licenses Under the Sherman Act*, 71 HARV. L. REV. 815 (1958).

A patentee could bargain successfully for such a powerful position only if his patent was both economically valuable and legally invulnerable.

It is in situations of patents which are not both valuable and invulnerable that price restrictions significantly affect industry performance. Control over price is used to achieve monopolistic pricing in the industry, but a large part of the profits remain in the pockets of licensees. The patentee does not extract it because he does not have the bargaining power to extract it. The price restriction is the backbone of a loose-knit cartel.

More specifically, a price restriction clause that is operative always has the primary consequence of dividing with the licensees, in proportion to their share of market sales, the monopoly profit potential of the endproduct. Because it does so it has several secondary consequences. First, licensees have no incentive to challenge validity of the patent. If the patent were held invalid, competition among the licensees, the patentee and perhaps new entrants would deprive licensees of their present share of monopoly profits. It is true that some outsider might challenge, or threaten to challenge the patent, but this fact does little to compensate for the elimination of licensee incentive to challenge: the licensees constitute the group most likely to possess knowledge revealing invalidity; and, moreover, an outside challenger can often be bought off by issuing him a license containing price restrictions and thus bringing him aboard the gravy train.

Second, the licensees have an incentive to suppress any unpatented product that competes with the patented product. An agreement to suppress such products is, of course, a prohibited conspiracy. But conspiracies are difficult to prove even if something sufficiently formal to be called an agreement does exist. And in any event the parties have no need of agreement: each will realize that the interests of all call for suppression. Again, introduction of such a competing product by an outsider is possible, but licensees constitute the group most likely to possess knowledge of the existence and market potential of such a product. Again, the outsider may be bought off by letting him into the club, and the outsider's incentive to break up the game is minimized by the facts that he would possess no shelter from competition and would suffer the hostility of the club members in the ensuing competition.

Third, licensee incentive to invent around the licensed patent is far less than if all monopoly profits were being absorbed by royalty obligations. Investment in experimentation that would be warranted by the prospect of avoiding royalty obligations and acquiring a monop-

oly position for himself is less attractive to a licensee who already has a shared monopoly and is paying low or nominal royalties.

Because of these effects on licensees, the price restrictive license destroys all assurance of comparability between the degree of monopoly restriction being imposed on customers and the value of the patentee's contribution. The suppression of potentially competitive products may afford a substantial degree of monopoly power to a patent which is invalid or which constitutes little or no contribution to knowledge. If the supressed products are old products on which the patented product is an "improvement," the conspiratorial overtones of the arrangement may be quite obvious.[105] Licenses will probably have to be issued to all existing producers of the old product before any licensee will commit himself to even a nominal royalty; and the awareness of customers and other outsiders of the existence of the "unimproved" product will generate resistance to substantial price increases for the new product, may lead to cheating by licensees, and may require elaborate policing steps by the patentee. These circumstances may generate sufficient evidence of "conspiracy" to support a § 1 charge. But where the patented product is new and the alternatives are suppressed by the failure to introduce them rather than by their withdrawal, proof of conspiracy will be difficult or impossible. Since the anticompetitive conduct is inherent in the price restriction, use of such clauses should be flatly prohibited. Not only do they threaten monopolistic exploitation far beyond that commensurate with the value of the patentee's contribution, but they also undermine the objectives of patent policy by weakening incentives of licensees to invent around the licensed patent since it is a source of benefits rather than of royalty obligations.

The position regarding price restrictions suggested by the Third Circuit, that such a clause is permissible if it appears in only one license, is unsound.[106] All the harmful effects described above can occur

105. *Cf.* United States v. United States Gypsum Co., 333 U.S. 364 (1948).

106. Newburgh Moire Co. v. Superior Moire Co., 237 F.2d 283 (3d Cir. 1956) (dictum). The court was led to this position not by any economic consideration but by attempting to parse language in earlier and inconsistent Supreme Court opinions. An intelligent comment on the case can be found in Note, *Patent Price Restrictions and the Antitrust Laws: A Balance Upset*, 67 YALE L.J. 700 (1958).

The conclusion reached in the Note is that the "narrow revision" of the law by judicial action is unsatisfactory and that "broad revision, requiring delicate adjustments" must be left to Congress. Narrow revision, including flat prohibition of price clauses, is found unsatisfactory because: (1) it would not prohibit other anticompetitive license practices that have similar effects, and (2) because it would affect some patentees more drastically than others.

The first objection to flat prohibition by judicial action ignores the ability of the

where there is only one licensee if he and the patentee hold a substantial portion of the relevant product market. A far more sensible distinction would permit price restrictions if the patentee and his licensees, whatever their number, enjoyed less than half the relevant market; for the existence of substantial competition from companies not being force-fed monopoly profits would minimize the potential damage of the price restriction. But in the absence of some demonstration that the clause yields patentees some advantage not obtainable by devices less destructive of competition, I see no reason ever to uphold such a clause.

In discussing royalty structures that increased more than proportionately with increases in licensee output, I suggested that such clauses were tolerable if they occurred in a single license. Such clauses are different from price restrictions because, unless there are multiple licensees, they benefit only the patentee. A price restriction would be more analogous to such clauses if the patentee had power to sell for any price he wished but the licensee could not sell below a set price; but the analogy would still be defective. Under the increasing royalty percentage the licensee's output is restricted only by the pinch of increasing royalty obligations and those obligations represent gains to the patentee. The licensee's willingness to incur those royalty obliga-

courts to control the use of other restrictive clauses and thus reestablish the logical symmetry of which the author is enamored.

The second objection, based on equity as between various types of patentees, is unpersuasive for more complex reasons. Patentees of consumer products would be more drastically affected than patentees of intermediate products, the Note argues, because demand for the former is more elastic than for the latter. The existence of this asserted difference in demand characteristics is neither documented empirically nor supported by persuasive theoretical argument. Demand elasticity for intermediate products attributable to input substitutability is ignored; and elasticity attributable to elasticity of demand for the end product to which the patented product is intermediate is acknowledged in the last sentence of a long footnote but is ignored in the text. 67 YALE L.J. at 709-10 n.31.

Patentees capable of producing the entire optimum volume of the patented product would be less drastically affected by flat prohibition than those not capable of doing so, the Note argues. The basis for this assertion is inadequate for two reasons. First, the Note relies upon advantages which would be held by patentees capable of large scale production whether or not price clauses are prohibited. See text at n.26. Second, the ability of the patentee to control industry output by using a price clause is assumed to be equivalent to is ability to maximize his own revenue. It probably is true that a patentee with power to vary selling price unilaterally can control industry price and output more delicately than one who can vary royalties; but except to the extent of his profits on direct sales, controlling industry output and price does him no good. See note 103 *supra*. He maximizes royalty income only to the extent he varies royalties. This fact is acknowledged cryptically in the last sentence of another long footnote, see 67 YALE L.J. at 707-08 n.25, but it is ignored in the text. The argument in the text is gravely impaired by the footnote acknowledgement.

Finally, I do not think that minor differences in impact upon different types of patentees—if any differences in impact exist—justify preservation of a licensing device that has the primary function of destroying comparability between the patentee's contribution and the degree to which monopoly restraint is imposed on the endproduct market.

tions gives some assurance of comparability, for he has received no assurance of monopoly profit from the face of the license. A price restriction not tied to the patentee's own prices also suggests that the patentee had a patent of some inherent value; for on its face it affords the licensee only the small comfort of high prices on whatever fraction of industry sales the patentee chooses to allow him. But a critical difference remains: under the price clause, to the extent the fixed licensee price exceeds that to which a greater royalty obligation could have forced it, the patentee is sharing the monopoly profits; and this fact suggests restraint unwarranted by the value of the invention.

A license restriction on licensee output—restriction to a set percentage of industry sales, or to a specific quantity quota or to a quota set from time to time by the patentee—should be treated just as a price restriction for the consequences are identical. Licensees share the benefit of restriction on industry output through resultant high prices; they have no incentive to market competitive products or to challenge validity of or to invent around the patent; therefore assurance of comparability is destroyed.

D. *Restrictions on the Field of Utilization*

The term "field restrictions" causes difficult problems of analysis primarily because it is used to refer to license clauses that have very different functions and that should be recognized as raising very different problems. For several reasons a patentee whose invention is susceptible to application in two or more independent areas of economic activity may wish to impose different terms on applications in the different areas. He may wish to charge different royalty rates or base royalty rates on different royalty bases in the different fields; or he may wish to vary as between fields the license terms in other respects—for example, to bargain for a grant-back clause from licensees in field #1 but to forego grant-backs and concentrate on bargaining for the highest possible royalty in field #2.

This general function of the field restriction is entirely appropriate. Use of differential royalties raises the pervasive problem of economic discrimination and requires further discussion; but assuming that in each type of license the royalty structure and any restrictions imposed are lawful, it is my conclusion that field restrictions used to perform this function should not be prohibited.

Assume a patentee whose invention is useful in fields #1 through #5 issues identical licenses to one hundred licensees, twenty of whom are in each of the five fields at the time of agreement. Each license

expressly contemplates that the licensee may enter any one or all of the other four fields and provides in substance the following arrangement: Royalties in field #1 shall be 5¢ per unit of input *a* consumed; in field #2 they shall be 10% of your expenditures for input *b*, and you shall issue to me a royalty free, nonexclusive license under any improvement patent you may acquire that is useful in field #2; royalties in field #3 shall be 7¢ per unit of input *c* consumed, etc.

No objection can be taken to this license pattern except on the ground that the patentee is maximizing his royalty income by setting rates on the basis of elasticity of demand for use of the invention in the different areas. For reasons previously given that objection is unpersuasive. To require a uniform royalty structure in different areas of economic activity would substantially diminish rewards to patentees to no useful purpose and often with harmful consequences. Uniform royalties would require not only the same rate but the same base, and a common base would not be available in the different areas in many cases unless an endproduct sales base were permitted. Such a base is unacceptable because of its effect on input substitution.

Even if a common base were available in a given case in the sense that there was a common input in the several production processes, it would not necessarily be an appropriate base in each area—appropriate in the sense of constituting an input closely related to the invention. And if the common input was an appropriate base in all areas, it would nevertheless be likely that producers in one area would use the base more intensively than those in other areas so that discrimination would not really be eliminated but merely disguised. Phonic but not economic equality would have been achieved. The very factor that makes an input base appropriate is its capacity to measure the utility which the licensee is deriving from the invention.

All the foregoing difficulties inhere in the fact that the classical concept of economic discrimination furnishes no adequate basis for normative rules when applied to a product which has no marginal cost.

Finally, if one managed, in some way I cannot conceive, to eliminate charges based on licensee utility, one would not have accomplished anything worthwhile. Forced to abandon charges based on utility in each field of application, a patentee would maximize against an aggregate demand for use of his invention and select a point on that demand curve approximating unitary elasticity. All potential applications of the invention having lesser utilities would be abandoned: they could not afford the uniform price which would exceed their old price. All applications with a higher utility would use the invention, many with

greater intensity than before. One cannot generalize that the resulting shift of resources from low utility uses to high utility uses would reflect an advance for the economy. One cannot even predict that more extensive use would be made of the invention; for although prohibiting the separation of markets diminishes income to the seller, it does not necessarily increase the quantity of the product that will be purchased.[107]

Field restrictions may be misused, however, to achieve an allocation of markets among actual or potential competitors. Like price restriction clauses, the greatest potential for causing economic harm exists when field restrictions are used in conjunction with patents of little commercial value. The economic loss caused by a price restriction is indicated by the difference between the fixed price and the price that would have been compelled by royalty obligations; the loss caused by a field restriction is the difference between the endproduct price that occurs as a result of the field restriction and the highest price that might have been compelled by royalty obligations. Since in both cases a practical limit on endproduct price is set by the relationship between demand for the endproduct and costs other than royalties, the potential of the restrictions for harm varies inversely with the value of the patentee's contribution and his resultant ability to bargain for royalties. Any monopoly income that prevailing endproduct price yields that does not flow to the patentee either through royalties or his own direct sales remains in the pocket of the licensee. Such income does not serve the purpose of the patent system and hence corresponds to unjustified restriction on output.

The foregoing conclusion can be illustrated by examining the cases, first, of a very valuable patent and, subsequently, of a trivial patent. Assume a patentee makes a startling breakthrough in laser technology which makes possible inexpensive and reliable motion picture photography and the reproduction of sharp images in three dimensions, a significant advance on present holography techniques; and assume, further, that viewers of all types find three dimensional reproductions so much more satisfactory than plane images that the technique has the potential to supplant totally present motion picture technology. The device would obviously have potential application in motion picture entertainment, industrial training films, military training films, advertising films, educational films and perhaps home movies. The elasticity of demand for use of the technique would differ in each of these fields, and the patentee might license its use at maximizing roy-

107. See Appendix, *infra*, at note 5.

alty R_1 in field #1, R_2 in field #2, etc., issuing licenses at those royalties to all applicants in each of the several fields. In each field there would be a number of licensed competitors. Each licensee would pay royalties according to the rate for his field and his individual rate of utilizing the royalty base for his field. From the several licensees in field #1 the patentee would receive a total sum of royalties, TR_1, from those in field #2, TR_2, etc., and his aggregate royalty income would be the sum of these totals. This pattern of licenses is wholly appropriate.

Assume, alternatively, that the patentee licensed only one user in each field at the highest royalty for which he can bargain, and the single licensee then, because of the commercial significance of the patent, displaced all unlicensed competition in his field. Several inquiries concerning this alternative are appropriate: (1) Is output in any given field likely to be more severely restricted under this alternative than under the first license pattern described, with the consequence that this license pattern may be labeled "undesirable"? (2) Is the total royalty income obtainable from any given field likely to be greater under this than the first described license pattern with the consequence that patentees will have an incentive to adopt this license pattern if allowed to do so?

During the life of the patent, output of the field is likely to be lower, but probably not significantly lower, under the one-per-field license pattern than under a multiple-license pattern, and this is true whether the patent covers the endproduct itself or only a vital input.[108] If the patent covers only a vital input, then to some extent other inputs will be substituted against it; but this will occur whether there are one or many licensees. If the single licensee selected in the one-per-field pattern is not as efficient as other potential licensees, output in the field will be further restricted by the licensee's higher costs; but since selection of an inefficient licensee is contrary to the patentee's interest, reasonable assurance against this possibility is afforded; and this possibility alone would not seem to warrant legal prohibition. To the extent that the patentee misjudges field demand or licensee costs and fails to bargain successfully for a royalty which extracts full monopoly potential, output is likely to be more restricted under the one-per-field pattern; for the single licensee presumably will capture the remaining monopoly

108. In my discussion at this point I assume that the patentee is confined to an appropriate royalty base, but the assumption is not essential to the analysis. If the patentee is free to use an endproduct royalty structure although the endproduct itself is not subject to the patent, then the degree of restriction under both patterns will be greater; but the comparative restriction attributable to the license pattern will not be affected.

potential himself, whereas competition between multiple licensees may cause the differential to be passed on. Multiple licensees may also be preferable because they will offer greater product differentiation and will have greater incentive to achieve further technological advance.

Consideration of the performance of the field after expiration of the patent affords a further basis for preferring multiple licensees. Competitive conditions are likely to be restored more quickly if the field has not been confined for the patent's life to a single licensee who, by the expiration of the patent, may have achieved a position which heightens entry barriers.

Finally, the strong ideological preference of our society for multiple participation as opposed to monopoly should not be disregarded if economic considerations are roughly in balance.

The second inquiry is whether one-per-field license patterns yield incremental returns to patentees. If the answer to this question were negative, then prohibition of such license patterns could easily be justified since the objectives of the patent system would not be thwarted in any way. However it is probable that such patterns yield incremental returns to the patentee even in the case of a valuable patent. If the patentee has correctly estimated and successfully bargained for the full monopoly potential of the patent in the field, the sole licensee is little benefited by his exclusive position during the life of the patent. While he enjoys his exclusive position, he has the quiet life of the monopolist but none of its economic fruits. It seems probable that the licensee will pay a premium royalty rate if, but only if, he has some assurance of the continuity of his exclusive position and in addition, one of the following factors is present: he values the prospect of establishing a dominant position in the field during patent life because of the possibility of exploiting that position subsequent to patent expiration; or he estimates demand and cost functions in the field differently than has the patentee, expects to be able to garner monopoly profits beyond those which would be absorbed by the royalty demanded for a nonexclusive position, and offers the premium royalty for the exclusive position as a mode of sharing with the patentee the additional monopoly profits the licensee expects.

To the extent that payment of the premium is induced by the licensee's long-run expectation of a dominant position sheltered by entry barriers of some sort, the premium does not constitute income of the kind contemplated by the patent system. Such income does not correspond to the value of the patentee's contribution measured during the permissible time period but is comparable to an extension of patent

life. To the extent the premium is attributable to what has been described as different estimates of cost and demand functions in the field, it is income that corresponds to patent value: the position of exclusivity is a factor strengthening the bargaining power of the patentee in his attempt to extract the full monopoly power of the patent as estimated in advance of agreement.

If one considered only cases involving very valuable patents, the arguments for and against toleration of one-per-field licensing patterns would thus be fairly closely balanced. While such patterns probably generate incremental income for patentees, the magnitude of the incremental income is likely to be small; and only some indeterminate portion of it will be attributable to invention value. In any case where there is incremental income there will be incremental restraint on output; and some indeterminate part of the restraint will be unjustifiable in terms of patent policy because it is attributable to monopoly profit shared with the licensee and licensee anticipation of post-patent market position.

Cases involving comparatively trivial patents must also be considered, however; and in such cases resultant restraint on output may substantially exceed that justifiable by invention value. An industry characterized by loose and fairly competitive oligopolistic structure can be restructured into a series of monopolistic satraps by parceling out exclusive fields within the industry to the existing participants. Here too the one-per-field license pattern, if permitted, will generate patentee income; for in addition to whatever value the invention may have, licensees will pay something for the service of carving up the market into monopolized fields. But patentee interest in such income plainly is not protected by patent policy, and the resultant restraint transgresses antitrust objectives.

To this analysis it may be objected that if the patent is not of great commercial significance, there is no real barrier to entry by outsiders into any one or all of the fields. While technically valid, the objection is not persuasive: it proves too much. On similar reasoning there is no reason to forbid price conspiracies or monopolization or horizontal mergers; for in theory the existence of monopoly profits will signal entrants who will erode the established market positions. But in practice the unjustified market positions endure long enough to make the game worth the candle for conspirators and monopolists. With one-per-field license patterns as with price restrictions, those in the best position to assess correctly the true value of the invention and the

practicality of operating without using it or of having the patent declared invalid are, by the licensing device, converted from potential attackers to committed defenders. The fact that some were willing to take licenses, pay royalties of some amount, and abandon alternative production routes to the endproduct creates a façade of legal invulnerability and economic importance for the patent which may be enough to deter more thorough examination by individuals whose primary efforts are directed elsewhere. Unless a device which has such adverse economic potential is also essential to achievement of desirable goals, the argument against blocking its use must fail.

In my opinion the limited utility of one-per-field licensing patterns to patentees as a bargaining device to extract the last drop of potential royalty revenue from valuable patents does not justify acceptance of misuse of the device; nor do I believe that proper and improper use can be identified by factual inquiry on a case-by-case basis.

In formulating rules governing this problem, however, care must be taken lest more be thrown out than bathwater. Any prohibition so sweeping in its scope as to block differential royalty rates applicable to different fields would be unwise. Nor is it realistic to require a patentee, in negotiating with a potential licensee presently interested in field #1, to anticipate all other possible fields of use and prescribe in the license what royalties would be if the licensee chose to expand his operations into other fields. The parties should be free under those circumstances to execute a license that applies only to field #1. Negotiations between them on a mutually acceptable royalty applicable to field #2 can be left for a future date when the licensee's interest in that field becomes less theoretical. Whether the license is merely silent as to other fields, leaving application by the licensee in other fields subject to infringement remedies, or contains promises that the licensee will not practice the invention in other fields, should be immaterial; neither a misuse defense nor an antitrust claim should be available to the licensee or to any third person on this state of facts alone.

The appropriate weapon with which to strike at the restricted field license is to require the patentee to issue to any qualified applicant a license to practice the invention in any field previously licensed on terms equal to those in existing licenses. A licensee under a currently effective license in any field, by reason of that fact alone, should be regarded as a qualified applicant with respect to any other field and as entitled, upon application, to be licensed to practice in such other field

if that other field has been licensed to any third person. Only a breach of his original license of sufficient importance to warrant its cancellation which the licensor has invoked as a basis for cancellation should exclude from eligibility a person to whom the patentee has voluntarily issued a license. An applicant who is not licensed under the patent in any field should be presumed qualified, subject to the patentee's ability to prove the existence of facts justifying a refusal to deal with him. Such facts would include prior acts of commercial dishonesty by the applicant or lack of financial solvency or other facts which raise substantial doubt about the applicant's willingness or ability to comply fully with the license terms. Notwithstanding proof of such facts, issuance of the license should be required if, in light of the facts shown, the court can issue an order which fully protects the interests of the patentee in faithful performance. If, for example, the refusal is based on undercapitalization of the applicant, the court should order the license issued on the condition that the applicant posts a bond or establishes a trust fund to insure royalty payments. Reasons for refusal other than financial would be more difficult to deal with by protective order; and if the facts asserted are found true and the refusal reasonable in light of the facts, issuance of a license would not be required.

Although this proposal may be said to involve "compulsory licensing" and has an objective embraced by earlier compulsory licensing proposals, it differs from them in two important respects. First, it avoids the nearly insuperable problem of deciding what constitutes a reasonable royalty. Issuance of a license would never be compelled until the patentee's resort to the market place on a voluntary basis had provided an answer to that question. Second, the patentee is not precluded from reserving for exploitation through his own direct production and sales the whole or any subpart of his invention's possible applications. He merely is precluded from auctioning off monopolistic positions to others, the restrictive impact of which may far exceed the real value of his invention because a division of markets rather than rights to practice the invention is the objective of the licensees.

For much the same reason, it cannot be objected that the definition of a "field" is difficult and that the proposal is unworkable for that reason. The patentee adequately defines the field in which the applicant has an option to practice when he issues the first license. Whether the field as defined in the first license constitutes a separable market in economic terms is an issue with which the court need not concern itself: it is enough that a first licensee wanted a license thus circumscribed and that the applicant now also wants such a license.

E. *Territorial Restrictions*

Most of what has been said regarding field limitations is equally applicable to territorial limitations. A pattern of licenses or assignments which carves the United States up into geographic subdivisions and thereby confers on licensees a monopoly position in their respective subdivisions has obvious potential for economic harm. Obviously, too, it makes possible imposition by the patentee of different royalty rates in different areas scaled to the demand elasticity in each area. The latter should not be a source of concern; the former should.

Again in the present context it is true that the potential for harm from market subdivision is inversely related to the economic value of the invention. The more valuable the invention, the more likely it is that royalty obligations would force an endproduct price approximating a monopoly price in the area; and the less valuable the invention, the more likely it is that licensees will be seeking not the right to use the invention so much as a legal umbrella for market division.

As with respect to field restrictions the optimum solution, in my opinion, is to compel the issuance of identical licenses to all qualified applicants with respect to any geographical area for which an assignment or a license has previously been issued. The patentee's power to auction off exclusive positions will thus be destroyed without impairing significantly his ability to capture through royalties the economic value of his innovative contribution.

Also left unimpaired is the patentee's ability to preserve for himself a territorial monopoly. Until he has issued at least one license or assignment with respect to a territory, there would be no obligation to issue any license for that territory. Hence it cannot be asserted persuasively that the proposed solution would bear more harshly on small than on large patentees; each would be able to reserve for himself an area suitable to his potential for direct exploitation and to gather royalty income from the balance of the United States. The patentee would also be able to provide for subsequent expansion of his exclusive area in anticipation of the growth of his own productive and distributive facilities. Licenses to adjoining territories could be made terminable on a given future date or at the patentee's option on or after a given date. Licenses then issued involuntarily to qualified applicants would similarly be terminable and the patentee's interest would be preserved.

Two points peculiar to territorial restrictions should be noted. The potential of territorial restrictions for economic harm has been limited by the long-standing doctrine that the patentee's monopoly is "ex-

hausted" by the first sale of a patented product.[109] Although the basis for this doctrine is probably to be found in the historic hostility to restraints on alienation[110] rather than judicial devotion to competitive processes, a major impact of the doctrine is to allow patented products to flow across the territorial boundaries of restrictive licenses. This erosion of insularity mitigates the restrictive effect of geographic monopolies. It also reduces the patentee's ability to impose differential royalty rates; to the extent there is inter-area competition through resales, disparities in demand elasticity in different areas cannot be exploited by local licensees and hence cannot be exploited through royalty differentials by the patentee.

But the check imposed by the resale doctrine is only partial. To the extent that the cost of transportation and of additional handling by the redistributor is a significant fraction of final price, a tariff wall remains to shelter the local licensee. The redistributor's costs, apart from transportation, may be substantial, among other reasons because some degree of surreptitiousness may be required to avoid impact of infringement and contributory infringement doctrines: the law is not entirely clear as to the significance of knowledge on the part of either the seller or redistributor of the fact that the latter intends to redistribute in an area foreclosed to the seller.[111] And since the courts have treated the "first sale" rule as one of metaphysics rather than of the purposeful implementation of social objectives,[112] it is difficult to predict results in any given case. Hence the existence of this conceptual doctrine, whatever its reach, should not be thought to make unnecessary or inappropriate purposeful doctrinal modifications which strike specifically at potential abuses of territorial restrictions.

A second peculiarity of the territorial restriction is the presence in

109. See, *e.g.*, Bloomer v. McQuewan, 55 U.S. (14 How.) 539 (1852); Adams v. Burke, 84 U.S. (17 Wall.) 453 (1873); Keeler v. Standard Folding Bed Co., 157 U.S. 659 1895); Bauer & Cie. v. O'Donnell, 229 U.S. 1 (1913). But see Skee Ball Co. v. Cohen, 286 Fed. 275 (1922).

110. Chafee, *Equitable Servitudes on Chattels*, 41 HARV. L. REV. 945, 999 (1928).

111. *Compare* Hobbie v. Jennison, 149 U.S. 355 (1893) (territorial assignee who knew that buyer intended to use product in another territory not liable for infringement by reason of sale) *with* Keeler v. Standard Folding Bed. Co., 157 U.S. 659, 666 (1895) ("Whether a patentee may protect himself and his assignees by special contracts brought home to the purchasers is not a question before us . . . . [H]owever . . . such a question would arise as a question of contract and not as one under the inherent meaning . . . of the patent laws.") See also General Talking Pictures Corp. v. Western Electric Co., 305 U.S. 124 (1938) where the Court declined to answer, because not posed by the record, the analogous question in the context of field restrictions.

112. "In the essential nature of things, when the patentee sells a machine . . . whose sole value is in its use . . . he parts with the right to restrict that use. The article passes without the limit of the monopoly," Adams v. Burke, 84 U.S. (17 Wall.) 453, 456 (1873).

the patent code of language asserted to be a statutory basis for such restrictions.[113] Indeed, it has been suggested that since territorial restrictions are no less harmful than price, output and field restrictions, all such restrictions must be tolerated so as not to create inconsistencies.[114]

Only by amateurish literalism or cynical distortion can it be argued that § 261 places a general imprimatur of legality on territorial restrictions. As the total content of the present section and its history shows, the purpose of the provision is totally irrelevant to the issues discussed in this paper; and the language of § 261 can be given its full intended effect quite consistently with the suggestions I have made.

Section 261 has its genesis in § 4 of the Patent Act of 1793[115] which merely authorized *assignment* by the patentee of his title and interest in his invention.[116] Statutory authorization was necessary "because since patent rights are creatures of statute and not of common law, the transfer of legal title thereto cannot be regulated by the rules of the latter system."[117]

The procedural and formal rigidities of the 19th century between law and equity, between legal and equitable title, and between contracts and conveyances lead to distinctions of substantial practical import between patent *assignments* and patent *licenses.* Made in compliance with the statute, an assignment vested in the assignee a legal title which was an adequate basis for the commencement by the assignee of infringement actions, the execution of licenses or further assignments, and application for extensions of the patent term.[118] A warranty of title was implied unless negated by the instrument.[119] Legal title prevailed over prior equitable incumbrances if obtained for value and without notice,[120] could be held by several persons as tenants

113. 35 U.S.C. § 261 (1964). "[P]atents, or any interest therein, shall be assignable in law by an instrument in writing. The patentee . . . may in like manner . . . convey an exclusive right under his . . . patents, to the whole or any specified part of the United States.

"A certificate of acknowledgement under the hand and official seal of a person authorized to administer oaths in the United States . . . shall be prima facie evidence of the execution of an assignment. . . .

"An assignment . . . shall be void as against any subsequent purchaser . . . for valuable consideration, without notice . . . unless it is recorded in the Patent Office . . . ."

114. Note, *Patent Price Restrictions and the Antitrust Laws: A Balance Upset,* 67 YALE L.J. 700, 706 (1958).

115. 1 Stat. 318, § 4 (1793).

116. *Ibid.* "[I]t shall be lawful for any inventor . . . to assign the title and interest in the said invention . . . and the assignee, having recorded the . . . assignment . . . shall thereafter stand in the place of the . . . inventor. . . ."

117. WALKER, PATENTS § 274 (1st ed. 1883).

118. *Id.* at §§ 272-95.

119. *Id.* at § 282.

120. *Id.* at § 286.

in common and perhaps by joint-tenancy,[121] and was subject to voluntary, but not involuntary, partition.[122]

But any transfer of interest that did not convey a sufficiently broad range of interest to the transferee was regarded not as an assignment but as a mere license.[123] Unlike an assignee, a licensee was merely a contracting party, not a property owner with title.[124] He received nothing more than the privilege of practicing the invention to the extent described in the license without being subject to liability for infringement.[125] Licenses could be written or oral and did not have to be recorded.[126] Unless the document expressly authorized transfer, a license was not transferable; and if made transferable, it was such only in its entirety unless expressly made transferable in parts and to a plurality of persons.[127]

Because of the important differences between assignment and license, it was important that there be a sharp definition of that range of interest which was sufficiently broad to be the subject of assignment as opposed to license. No such definition was written into the 1793 provision.[128] This deficiency was corrected in the 1836 act which assimilated to the assignment category "every grant . . . of the exclusive right . . . to make and use, and to grant to others to make and use the thing patented within and throughout any specified part . . . of the United States."[129] Transfers of that breadth were required to be recorded[130] and were regarded as the minimum quantum of interest to which the qualities of "property" and "title" attached; and with those qualities went a complex body of interpretive rules.[131] Finally in

121. *Id.* at § 292.
122. *Id.* at § 295.
123. *Id.* at § 296.
124. *Id.* at §§ 305, 306.
125. *Id.* at §§ 296-300.
126. *Id.* at §§ 303, 304.
127. *Id.* at § 310.
128. See note 116 *supra.*
129. Act of July 4, 1836, ch. 357, § 11, 5 Stat. 117 (1836): "[E]very patent shall be assignable in law, either as to the whole interest, or any undivided part thereof, by any instrument in writing; which assignment, and also every grant . . . of the exclusive right . . . to make and use, and to grant to others to make and use, the thing patented within and throughout any specified part . . . of the United States, shall be recorded in the Patent Office. . . ."
130. See note 129 *supra.*
131. See, *e.g.*, Waterman v. Mackenzie, 138 U.S. 252 (1891). See WALKER, PATENTS § 296 (1st ed. 1883).
Even today it is not unusual for courts to allow characterizations of legal relationships to control the answer to questions bearing little apparent relationship to those characterizations. See, *e.g.*, Ohio Valley Advertising Corp. v. Linzell, 107 Ohio App. 351, 152 N.E.2d 380 (1957), *aff'd*, 168 Ohio St. 259, 153 N.E.2d 773 (1958) (plaintiff had received "mere licenses" rather than leases to maintain signs on property subsequently taken by eminent domain and therefore was not entitled to compensation.)

1870 an express statement of the sanction for failure to record, nullity as against a bona fide purchaser, was added.[132] When reprinted in the Revised Statutes of 1875, the section was broken into two sentences;[133] in 1897, the sentence making certificates of notaries prima facie evidence of execution of an assignment was added;[134] in 1941 the section was expanded to include applications as well as patents;[135] and in 1952 the several sentences were rearranged into their present order, placing the paragraph relating to certificates between the authorization to assign and the requirement of recordation.[136]

As a consequence of these technical modifications the basic authorization to transfer by assignment rather than by license interests of a minimal quantum now stands alone in the first principal paragraph of § 261 and is susceptible to being misread as a legalization for all purposes of assignments of exclusive right to a specified part of the United States. But that reading is wholly erroneous. The section remains today what it has been for over a century: a demarcation of the dividing line between interests sufficiently extensive to be transferred by assignment and those which are to be transferred by license. The modern reader, less sensitive than his predecessors to refined distinctions between contract and conveyance, legal and equitable title, and persons who are and are not proper parties in interest, is liable to err; but internal corroboration of the original purpose is present in the section: It deals with assignments, not licenses; and the sentence that is critical for present purposes still reads, "The . . . patentee . . . may, *in like manner* (*i.e.*, by assignment), grant an exclusive right . . . to . . . any specified part . . . ."

The section should be given the full import intended for it. The assignee of an exclusive territorial right has "property" and may sue for infringement within his territory. His document should be interpreted by the suppletive rules appropriate to assignments, and so forth. But to allow this section to preclude intelligent harmonization of divergent objectives of the patent and antitrust laws would be to parody the process of statutory interpretation. The statutory phrase antedates the Sherman Act by more than fifty years, and neither at the time of

132. Act of July 8, 1870, ch. 230, § 36, 16 Stat. 198: ". . . part of the United States; and said assignment . . . shall be void as against any subsequent purchaser . . . unless it is recorded in the Patent Office. . . ."
133. REV. STAT. § 4898 (1875): ". . . part of the United States. An assignment . . . shall be void . . . ."
134. Act of March 3, 1897, ch. 391, § 5, 29 Stat. 692.
135. Act of Aug. 18, 1941, ch. 370, 55 Stat. 634: "Every application for patent or patent . . . shall be assignable . . . ."
136. 35 U.S.C. § 261 (1964).

its passage nor at any time during the interim hundred forty years has Congress given any evidence that the language represented a considered judgment on the substantive propriety of territorial restrictions and their economic consequences.[137] On its face the section leaves untouched the validity of territorial restrictions in licenses; it should be construed as irrelevant to the substantive validity of similar restrictions in assignments.

VIII. In Anticipation of Rebuttal

In retrospect it can be seen that a major theme of the proposals made in this paper is that patentees should be blocked from conferring upon licensees the ability to monopolize generic endproducts and thus to capture the monopoly profit potential in the demand for the endproducts rather than the profit potential in the demand for the invention. I use the term "generic endproduct" to emphasize that the theme has embraced situations where the endproduct itself was patented. A patented endproduct may have more or less close functional substitutes, and if licensees are deprived of incentive to substitute functionally equivalent endproducts against the patented endproduct, the situation is analogous to those in which incentive to substitute unpatented inputs for the patented input is blocked.

To this theme it might be objected that the patentee who licenses would be put in a less advantageous position than one who exploits directly through his own manufacture and sales. In the case of direct exploitation, a patentee will not substitute against his patented input or against his patented endproduct. Surely the economy is no worse off, the argument would run, if licensees are enabled to behave in essentially the same way. Indeed, it could be said, the economic value of the invention is best measured by the extent of restriction and profits that would attend direct patentee exploitation; and to prevent the patentee from capturing a part of these incremental profits by splitting them with licensees on whom he has conferred sheltered positions is to deprive the patentee of part of that value. The argument concludes

137. I have been unable to find any discussion of the territorial assignment language in § 261 either in Congressional reports or in floor debate. The general references that have been made to the section as a whole confirm the view that it has been carried forward without any purpose to alter the meaning it had in 1836.

See, *e.g.*, CONG. GLOBE, 41st Cong., 2d Sess. 2683 (1870) ("[t]here are not new nor additional provisions. . . ."); 29 CONG. RECORD 901 (1897) (the only change is "a matter of detail."); H.R. REP. No. 790, 77th Cong., 1st Sess. 1-2 (1941) (the only change intended was to make § 261 . . . "applicable to assignments of application as well as of patents."); H. R. REP. No. 1923, 82d Cong., 2d Sess. 3, 5, 9 (1952) (purpose of the bill was the codification and enactment of title 35 with only minor procedural changes).

by asserting that an important consequence of depriving patentees of the ability to auction off positions of market power over generic endproducts is that fewer licenses will be issued and exploitation by direct patentee sales will be more frequent with the ultimate consequence of greater concentration and diminished opportunities for small potential licensees.

Although this line of objection cannot be dismissed as patently frivolous, I find it wholly unpersuasive for several reasons. First, it is inconsistent with the basic structure of the patent system. The only possible justification for adopting the private monopoly device as a mode of subsidizing innovative activity is to take advantage of free market assessment of the value of a patentee's contribution. A direct system of government subsidy would avoid all the economic damage of monopolistic restraint and would certainly be preferable but for the enormous political and administrative difficulties of determining an appropriate reward.

But free market assessment is thwarted if the patentee is permitted to sell not the bare right to use the invention but the right to monopolize generic endproducts in which it is useful. The two things are not the same, the latter always being as great or greater in value as the former; and moreover they do not bear any constant proportion to one another. To permit patentees to sell endproduct monopoly positions both impairs the degree of precision with which comparative invention values are assessed and generally increases the values assigned.

I do not imply that the market for innovation is perfect and would make highly refined assessments but for license restrictions. It is not. As Professors Kaysen and Turner have argued with admirable style, the market is highly imperfect: "innovation is a lottery, and it is the high prizes that count.[138] But that unavoidable fact does not justify the further avoidable impairment of market operation that results from requiring that it assess invention value by assessing generic endproducts.[139]

Even when the invention is "essential" to the generic endproduct in the sense that without use of the invention no functionally equivalent

138. KAYSEN & TURNER, ANTITRUST POLICY 163 (1959).

139. A footnote is the appropriate place to anticipate the observation that the ultimate logical conclusion to which my position leads is that patentees should be required to license in all cases so that they will be denied the possibility of endproduct monopolization through self-production. The observation recalls a parable which ends, "I know how to spell 'banana' but I don't know when to stop."

This last, logical, step would require, for reasons of administration, denial of self-utilization until one or more licenses were issued.

product can be produced at a cost anyone is willing to pay, the value of the invention is what the market places on use of the invention, not on the endproduct. For the susceptibility of the invention to input substitution in the production process is an unavoidable element of value. *Hamlet* cannot be played without Hamlet; but the merit of the performance will depend heavily on the rest of the cast, the scenery, and even the padding on the seats.

In the more frequent situation where the invention is not essential to the generic endproduct because of the availability of either old or easily discovered new substitutes for it, it is of vital importance for market assessment to focus on invention use. If the patentee is permitted to share a monopoly over the endproduct, incentive, among licensees to bring those substitutes into play is diminished or destroyed; and since the licensees will usually be those in the best technological position to learn of the substitutes and assess them comparatively with the invention, alternative routes to the generic endproduct will often be obliterated.

When the alternative route to the endproduct is already known, restrictive license patterns blend indistinguishably with the classic patent-umbrella cartel. And distinction is difficult between what is already known, what is readily perceptible and what, itself, ranks as invention. To require that invention utilization rather than endproducts be the focus of market assessment would restore incentive to resort to alternatives and would constitute a major step toward eliminating this type of horizontal parallelism. That step is important because of its direct economic consequences in product markets. It is important because of the potential savings of scarce human resources now invested in antitrust enforcement. And it is important because decisions in such cases, turning as they must on highly speculative assessments of the comparability of patent value with obversed restraint, are a source of considerable business community disaffection for the antitrust laws generally.

Finally, the prediction that substantially fewer patentees will issue any licenses does not trouble me at all. Its accuracy is doubtful. Holders of strong and valuable patents—patents that will serve as the cornerstone for an industry—who have or can acquire the capacity to exploit the invention fully by self-production probably do so now. That course yields monopoly profits for patent life, affords a strong market position thereafter, and avoids the administrative costs of establishing and policing a license system. Those who neither have nor can acquire adequate capacity will license because they must.

There may be a shift at the margin, but there is little reason to think the shift will be substantial. Any negative value that reasonably may be assigned to that shift will be far outweighed, I believe, by the impact of the proposals on markets affected by patents which are not both legally invulnerable and economically valuable.

It is wholly unimportant whether willingness to license is diminished on the part of holders of patents which are either legally vulnerable or of minor economic value; and the vast preponderance of patents falls in these categories. No shift even at the margin is predictable, for to insist upon exclusive utilization will provoke assaults on validity of patents in the first category and investigation of alternative methods so as to circumvent those in the second category. If the patentees choose to license, they will do so for royalties commensurate with invention value and the probability of validity. With rare exceptions readiness of patentees to license is a mixed blessing for which no substantial price should be paid.

IX. In Conclusion

Subsidization of innovative activity is probably necessary to achieve the devotion to that activity of an optimum flow of resources.[140] Adoption of the monopoly device for this purpose has the great disadvantage of underutilization of all significant inventions but may be justified by the preference for free market assessment of invention value over administrative assessment.[141] It is essential to market assessment that comparability exist between the benefits that flow to the patentee and the burden imposed upon licensees; if in their bargaining process they are allowed to externalize the burden of payment by constraints on licensee conduct or by adoption of certain royalty structures, the adverse interest of the licensee no longer serves to check the degree of monopoly and restraint imposed.[142]

Apart from blatant techniques such as the physical destruction of competitors, impairment of comparability occurs through economic arrangements which unnecessarily extend the restraining impact of high price to goods and services other than the invention itself. Some minimal familiarity with price theory is essential to understanding the impact of different arrangements.[143] In particular two economic phenomena recur repeatedly in analysis of licensing problems: that of

140. Pp. 267-72 *supra.*
141. Pp. 272-75 *supra.*
142. Pp. 277-79 *supra.*
143. Pp. 279-80 *supra.*

differential pricing in accordance with demand elasticity, and that of arrangements which affect substitution of inputs for the patented input. Because incremental use of an existing invention is costless, the classic concept of economic discrimination cannot be applied in this context; and with limited exceptions the law should not attempt to control differential pricing practices.[144] Arrangements which block the substitution for the patented product of other factor inputs, on the other hand, is a serious problem to which inadequate attention has been given.[145] These considerations lead to a number of specific proposals for judicial and legislative action.

Judicial Action

1. Royalty structures based on licensee sales of unpatented endproducts should be prohibited on the ground that they restrain outputs and increase prices of the endproducts to an extent not comparable with the value of invention.[146] Adoption of such a royalty structure should be held to constitute misuse of the licensed patents and a violation of § 1 of the Sherman Act.

2. Royalty structures, other than those imposing a constant charge per unit of an appropriate input, which, when expressed as a percentage of licensee sales, increase as licensee output increases, should be prohibited if there exists a licensee other than the one to whom the royalty structure applies who is in competition with the one to whom the structure applies.[147]

3. An exception to the prohibition against tie-in sales should be recognized where the tying product is patented, the total market for the tied product substantially exceeds the submarket for tied sales and the tied product is sold for prevailing market price.[148]

4. Package licensing of patents to which a common input royalty base is appropriate should not be prohibited provided that the license gives the licensee an option to terminate the license on the date of expiration of any one patent included in the package.[149]

5. Restrictions on either a licensee's price or output should be prohibited.[150]

6. Restrictions on the field in which the licensed patent can be

144. Pp. 280-99, 306-12 *supra*
145. Pp. 299-306 *supra.*
146. Pp. 299-312 *supra.*
147. Pp. 314-18 *supra.*
148. Pp. 318-22 *supra.*
149. Pp. 322-29 *supra.*
150. Pp. 329-39 *supra.*

applied should not be prohibited, but when such a clause is used, a subsequent refusal to issue to any qualified applicant a license identical to any license in which a field restriction appears should be held a misuse of the licensed patent.[151]

7. Restrictions on the geographic territory in which an invention can be practiced should not be prohibited. But when a license containing such a restriction is issued, a subsequent refusal by the licensor to issue to any qualified applicant a substantially identical license should be held a misuse of the licensed patent by the licensor. And when an exclusive territorial assignment is executed, a subsequent refusal by the assignee to issue a license substantially identical to his assignment should be held a misuse of the assigned patent by the assignee.[152]

Legislative Action

Legislative enactment of any of the foregoing proposals would be wholly appropriate although each of them can also be implemented appropriately by judicial action. Legislation defining the characteristics which make an applicant a "qualified applicant" for the purposes of proposals 6 and 7 would be particularly helpful.[153]

If the courts continue to exhibit blind literalism in construing the conveyancing clause in § 261, an amendment explicitly limiting the effect of that language to its originally intended purpose will be appropriate.[154]

Only one of the proposals offered—one offered rather tentatively—is necessarily dependent on legislation for its implementation. If it is thought important that resort to differential royalties based on elasticity differences be subjected to legal controls in those instances when the practice arguably causes economic harm, Congress should establish administrative machinery through which to permit exceptions to the general rule against unpatented endproduct royalties. Exceptions should be narrowly confined to those situations where, by reason of absorption of substantial real cost advantages through differential royalties, there is reason to believe that aggregate output of an industry is more restricted by the differential royalty structure than by blocking input substitution.[155]

151. Pp. 339-46 *supra.*
152. Pp. 347-52 *supra.*
153. Pp. 345-46 *supra.*
154. Pp. 348-52 *supra.*
155. P. 310 *supra.*

Appendix

Some Basic Economics

For any particular product, a schedule can be made showing the quantity that will be purchased in a unit of time if various prices per unit of product are charged. Such a schedule, when represented graphically, yields a demand curve for the product. Price per unit is measured on the vertical axis; units of product sold per unit of time are measured on the horizontal axis. Since more units of product can be sold in a unit of time at lower prices per unit of product than at higher prices, the demand curve will slope downward from left to right.

The curve will slope downward to the right for two reasons. First, customers who would buy some, but only a few, units at very high prices, will buy a larger number of units at lower prices. Second, other customers will not buy any units at very high prices but will buy some units at a lower price. Finally, at some quite low price, perhaps at zero, all potential customers would be buying as many units as they have any use for; they could be induced to take more only if they were paid to do so.

The slope of the curve may be gradual or steep, depending on the characteristics of the product and on whether reasonably close substitutes for the product are available. If the product is medicine, for example, customers will pay high prices for the prescribed dosage when they are ill; but they have no use at all for larger amounts even when ill, and they have no use for any at all when they are well. The quantity of medicine that can be sold depends principally on population size and frequency of disease and only secondarily on price. The slope of the demand curve for medicine will be quite steep at all realistic prices. Phonograph records, on the other hand, might have a gradually sloped curve at all but extremely high or extremely low prices. People can get along without them altogether; but at low prices many people would accumulate very large numbers.

The slope of the curve at any given point may be referred to as its elasticity and is measured by comparing the percentage change in quantity purchased which results from a percentage change in price. If a 2 per cent change in quantity is brought about by a 1 per cent change in price, the curve, or demand, in that price range is elastic and has an elasticity of 2.0; if the same 2 per cent change in quantity is produced by a 2 per cent price change, elasticity is unitary, and equals 1.0; and if a 4 per cent change in price produces that result, demand is inelastic and equals 0.5.

The curve discussed in the preceding paragraph represents demand for a *product.* A seller of that product will not necessarily face the same demand curve. The schedule of how much a seller can sell at various prices will be the same as the schedule for the product if, but only if, he is the only seller of that product—if he has a monopoly on the product. If there are other sellers of the same product, he will obviously have to share the market and will only sell some fraction

of the total quantity purchased at any given price. To characterize the situation as "sharing" might suggest that the demand curve facing the seller would be a new curve representing a constant fraction, at every point, of the product curve and lying somewhat to the left of it. But the existence of other independent sellers has another, more significant, impact: it assures the availability to consumers of the same product at roughly the prevailing price; and should one seller raise his price significantly, consumers will buy more from other sellers and the first seller will sell substantially less. The "sharing" need not occur in a stable proportion. Hence the existence of other independent sellers of the same product not only means that each seller's demand curve will lie to the left of the product curve but also that, in general, it will be more elastic, at least for prices higher than that presently prevailing.

The last qualification is made necessary by the possibility that other sellers will learn of a price reduction by the first seller and respond by reducing their prices. If this occurs, the first seller will not increase his proportion of the total product sales; but he will sell a larger absolute quantity since total sales will increase at the lower common price in accordance with the demand schedule for the product. It will be noted that the preceding paragraph tacitly assumed that other sellers would not respond to a price increase by the first seller and hence that his proportion as well as his absolute quantity of sales would decrease. This assumption is generally realistic even if other sellers do learn of the price change by the first seller; for unless his change was motivated by a cost increase which his competitors also experienced, the competitors need only continue selling at a price heretofore found satisfactory in order to increase their proportion of the market at his expense. Even more obviously the assumption is realistic if other sellers do not learn of the price change of the first seller.

On the basis of the content of the preceding paragraphs demand curves of several basic types may be drawn. In figures 1 through 3 the lines on the right are hypothetical product curves. Some are more steeply sloped than others, and these differences may be thought to represent variations in the degree to which the products are important to consumers both because they satisfy needs that are, or are not, fundamental and because reasonably adequate substitute products satisfying those needs are, or are not, available.

In each of the three figures the curves on the left are hypothetical demand curves for individual sellers. No separate seller's curve has been drawn for the monopolistic seller—unlike other sellers his curve is the product curve of the product he monopolizes. The seller's demand curve in figure 1 is the curve facing a seller in pure competition and is drawn over the range of outputs he might reasonably contemplate. In this market there is a very large number of sellers of a homogeneous product, each of whom sells a very small part of total product output. Each can sell all he produces at the existing price and can sell nothing at higher prices. His curve is perfectly, or infinitely, elastic.

In figure 2 the left hand curve is one that might represent the situ-

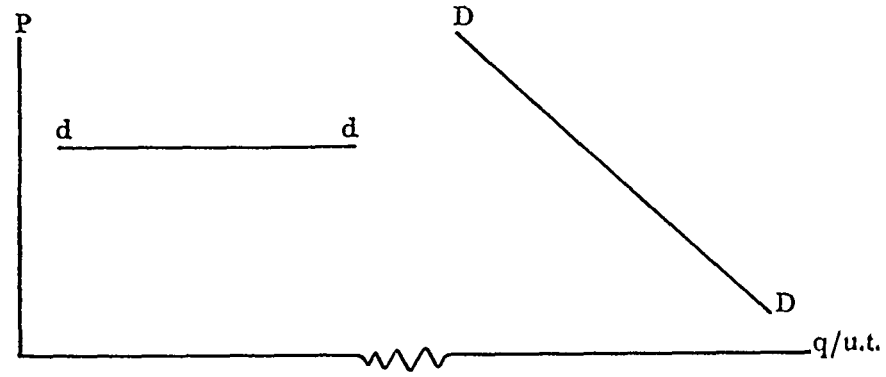

FIG. 1. Demand curves for seller in perfect competition (d-d) and for his industry or product (D-D).

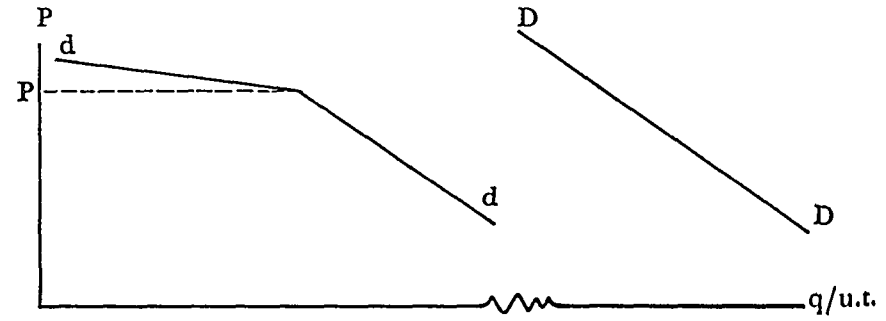

FIG. 2. Demand curves for seller in oligopoly (d-d) and for his industry or product (D-D).

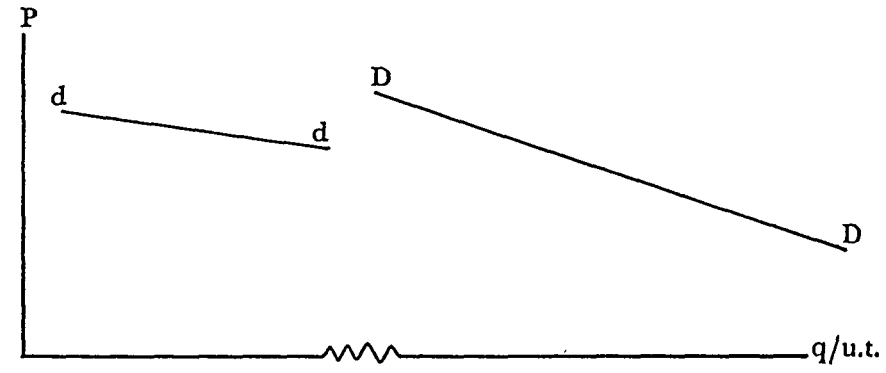

FIG. 3. Demand curves for seller in imperfect competition (d-d) and for his industry or product (D-D).

ation of a seller in an oligopolistic industry, one comprised of a small number of sellers, three, seven, perhaps as many as ten. The kink in the curve occurs at the existing price, and hence a price line has been added in this diagram. Below the kink the curve roughly parallels the

product curve reflecting the probability that competitors will match price reductions; above the kink the curve is more flattened, reflecting the probability that competitors will not match price increases.[1]

In figure 3 the left hand curve represents an imperfectly competitive market, perhaps one possessing too many sellers to behave in the oligopolistic manner illustrated in figure 2, or perhaps one where product differentiation is sufficient to soften extreme oligopolistic interdependencies, but not one possessing both product homogeneity and the very large number of very small (relatively) firms necessary for perfect competition.

Generally within a short time period, sellers will sell each unit of product at the same price for which they sell every other unit of product, either because of legal constraint on their conduct or because competition forces them to do so. They have no wish to sell at prices lower than the lowest profitable price, and they cannot extract any higher price from any one consumer because a competitor would promptly offer to satisfy that consumer's needs at a lower but still profitable price. If a seller must sell each unit at the same price as every other unit, the slope of the demand curve he faces becomes critical. A contemplated price reduction motivated by the desire to make more sales per month must be evaluated by the seller very differently than if the new lower price would apply only to the new sales that would not be made during the month but for the price cut. The price reduction will apply to and reduce the profitability of all the other units of product which could have been sold at higher prices. Hence the net revenue that will be garnered by the sale of one additional unit per unit of time will always be less than the price of that unit if an across-the-board price reduction is necessary to make that sale; and a general reduction will be necessary if the seller's demand curve is sloped at all and if all units must be sold at the same price per unit.

For any demand curve it is possible to draw another curve representing the net effect on revenues of making ever larger numbers of sales. The sale of one unit per month at the highest price anyone will pay for it garners the price of the unit. But the sale of two units per month, necessarily at a slightly lower price, garners twice the price of each unit which is less than the sum of the one-unit-sale-price and the two-unit-sale-price. So it is with each successive number of units; the return to revenue as a consequence of the last, or marginal, sale always being less than the price at which the sale was made. Since the demand

1. This paragraph in particular involves extreme oversimplification. The critical feature of an oligopolistic industry is that the number of firms is sufficiently small that the conduct of one firm significantly affects, and is mutually recognized as affecting, the situation of each other firm. To make any general statement about the demand curve facing a firm in such an industry it is necessary to make some assumption about the way other firms will react to the first firm's conduct. The assumption in the text, that other firms will match price reductions but not price increases, is the assumption most frequently made in the literature; and the curve illustrated corresponds to that assumption.

curve represents the *price* at which each of these successively larger quantities can be sold in a unit of time, the curve representing the marginal revenue to be derived from the successively larger quantities will lie below the demand curve at every quantity after the first unit; and the marginal revenue curve will itself slope downward so long as the demand curve follows a linear descent.

Reflection on the foregoing paragraphs will now make apparent several features about marginal revenue curves generally. First, if a seller's demand curve is horizontal, indicating that he can sell any quantity he wishes to produce at the same price, his marginal revenue curve will be coincident with his demand curve, reflecting his ability to add to his revenues the full amount of each successive sale price. Second, the marginal revenue curve is always coincident with the demand curve at the first unit of quantity. Third, if the demand curve slopes down smoothly, marginal revenue will become equal to zero and then turn negative (cross the horizontal axis) at that quantity at which the demand curve achieves unitary elasticity; for beyond that point a given percentage decrease in price will yield a lesser percentage increase in quantity sold, and the effect of any such change will be to decrease aggregate revenue.

In figures 4 and 5 are shown the same seller demand curves as were shown in figures 2 and 3 together with their corresponding marginal revenue curves. Note that there is a discontinuity in the marginal revenue curve at the point where the oligopolist's demand curve kinks.

In describing a seller's economic situation and predicting his responses to that situation, the seller's cost patterns are of importance equal to that of demand characteristics for his product. Demand curves describe graphically how much product he can sell at various prices; cost curves describe graphically how much product he can produce at various costs; the relationship between the two sets of curves indicates the level of operation that will take fullest economic advantage of his situation.

Analysis of cost characteristics is facilitated by separating the cost elements into two basic categories: fixed costs are those that remain constant in amount regardless of how the seller changes his level of operations within the output range available to him with his existing physical facilities and long-term contractual arrangements; and variable costs are those that will increase or decrease as he chooses to operate at higher or lower levels of output. The line between the two categories is not sharp, even if some precise time period is arbitrarily selected as the period within which to test for variability. And as the time period being considered lengthens, larger portions of costs will fall into the "variable" category. Nevertheless, the categories are analytically useful.

Each category of costs may be described in either of two ways: in aggregate or total amount, or in relation to units of output, as proves to be more helpful for any given purpose. By definition, the aggregate of fixed costs is the same at all levels of output. Obviously most items of variable costs will increase as output increases and aggregate vari-

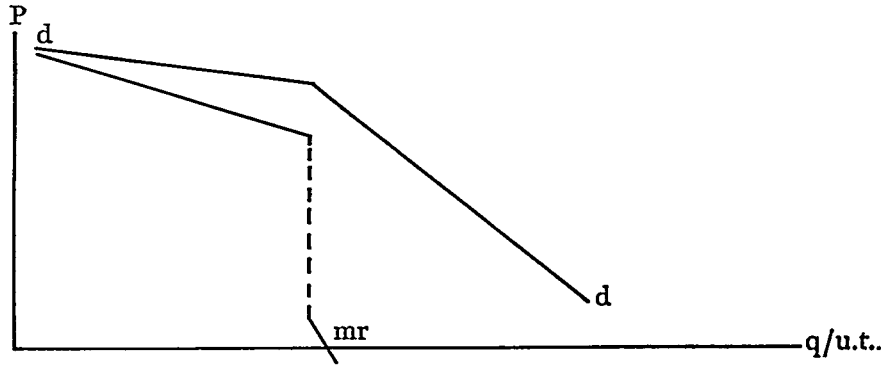

FIG. 4. Demand curve for seller in oligopoly (d-d) and corresponding marginal revenue curve.

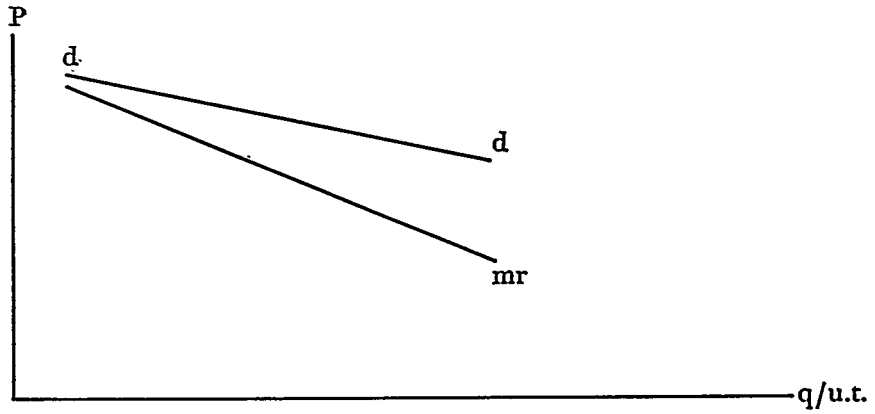

FIG. 5. Demand curve for seller in imperfect competition and corresponding marginal revenue curve.

able costs will increase as output increases. Representations of total fixed cost, total variable cost and total cost are shown in figure 6. Average fixed costs, average variable costs and average total costs are obtained by dividing each category of aggregate cost by the number of units of output. Representations of those functions are shown in figure 7.

Another way, and a very useful way, of thinking about and representing variable costs is to focus on the *additional* amount of cost necessitated by the production of each additional unit of product. For any unit of product, the additional cost attributable to its production is called its marginal cost. Since fixed costs do not change, the additional cost of producing one more unit is necessarily attributable to variable cost changes; and hence the marginal cost curve bears a close relationship to the average variable cost curve. The average variable cost of the first unit of product obviously is equal to total variable cost at that production level and is also equal to the

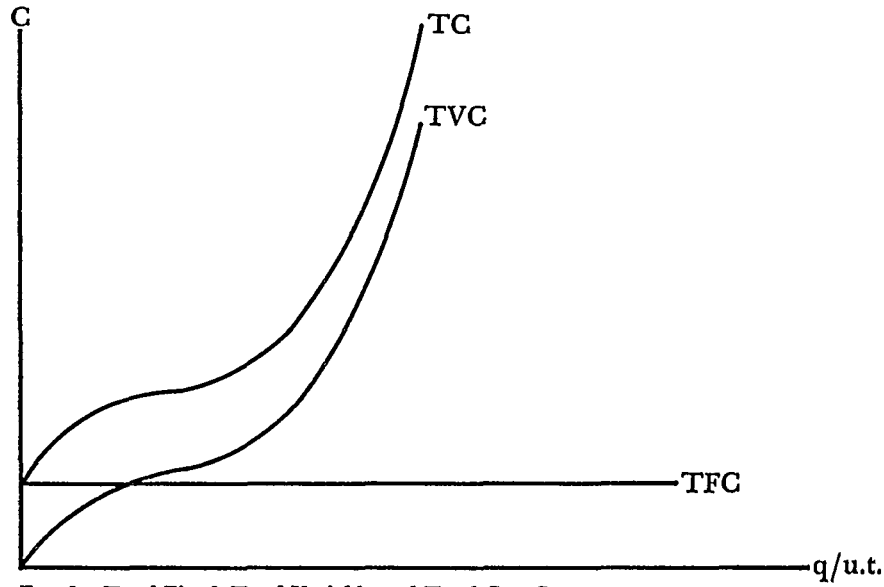

FIG. 6. Total Fixed, Total Variable and Total Cost Curves.

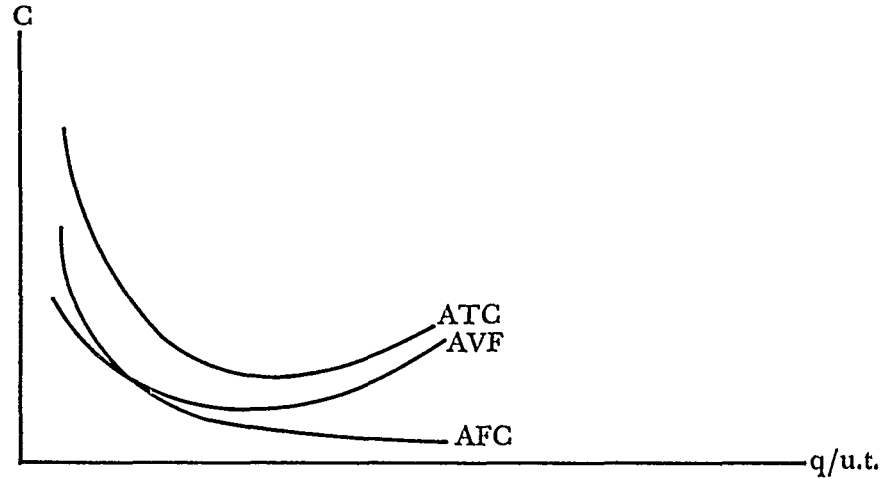

FIG. 7. Average Fixed, Average Variable and Average Total Cost Curves.

additional cost of moving from no production to the one-unit level. Hence these three presentations of variable cost all start at the same point. As successively larger output levels are considered, the following relation will be seen to exist: so long as the incremental cost of producing a unit is less than the incremental cost of producing the preceding unit, both average variable cost and marginal cost will fall; and since average variable cost is affected by the higher incremental costs of earlier units but marginal cost is not so affected, marginal cost will be less than average variable cost. At that output level where returns to larger scale cease to yield successively lower incremental costs for

successive units, marginal cost will reach its minimum; it will still be below average variable cost, and as the marginal cost curve flattens and first starts to rise, average variable cost will continue to fall. At that output level where the incremental cost of a unit exceeds the average variable cost of all preceding units, average variable cost will start to rise. Hence the marginal cost curve reaches its minimum at a smaller output level than will the average variable cost curve; it will rise and intersect the average variable cost curve at the minimum of the latter; and it will lie above the average variable cost curve at all outputs where the latter is rising and below it at all outputs when the latter is falling.

These precise statements about the relationship of marginal cost to average variable cost can be made without qualification because the relationship is a mathematical one implicit in the definitions of the two concepts. The statements about the relationship are "true" in the absolute but limited sense that two plus two may be said to equal four. No empirical data is necessary to make the statements; on the other hand the statements tell us nothing about the real world. The relationship between demand and marginal revenue is also of this mathematical character. Statements of this character are to be contrasted with statements about observable phenomena which are susceptible to being shown true or false by empirical data. The statement that a product demand curve always slopes downward to the right is of this latter type, and it is only the factual validity of that statement that gives the related mathematical relationships practical utility.

Empirical statements about costs similarly are prerequisite to useful application of the curve relationships described above. Despite their empirical quality, I will make such statements without citation to supporting data on the grounds that they are intuitively probable, that to varying degrees they have been validated elsewhere and that they are regarded as common knowledge in economic literature.

Over that range of outputs available to the firms within a time period too short to permit major alterations of existing physical plant and long-term contractual arrangements, average variable cost usually will be high at very small and at very large outputs and will reach a minimum at an intermediate point. This is to say only that the firm will be most efficient at some fairly narrow range of outputs, presumably that range contemplated when its plant and long-term contractual arrangements were made. If the firm is viewed prospectively, however, over a time period long enough to permit more fundamental rearrangements, the range over which efficient production may be achieved will broaden. In almost all industries, there are some returns to scale; hence even in the long run average cost will be high and declining at very low outputs. If unavoidable inefficiencies of very large scale are encountered, the curve will rise again at those large outputs. The long-run average cost curve is one which is tangent to a series of alternative short-run curves. Figure 8 illustrates a long run cost curve that might be abstracted from a series of short run alternatives.

Profit maximization over some time span is assumed to be the basic

motivating factor of all economic behavior. The course of action which will maximize profit in any time span is indicated by superimposing the cost and demand curves appropriate for that period. To produce a unit of output for which the incremental revenue yielded by its sale would be less than the incremental cost of producing the unit would be inconsistent with the purpose to maximize profit. Similarly, that purpose would be thwarted by failure to produce a unit whose incremental cost was less than the incremental income its sale would bring. Figure 9 illustrates the point. Hence the firm will

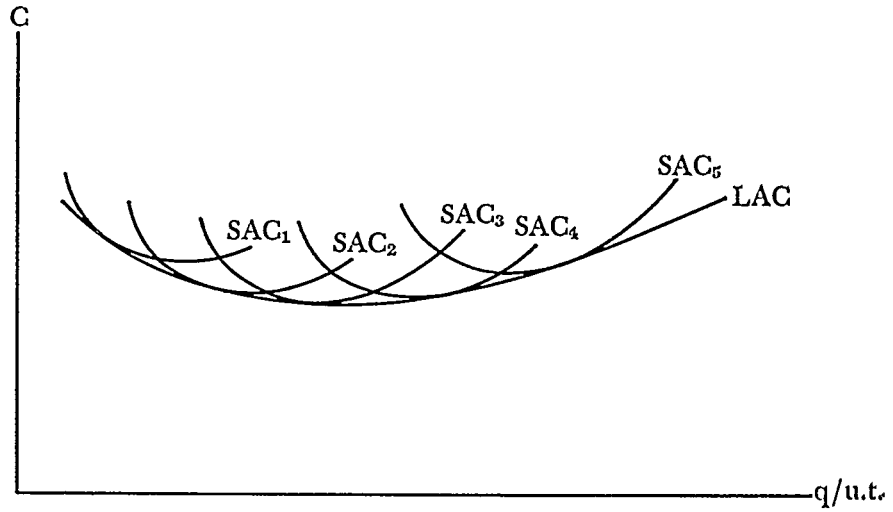

FIG. 8. Long Run Cost curve derived from a series of alternative short run possibilities.

operate at the output level where marginal cost equals marginal revenue; it will charge a price indicated by the demand curve at that output level; its gross revenue will be q units multiplied by $\$p$, the area of the rectangle $oqp'p$; its total cost will be that portion of the gross profit rectangle which lies below the line $TC\ TC'$. Since, as a matter of definition in economic analysis, a fair return to invested capital is included in "cost"—the "opportunity cost" of the capital involved—any profit in excess of cost thus defined, here the part of the gross profit rectangle $pp'TC'TC$, is defined as monopoly profit.

The fact that only marginal cost is relevant to profit maximization deserves emphasis. The point p' in the demand curve directly above the intersection of marginal cost and marginal revenue is the optimum price for the firm whether that point is above average total cost or below it. If p' is above ATC, a monopoly profit will be earned at that price. If p' equals ATC normal returns to capital will be earned. If p' is less than ATC losses will be sustained; but to adopt any other price and output would merely increase the amount of loss. Fixed costs thus are relevant to the ultimate profit or loss position of the firm, for they make up part of total costs; but they are wholly irrelevant to the optimum level of operation in the short run, for by definition they will be incurred whatever output level is adopted.

The significance of monopoly can now be stated in two alternate terms. In the more traditional focus—the impact of monopoly on the economic system—monopoly is undesirable because it results in a lesser output than would be forthcoming under competitive conditions assuming production costs are the same under competition and monopoly. Since the demand curve for the monopolistic firm is the same curve that pertains to the product produced, it will slope downward

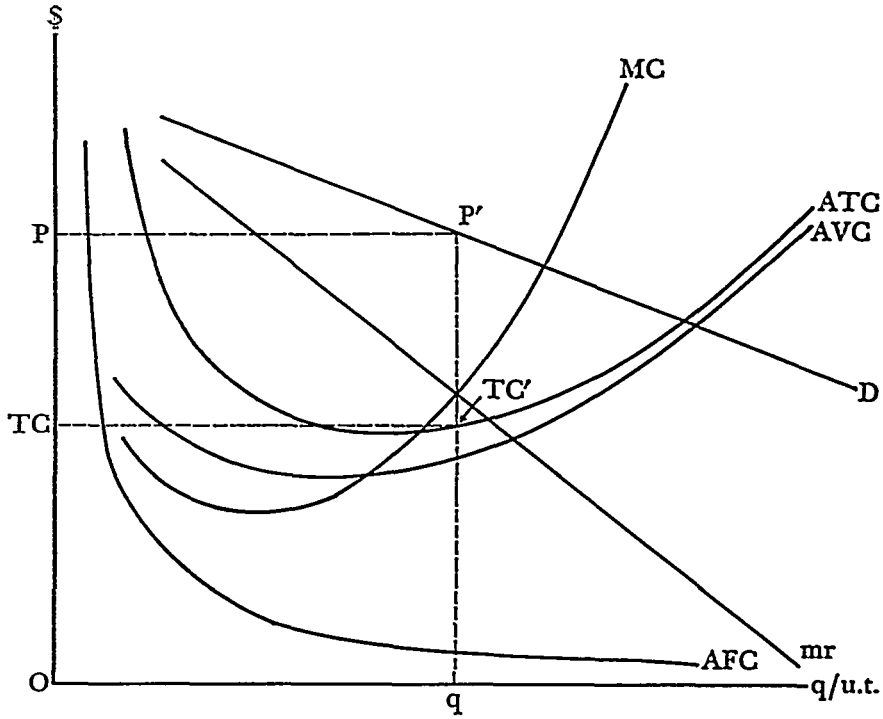

FIG. 9. Profit maximization under non-competitive conditions.

to the right; and the marginal revenue curve that corresponds to it generally will slope even more sharply. Hence the marginal revenue curve will intersect the marginal cost curve at an output less than would be dictated by a less steeply sloped demand function. More of the product could be produced by the monopolist at costs less than the prices buyers would be willing to pay for additional units; hence there is underutilization in the economy of the monopolized product itself and of the labor, raw materials and other production factors consumed in its production.

Alternatively the significance of monopoly can be described from the standpoint of the monopolist. It benefits him in several respects. He need not share but can garner to himself all the profits that are potential in the market for the product. To the extent there are economies of scale obtainable within the range of output called for, he can obtain them without concern that a competitor's efforts to follow

a similar course will result in excess capacity. Unless the total market is smaller than the output of an efficient plant, his marginal costs will equal or exceed his total costs, and he will earn an unusually large percentage return on his investment because price will exceed marginal cost. These are the private gains traditionally associated with monopoly.

It is not true, however, that restricting production to that indicated by the equality of marginal revenue and marginal cost is inherently advantageous to the monopolist. The monopolist would profit even more if he produced a larger output per unit of time, that indicated by the intersection of marginal cost and demand, provided only that he could prevent the lower prices at which the additional output must be sold from affecting his ability to sell at higher prices the other units of output. Up to that output where marginal cost equals demand (price) there are profits to be made. He could achieve this ideal situation if he were able to extract from each buyer the largest amount that buyer would be willing to pay for the first unit the buyer needed, and then to extract the largest amount the buyer would be willing to pay to get a second unit, and so forth for each unit and each buyer until he was no longer able to sell another unit to any buyer at a price exceeding marginal cost. Under these highly theoretical conditions the monopolist would maximize his returns, and he would also produce a larger output such that price equaled marginal cost. Hence restriction of production is not inherently advantageous to the monopolist but rather is the tactic which maximizes his return so long as he is subject to the constraint, practical or legal, that a single, uniform price be charged for all units of output.

Although the possibility just described is wildly theoretical, its implications have practical applications. The fact that the demand curve does lie, in part, above the level of uniform price indicates that there are some customers who are willing to pay more than they are presently paying for at least some units they are presently buying. Those buyers, with respect to those units, are receiving benefit from the "one uniform price" constraint and from the fact that the monopolist's self-interest dictates that he produce up to the output where marginal cost equals marginal revenue. The aggregate of this benefit to all who receive it is called consumer surplus and is represented graphically by the triangular area enclosed by the vertical axis, the price line and the demand curve in figure 9.

A firm is said to practice economic, or monopolistic, discrimination if it charges different prices for different units of the same product, or charges prices $p1$ and $p2$ for different but related products having marginal costs of $mc1$ and $mc2$ such that the ratio of $p1$ to $mc1$ is unequal to the ratio of $p2$ to $mc2$. If there is competition for the sale of each unit, discrimination cannot occur;[2] for a firm's attempt to extract

2. One type of discrimination, as defined, can occur even under "perfect competition." It is illustrated by "off-season" rates by resort area hotels: Marginal cost of accommodating off-season customers is no less, but they are charged less than peak-season guests. It may be objected that the example is not one of perfect competition, and in a

more than the prevailing price for any unit will be defeated by other firms' willingness to sell at prevailing price. Hence the existence of discrimination depends upon the lack of alternatives available to the disfavored buyer and hence upon some degree of monopoly.

Generally speaking, the nature of the practical constraint that prevents the monopolist from charging different prices for different units

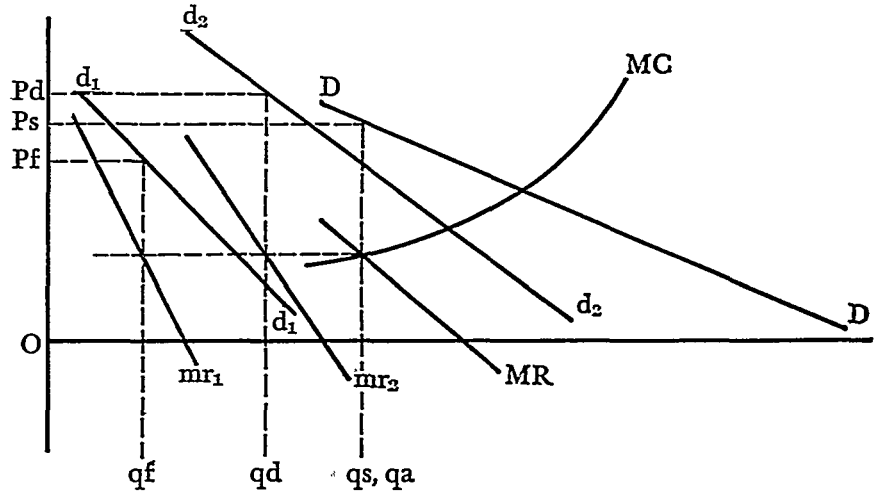

FIG. 10. The two submarkets are represented by demand curves d_1d_1 and d_2d_2 and their corresponding marginal revenue curves mr_1 and mr_2. Together they constitute the total market DD and its marginal revenue curve MR. Marginal cost at the appropriate total output is indicated by the intersection of MR and MC, and its value is superimposed on the segregated marginal revenue curves by the broken construction line. Under uniform pricing quantity qs would be sold at price Ps. Under discriminatory pricing quantity qf is sold at price Pf in the favored market, elasticity of demand being greater in that market in the relevant price range; and qd is sold at Pd in the disfavored market. Revenue is always increased by discrimination: the areas $(Pf \times qf) + (Pd \times qd) > Ps \times qs$. In the case shown, since the several demand curves are straight lines, the rates of change of the several elasticities are the same and output remains unchanged: $qf + qd = qa = qs$.

of product, thus capturing consumer surplus and augmenting profit, is the potential resale of his product by those who buy cheap to those who buy dear. To the extent that the monopolist by some device can segment his customers into groups which differ from one another in their demand characteristics and can prevent retransfers of product between the groups, he can increase his profits and diminish consumer surplus. This segmentation may be possible, for example, on a geographic basis if transportation costs are a substantial part of total cost of the product; for inter-area resale will not then be profitable. Where a service rather than a commodity is the firm's product, segmentation is generally feasible because retransfer between segments cannot occur.

To the extent the monopolist is able to segregate his customers,

he maximizes his return by selling to each segment of the market that quantity of product necessary to make the marginal revenue in each segment equal to the marginal cost of the entire output. More generally stated, the discriminating monopolist continues to regard his supply (cost) situation as a single, aggregate function; but he calculates separate demand and marginal revenue curves for each of the segregated submarkets and disregards the aggregate demand curve which the sum of the submarkets would yield. He then notes the price and output in each submarket at which the marginal revenue curve in that submarket equals marginal cost of total output; and in each submarket he offers the indicated portion of output at the indicated price. Since the marginal revenue curve will lie farther below the demand curve in submarkets with less elastic demand than in those with more elastic demand, the consequence will be to offer lower quantities at higher prices in less elastic submarkets and larger quantities at lower prices in more elastic submarkets than would be appropriate if a uniform price were charged. Revenue will be increased and consumer surplus will be decreased. A simple example is shown in figure 10.

The effect on aggregate output is less certain: it depends, obviously, on whether the increased sales in more elastic markets exceed, equal or are exceeded by the decreased sales in less elastic markets. Because output is necessarily greater under perfect discrimination[3] than under single price monopoly, it might be supposed on intuitive grounds that each additional subdivision of markets approached that end position and would result in some increase in output.[4] But the assumption is false; the effect on aggregate output will depend on the comparative rates of change of elasticity in the demand curves of the several segregated markets over those price ranges involved in the shift from uniform to discriminatory pricing.[5] Output will increase if demand in the favored market is becoming more elastic faster than demand in the disfavored market is becoming inelastic; it will decrease if these conditions are reversed.

theoretical sense the objection is valid. Instantaneous entry and exit of firms from the industry in response to week-to-week changes in rates of return on capital would eliminate the discrimination, but as a practical matter they do not occur. Nevertheless the industry approximates perfect competition about as closely as any actual industry.

3. The theoretical possibility is called perfect discrimination.

4. This erroneous assumption is not uncommon and may underlie, implicitly, some of the dispute about the social desirability or undesirability of economic discrimination.

5. The classic treatment of the subject is ROBINSON, ECONOMICS OF IMPERFECT COMPETITION 188-95 (1959). That output must be larger when the point of perfect discrimination is reached is not inconsistent: at that point each unit of product is being sold in accordance with the dictates of a separate "demand curve" which is a point adjoining the price axis where the rate of change of elasticity in infinite.