



CPI COMPETITION POLICY
INTERNATIONAL

Volume 6 | Number 2 Autumn 2010

Theory of Monopoly

George Stigler

A Theory of Oligopoly

George J. Stigler*

No one has the right, and few the ability, to lure economists into reading another article on oligopoly theory without some advance indication of its alleged contribution. The present paper accepts the hypothesis that oligopolists wish to collude to maximize joint profits. It seeks to reconcile this wish with facts, such as that collusion is impossible for many firms and collusion is much more effective in some circumstances than in others. The reconciliation is found in the problem of policing a collusive agreement, which proves to be a problem in the theory of information. A considerable number of implications of the theory are discussed, and a modest amount of empirical evidence is presented.

*This article was originally published in *The Journal of Political Economy*, Vol. 72, No. 1 (Feb., 1964), pp. 44-61. It is reprinted by permission of the University of Chicago Press. In 1964, George Stigler was a Professor of Economics at the University of Chicago; he won the Nobel Memorial Prize in Economics Sciences in 1982. His original author's footnote to the article reads, "University of Chicago. I am indebted to Claire Friedland for the statistical work and to Harry Johnson for helpful criticisms."

I. The Task of Collusion

A satisfactory theory of oligopoly cannot begin with assumptions concerning the way in which each firm views its interdependence with its rivals. If we adhere to the traditional theory of profit-maximizing enterprises, then behavior is no longer something to be assumed but rather something to be deduced. The firms in an industry will behave in such a way, given the demand-and-supply functions (including those of rivals), that their profits will be maximized.

The combined profits of the entire set of firms in an industry are maximized when they act together as a monopolist. At least in the traditional formulation of the oligopoly problem, in which there are no major uncertainties as to the profit-maximizing output and price at any time, this familiar conclusion seems inescapable. Moreover, the result holds for any number of firms.

Our modification of this theory consists simply in presenting a systematic account of the factors governing the feasibility of collusion, which like most things in this world is not free. Before we do so, it is desirable to look somewhat critically at the concept of homogeneity of products, and what it implies for profit-maximizing. We shall show that collusion normally involves much more than “the” price.

Homogeneity is commonly defined in terms of identity of products or of (what is presumed to be equivalent) pairs of products between which the elasticity of substitution is infinite. On either definition it is the behavior of buyers that is decisive. Yet it should be obvious that products may be identical to any or every buyer while buyers may be quite different from the viewpoint of sellers.

This fact that every transaction involves two parties is something that economists do not easily forget. One would therefore expect a definition of homogeneity also to be two-sided: if the products are what sellers offer, and the purchase commitments are what the buyers offer, full homogeneity clearly involves infinite elasticities of substitution between both products and purchase commitments. In other words, two products are homogeneous to a buyer if he is indifferent between all combinations of x of one and (say) $20 - x$ of the other, at a common price. Two purchase commitments are homogeneous to a seller if he is indifferent between all combinations of y of one and (say) $20 - y$ of the other, at a common price. Full homogeneity is then defined as homogeneity both in products (sellers) and purchase commitments (buyers).

The heterogeneity of purchase commitments (buyers), however, is surely often at least as large as that of products within an industry, and sometimes vastly larger. There is the same sort of personal differentia of buyers as of sellers—ease in making sales, promptness of payment, penchant for returning goods, likelihood of buying again (or buying other products). In addition there are two differences among buyers which are pervasive and well recognized in economics:

1. The size of purchase, with large differences in costs of providing lots of different size.
2. The urgency of purchase, with possibly sufficient differences in elasticity of demand to invite price discrimination.

It is one thing to assert that no important market has homogeneous transactions, and quite another to measure the extent of the heterogeneity. In a regime of perfect knowledge, it would be possible to measure heterogeneity by the variance of prices in transactions; in a regime of imperfect knowledge, there will be dispersion of prices even with transaction homogeneity.¹

The relevance of heterogeneity to collusion is this: It is part of the task of maximizing industry profits to employ a price structure that takes account of the larger differences in the costs of various classes of transactions. Even with a single, physically homogeneous product the profits will be reduced if differences among buyers are ignored. A simple illustration of this fact is given in the Appendix; disregard of differences among buyers proves to be equivalent to imposing an excise tax upon them, but one which is not collected by the monopolist. A price structure of some complexity will usually be the goal of collusive oligopolists.

II. The Methods of Collusion

Collusion of firms can take many forms, of which the most comprehensive is outright merger. Often merger will be inappropriate, however, because of diseconomies of scale,² and at certain times and places it may be forbidden by law. Only less comprehensive is the cartel with a joint sales agency, which again has economic limitations—it is ill suited to custom work and creates serious administrative costs in achieving quality standards, cost reductions, product innovations, etc. In deference to American antitrust policy, we shall assume that the collusion takes the form of joint determination of outputs and prices by ostensibly independent firms, but we shall not take account of the effects of the legal prohibition until later. Oligopoly existed before 1890, and has existed in countries that have never had an antitrust policy.

The colluding firms must agree upon the price structure appropriate to the transaction classes which they are prepared to recognize. A complete profit-maximizing price structure may have almost infinitely numerous price classes: the firms will have to decide upon the number of price classes in the light of the costs and returns from tailoring prices to the diversity of transactions. We have already indicated by hypothetical example (see Appendix) that there are net profits to be obtained by catering to differences in transactions. The level of collusive prices will also depend upon the conditions of entry into the industry as well as upon the elasticities of demand.

Let us assume that the collusion has been effected, and a price structure agreed upon. It is a well-established proposition that if any member of the agreement can secretly violate it, he will gain larger profits than by conforming to it.³ It is, moreover, surely one of the axioms of human behavior that all agreements whose violation would be profitable to the violator must be enforced. The literature of collusive agreements, ranging from the pools of the 1880's to the electrical conspiracies of recent times, is replete with instances of the collapse of conspiracies because of "secret" price-cutting. This literature is biased: conspiracies that are successful in avoiding an amount of price-cutting which leads to collapse of the agreement are less likely to be reported or detected. But no conspiracy can neglect the problem of enforcement.

Enforcement consists basically of detecting significant deviations from the agreed-upon prices. Once detected, the deviations will tend to disappear because they are no longer secret and will be matched by fellow conspirators if they are not withdrawn. If the enforcement is weak, however—if price-cutting is detected only slowly and incompletely—the conspiracy must recognize its weakness: it must set prices not much above the competitive level so the inducements to price-cutting are small, or it must restrict the conspiracy to areas in which enforcement can be made efficient.

Fixing market shares is probably the most efficient of all methods of combating secret price reductions. No one can profit from price-cutting if he is moving along the industry demand curve,⁴ once a maximum profit price has been chosen. With inspection of output and an appropriate formula for redistribution of gains and losses from departures from quotas, the incentive to secret price-cutting is eliminated. Unless inspection of output is costly or ineffective (as with services), this is the ideal method of enforcement, and is widely used by legal cartels. Unfortunately for oligopolists, it is usually an easy form of collusion to detect, for it may require side payments among firms and it leaves indelible traces in the output records.

Almost as efficient a method of eliminating secret price-cutting is to assign each buyer to a single seller. If this can be done for all buyers, short-run price-cutting no longer has any purpose. Long-run price-cutting will still be a serious possibility if the buyers are in competition: lower prices to one's own customers can then lead to an expansion of their share of their market so the price-cutter's long-run demand curve will be more elastic than that of the industry. Long-run price-cutting is likely to be important, however, only where sellers are providing a major cost component to the buyer.

There are real difficulties of other sorts to the sellers in the assignment of buyers. In general the fortunes of the various sellers will differ greatly over time: one seller's customers may grow threefold, while another seller's customers shrink by half. If the customers have uncorrelated fluctuations in demand, the various sell-

ers will experience large changes in relative outputs in the short run.⁵ Where the turnover of buyers is large, the method is simply impracticable.

Nevertheless, the conditions appropriate to the assignment of customers will exist in certain industries, and in particular the geographical division of the market has often been employed. Since an allocation of buyers is an obvious and easily detectable violation of the Sherman Act, we may again infer that an efficient method of enforcing a price agreement is excluded by the antitrust laws. We therefore turn to other techniques of enforcement, but we shall find that the analysis returns to allocation of buyers.

In general the policing of a price agreement involves an audit of the transactions prices [sic]. In the absence or violation of antitrust laws, actual inspection of the accounting records of sellers has been employed by some colluding groups, but even this inspection gives only limited assurance that the price agreement is adhered to.⁶ Ultimately, there is no substitute for obtaining the transaction prices from the buyers.

An oligopolist will not consider making secret price cuts to buyers whose purchases fall below a certain size relative to his aggregate sales. The ease with which price-cutting is detected by rivals is decisive in this case. If p is the probability that some rival will hear of one such price reduction, $1 - (1 - p)^n$ is the probability that a rival will learn of at least one reduction if it is given to n customers. Even if p is as small as 0.01, when n equals 100 the probability of detection is .634, and when n equals 1000 it is .99996. No one has yet invented a way to advertise price reductions which brings them to the attention of numerous customers but not to that of any rival.⁷

It follows that oligopolistic collusion will often be effective against small buyers even when it is ineffective against large buyers. When the oligopolists sell to numerous small retailers, for example, they will adhere to the agreed-upon price, even though they are cutting prices to larger chain stores and industrial buyers. This is a first empirical implication of our theory. Let us henceforth exclude small buyers from consideration.

The detection of secret price-cutting will of course be as difficult as interested people can make it. The price-cutter will certainly protest his innocence, or, if this would tax credulity beyond its taxable capacity, blame a disobedient subordinate. The price cut will often take the indirect form of modifying some non-price dimension of the transaction. The customer may, and often will, divulge price reductions, in order to have them matched by others, but he will learn from experience if each disclosure is followed by the withdrawal of the lower price offer. Indeed the buyer will frequently fabricate wholly fictitious price offers to test the rivals. Policing the collusion sounds very much like the subtle and complex problem presented in a good detective story.

There is a difference: In our case the man who murders the collusive price will receive the bequest of patronage. The basic method of detection of a price-cutter must be the fact that he is getting business he would otherwise not obtain. No promises of lower prices that fail to shift some business can be really effective—either the promised price is still too high or it is simply not believed.

Our definition of perfect collusion, indeed, must be that no buyer changes sellers voluntarily. There is no competitive price-cutting if there are no shifts of buyers among sellers.

To this rule that price-cutting must be inferred from shifts of buyers there is one partial exception, but that an important one. There is one type of buyer who usually reveals the price he pays, and does not accept secret benefices: the government. The system of sealed bids, publicly opened with full identification of each bidder's price and specifications, is the ideal instrument for the detection of price-cutting. There exists no alternative method of secretly cutting prices (bribery of purchasing agents aside). Our second empirical prediction, then, is that collusion will always be more effective against buyers who report correctly and fully the prices tendered to them.⁸

It follows from the test of the absence of price competition by buyer loyalty—and this is our third major empirical prediction—that collusion is severely limited (under present assumptions excluding market-sharing) when the significant buyers constantly change identity. There exist important markets in which the (substantial) buyers do change identity continuously, namely, in the construction industries. The building of a plant or an office building, for example, is an essentially non-repetitive event, and rivals cannot determine whether the successful bidder has been a price-cutter unless there is open bidding to specification.

The normal market, however, contains both stability and change. There may be a small rate of entry of new buyers. There will be some shifting of customers even in a regime of effective collusion, for a variety of minor reasons we can lump together as “random factors.” There will often be some sharing of buyers by several sellers—a device commending itself to buyers to increase the difficulty of policing price agreements. We move then to the world of circumstantial evidence, or, as it is sometimes called, of probability.

III. The Conditions for Detecting Secret Price Reductions

We shall investigate the problem of detecting secret price-cutting with a simplified model, in which all buyers and all sellers are initially of equal size. The number of buyers per seller—recalling that we exclude from consideration all buyers who take less than (say) 0.33 per cent of a seller's output—will range from 300 down to perhaps 10 or 20 (since we wish to avoid the horrors of full bilateral oli-

gopoly). A few of these buyers are new, but over moderate periods of time most are “old,” although some of these old customers will shift among suppliers. A potential secret price-cutter has then three groups of customers who would increase their patronage if given secret price cuts: the old customers of rivals; the old customers who would normally leave him; and new customers.

Most old buyers will deal regularly with one or a few sellers, in the absence of secret price-cutting. There may be no secret price-cutting because a collusive price is adhered to, or because only an essentially competitive price can be obtained. We shall show that the loyalty of customers is a crucial variable in determining which price is approached. We need to know the probability that an old customer will buy again from his regular supplier at the collusive price in the absence of secret price-cutting.

The buyer will set the economies of repetitive purchase (which include smaller transaction costs and less product-testing) against the increased probability of secret price-cutting that comes from shifting among suppliers. From the viewpoint of any one buyer, this gain will be larger the larger the number of sellers and the smaller the number of buyers, as we shall show below. The costs of shifting among suppliers will be smaller the more homogenous the goods and the larger the purchases of the buyer (again an inverse function of his size). Let us label this probability of repeat purchases p . We shall indicate later how this probability could be determined in a more general approach.

The second component of sales of a firm will be its sales to new buyers and to the floating old customers of rivals. Here we assume that each seller is equally likely to make a sale, in the absence of price competition.

Let us proceed to the analysis. There are n_o “old” buyers and n_n new customers, with $n_n = \lambda n_o$ and n_s sellers. A firm may look to three kinds of evidence on secret price-cutting, and therefore by symmetry to three potential areas to practice secret price-cutting.

1. *The behavior of its own old customers.*—It has, on average, n_o/n_s such customers, and expects to sell to $m_1 = pn_o/n_s$ of them in a given round of transactions, in the absence of price cutting. The variance of this number of customers is $\sigma_1^2 = \frac{(1-p)pn_o}{n_s}$. The probability of the firm losing more old customers than $\frac{(1-p)pn_o}{n_s} + k\sigma_1$ is given by the probability of values greater than k . The expected number of these old customers who will shift to any one rival is, say, $m_2 = \frac{1}{n_s - 1} \left[\frac{(1-p)n_o}{n_s} + k\sigma_1 \right]$, with a variance $\sigma_2^2 = \frac{n_s - 2}{(n_s - 1)^2} \left[\frac{(1-p)n_o}{n_s} + k\sigma_1 \right]$. The probability that any rival will obtain more than $m_2 + r\sigma_2$ of these customers is determined by r . We could now choose those combinations of k and r that fix a level of probability for the loss of a given number of old customers to any one

rival beyond which secret price-cutting by this rival will be inferred. This is heavy arithmetic, however, so we proceed along a less elegant route.

Let us assume that the firm's critical value for the loss of old customers, beyond which it infers secret price-cutting, is

$$\frac{(1-p)n_o}{n_s} + \sigma_1 = \frac{(1-p)n_o}{n_s} \left[1 + \sqrt{\left(\frac{p}{1-p} \frac{n_s}{n_o} \right)} \right] = \frac{(1-p)n_o}{n_s} (1 + \theta),$$

that is, one standard deviation above the mean. Any one rival will on average attract $m_2 = \frac{1}{n_s - 1} \left[\frac{(1-p)n_o}{n_s} + \sigma_1 \right]$ of these customers, with a variance of $\sigma_2^2 = \frac{n_s - 2}{(n_s - 1)^2} \left[\frac{(1-p)n_o}{n_s} + \sigma_1 \right]$. Let the rival be suspected of price-cutting if he obtains more than $(m_2 + \sigma_2)$ customers, that is, if the probability of any larger number is less than about 30 per cent. The joint probability of losing one standard deviation more than the average number of old customers and a rival obtaining one standard deviation more than his average share is about 10 per cent. The average sales of a rival are n_o/n_s , ignoring new customers. The maximum number of buyers any seller can obtain from one rival without exciting suspicion, minus the number he will on average get without price-cutting $([1-p]n_o/n_s[n_s - 1])$, expressed as a ratio to his average sales is $\frac{[\theta(1-p)n_o/(n_s - 1)n_s + \sigma_2]}{n_o/n_s}$. This criterion is tabulated in Table 1.

Table 1

Percentage Gains
in Sales from
Undetected
Price-Cutting
by a Firm

Criterion I: $\frac{1}{(n_s - 1)} \left[\theta(1-p) + \sqrt{\frac{n_o(n_s - 2)(1-p)(1 + \theta)}{n_o}} \right] \theta = \sqrt{\frac{p}{1-p} \frac{n_s}{n_o}}$

Probability of Repeat Sales (<i>p</i>)	No. of Buyers (<i>n_o</i>)	No. of Sellers					
		2	3	4	5	10	20
<i>p</i> = 0.95	20	6.9	11.3	11.3	11.4	11.8	12.7
	30	5.6	8.9	8.8	8.8	9.0	9.6
	40	4.9	7.5	7.4	7.4	7.5	7.9
	50	4.4	6.6	6.5	6.4	6.5	6.8
	100	3.1	4.4	4.3	4.3	4.2	4.4
	200	2.2	3.0	2.9	2.8	2.8	2.8
	400	1.5	2.1	2.0	1.9	1.8	1.8
<i>p</i> = 0.90	20	9.5	14.8	14.7	14.6	14.8	15.7
	30	7.8	11.7	11.5	11.4	11.4	12.0
	40	6.7	10.0	9.7	9.6	9.5	9.9
	50	6.0	8.8	8.6	8.4	8.3	8.6
	100	4.2	6.0	5.8	5.6	5.4	5.5
	200	3.0	4.1	3.9	3.8	3.6	3.6
	400	2.1	2.8	2.7	2.6	2.4	2.4

Continued on next page

Table 1,
continued

Probability of Repeat Sales (p)	No. of Buyers (n_0)	No. of Sellers					
		2	3	4	5	10	20
$p = 0.80$	20	12.6	19.3	18.9	18.7	18.6	19.4
	30	10.3	15.4	15.0	14.7	14.5	15.0
	40	8.9	13.1	12.7	12.5	12.2	12.5
	50	8.0	11.6	11.2	11.0	10.6	10.8
	100	5.7	8.0	7.7	7.4	7.1	7.1
	200	4.0	5.5	5.3	5.1	4.8	4.7
	400	2.8	3.8	3.6	3.5	3.2	3.2
$p = 0.70$	20	14.5	22.3	21.8	21.5	21.2	21.9
	30	11.8	17.8	17.3	17.0	16.6	16.9
	40	10.2	15.2	14.8	14.5	14.0	14.2
	50	9.2	13.5	13.1	12.8	12.3	12.4
	100	6.5	9.3	9.0	8.7	8.2	8.2
	200	4.6	6.5	6.2	6.0	5.6	5.5
	400	3.2	4.5	4.3	4.2	3.8	3.7

The entries in Table 1 are measures of the maximum additional sales obtainable by secret price-cutting (expressed as a percentage of average sales) from any one rival beyond which that rival will infer that the price-cutting is taking place. Since the profitability of secret price-cutting depends upon the amount of business one can obtain (as well as upon the excess of price over marginal cost), we may also view these numbers as the measures of the incentive to engage in secret price-cutting. Three features of the tabulation are noteworthy:

- a) The gain in sales from any one rival by secret price-cutting is not very sensitive to the number of rivals, given the number of customers and the probability of repeat sales. The aggregate gain in sales of a firm from price-cutting—its total incentive to secret price-cutting—is the sum of the gains from each rival, and therefore increases roughly in proportion to the number of rivals.
- b) The incentive to secret price-cutting falls as the number of customers per seller increases—and falls roughly in inverse proportion to the square root of the number of buyers.
- c) The incentive to secret price-cutting rises as the probability of repeat purchases falls, but at a decreasing rate.

We have said that the gain to old buyers from shifting their patronage among sellers will be that it encourages secret price-cutting by making it more difficult to detect. Table 1 indicates that there are diminishing returns to increased shifting: The entries increase at a decreasing rate as p falls. In a fuller model we could intro-

duce the costs of shifting among suppliers and determine p to maximize expected buyer gains. The larger the purchases of a buyer, when buyers are of unequal size, however, the greater is the prospect that his shifts will induce price-cutting.

In addition it is clear that, when the number of sellers exceeds two, it is possible for two or more firms to pool information and thus to detect less extreme cases of price-cutting. For example, at the given probability levels, the number of old customers that any one rival should be able to take from a firm was shown to be at most $(1-p) \frac{n_o(1+\theta)}{n_s-1}$, with variance $\frac{(n_s-2)(1-p)(1+\theta)}{(n_s-1)^2} n_o$. At the same probability level, the average number of old customers that one rival should be able to take from T firms is at most $\frac{T(1-p)n_o}{n_s-T} \left(1 + \frac{\theta}{\sqrt{T}}\right)$, with the variance $\frac{(n_s-T-1)}{(n_s-T)^2} (1-p) \left(1 + \frac{\theta}{\sqrt{T}}\right) n_o T$. Each of these is smaller than the corresponding expression for one seller when expressed as a fraction of the customers lost by each of the firms pooling information.

There are of course limits to such pooling of information: not only does it become expensive as the number of firms increases, but also it produces less reliable information, since one of the members of the pool may himself be secretly cutting prices. Some numbers illustrative of the effect of pooling will be given at a later point.

2. *The attraction of old customers of other firms is a second source of evidence of price-cutting.*—If a given rival has not cut prices, he will on average lose $(1-p)$ (n_o/n_s) customers, with a variance of σ_1^2 . The number of customers he will retain with secret price-cutting cannot exceed a level at which the rivals suspect the price-cutting. Any one rival will have little basis for judging whether he is getting a fair share of this firm's old customers, but they can pool their information and then in the aggregate they will expect the firm to lose at least $(1-p)(n_o/n_s) - 2\sigma_1$ customers, at the 5 per cent probability level. Hence the secret price-cutter can retain at most $2\sigma_1$ of his old customers (beyond his average number), which as a fraction of his average sales (ignoring new customers) is $\frac{2\sigma_1}{n_o/n_s} = 2\sqrt{\frac{(1-p)pn_s}{n_o}}$. This is tabulated as Table 2.

Table 2

Old Customers
That a Secret
Price-Cutter Can
Retain, as a
Percentage of
Average Sales

Criterion II: $2\sqrt{\frac{p(1-p)}{2} \frac{n_s}{n_o}}$

Probability That Old Customer Will Remain Loyal (p)	No. of Old Customers per Seller (n_o/n_s)			
	10	20	50	100
0.95	13.8	9.7	6.2	4.4
.90	19.0	13.4	8.5	6.0

Continued on next page

Table 2,
continued

Probability That Old Customer Will Remain Loyal (p)	No. of Old Customers per Seller (n_o/n_s)			
	10	20	50	100
0.85	22.6	16.0	10.1	7.1
.80	25.3	17.9	11.3	8.0
.75	27.4	19.4	12.2	8.7
.70	29.0	20.5	13.0	9.2
.65	30.2	21.3	13.5	9.5
.60	31.0	21.9	13.9	9.8
.55	31.5	22.2	14.1	10.0
0.50	31.6	22.4	14.1	10.0

If the entries in Table 2 are compared with those in Table 1,⁹ it is found that a price-cutter is easier to detect by his gains at the expense of any one rival than by his unusual proportion of repeat sales. This second criterion will therefore seldom be useful.

3. *The behavior of new customers is a third source of information on price-cutting.*— There are n_n new customers per period,¹⁰ equal to λn_o . A firm expects, in the absence of price-cutting, to sell to $m_3 = \frac{1}{n_s} \lambda n_o$ of these customers, with a variance of $\sigma_3^2 = \left(1 - \frac{1}{n_s}\right) \frac{\lambda n_o}{n_s}$. If the rivals pool information (without pooling, this area could not be policed effectively), this firm cannot obtain more than $m_3 + 2\sigma_3$ customers without being deemed a price-cutter, using again a 5 per cent probability criterion. As a percentage of the firm's total sales, the maximum sales above the expected number in the absence of price cutting[sic] are then $\frac{2\sigma_3}{n_o(1 + \lambda)/n_s} = \frac{2}{1 + \lambda} \sqrt{\frac{\lambda(n_s - 1)}{n_o}}$. We tabulate this criterion as Table 3.

Table 3

Maximum Additional
New Customers
(as a Percentage
of Average Sales)
Obtainable by
Secret Price-Cutting

Rate of Appearance of New Buyers (λ)	No. of Old Buyers (n_o)	No. of Sellers					
		2	3	4	5	10	20
1/100	20	4.4	6.3	7.7	8.9	13.3	19.3
	30	3.6	5.1	6.3	7.2	10.8	15.8

Continued on next page

Table 3,
Continued

Rate of Appearance of New Buyers (λ)	No. of Old Buyers (n_0)	No. of Sellers					
		2	3	4	5	10	20
	40	3.1	4.4	5.4	6.3	9.4	13.6
	50	2.8	4.0	4.8	5.6	8.4	12.2
	100	2.0	2.8	3.4	4.0	5.9	8.6
	200	1.4	2.0	2.4	2.8	4.2	6.1
	400	1.0	1.4	1.7	2.0	3.0	4.3
1/10	20	12.9	18.2	22.3	25.7	38.6	56.0
	30	10.5	14.8	18.2	21.0	31.5	45.8
	40	9.1	12.9	15.8	18.2	27.3	39.6
	50	8.1	11.5	14.1	16.3	24.4	35.4
	100	5.8	8.1	10.0	11.5	17.2	25.1
	200	4.1	5.8	7.0	8.1	12.2	17.7
	400	2.9	4.1	5.0	5.8	8.6	12.5
1/5	20	16.7	23.6	28.9	33.3	50.0	72.6
	30	13.6	19.2	23.6	27.2	40.8	59.3
	40	11.8	16.7	20.4	23.6	35.4	51.4
	50	10.5	14.9	18.3	21.1	31.6	46.0
	100	7.4	10.5	12.9	14.9	22.4	32.5
	200	5.3	7.4	9.1	10.5	15.8	23.0
	400	3.7	5.3	6.4	7.4	11.2	16.2
1/4	20	17.9	25.3	31.0	35.8	53.7	78.0
	30	14.6	20.7	25.3	29.2	43.8	63.7
	40	12.6	17.9	21.9	25.3	38.0	55.1
	50	11.3	16.0	19.6	22.6	33.9	49.3
	100	8.0	11.3	13.9	16.0	24.0	34.9
	200	5.7	8.0	9.8	11.3	17.0	24.7
	400	4.0	5.7	6.9	8.0	12.0	17.4

Two aspects of the incentive to cut prices (or equivalently the difficulty of detecting price cuts) to new customers are apparent: the incentive increases rapidly with the number of sellers¹¹ and the incentive increases with the rate of entry of new customers. As usual the incentive falls as the absolute number of customers per seller rises. If the rate of entry of new buyers is 10 per cent or more, price-cutting to new customers allows larger sales increases without detection that can be obtained by attracting customers of rivals (compare Tables 1 and 3).

Of the considerable number of directions in which this model could be enlarged, two will be presented briefly.

The first is inequality in the size of firms. In effect this complication has already been introduced by the equivalent device of pooling information. If we tabulate the effects of pooling of information by K firms, the results are equivalent to having a firm K times as large as the other firms. The number of old customers this large firm can lose to any one small rival (all of whom are equal in

size) is given, in Table 4, as a percentage of the average number of old customers of the small firm; the column labeled $K = 1$ is of course the case analyzed in Table 1.

Table 4

Percentage Gains in Sales from Undetected Price-Cutting by a Small Firm

$$\text{Criterion IV: } \frac{1}{n_s - K} \left[\theta(1-p)\sqrt{K} + \sqrt{\frac{n_s K(1-p)(n_s - K - 1)(1 + \theta\sqrt{K})}{n_o}} \right]$$

$$\theta = \sqrt{\frac{p}{1-p} \frac{n_s}{n_o}}$$

Probability of Repeat Sales (p)	No. of Firms ($n_s - K + 1$)	Buyers per Small Seller (n_o/n_s)	Size of Large Firm (K)			
			1	2	5	9
$p = 0.9$	2	10	9.5	13.4	21.2	28.5
		30	5.5	7.7	12.2	16.4
		50	4.2	6.0	9.5	12.7
	3	10	11.7	15.8	23.9	31.4
		30	6.3	8.7	13.3	17.6
		50	4.8	6.6	10.2	13.5
	4	10	9.7	13.1	19.7	25.7
		30	5.2	7.1	10.9	14.4
		50	4.0	5.4	8.3	11.0
	10	10	5.4	7.2	10.7	14.0
		30	2.9	3.9	5.9	7.7
		50	2.2	2.9	4.5	5.9
$p = 0.8$	2	10	12.6	17.9	28.3	37.9
		30	7.3	10.3	16.3	21.9
		50	5.7	8.0	12.6	17.0
	3	10	15.4	21.0	32.1	42.3
		30	8.4	11.6	18.0	23.9
		50	6.4	8.9	13.8	18.4
	4	10	12.7	17.3	26.3	34.7
		30	6.9	9.5	14.7	19.5
		50	5.3	7.3	11.3	15.0
	10	10	7.1	9.5	14.4	18.9
		30	3.8	5.2	8.0	10.6
		50	2.9	4.0	6.1	8.1

The effects of pooling on the detection of price-cutting are best analyzed by comparing Table 4 with Table 1. If there are 100 customers and 10 firms (and $p = 0.9$), a single firm can increase sales by 5.4 per cent by poaching on one rival, or about 50 per cent against all rivals (Table 1). If 9 firms combine, the maximum amount the single firm can gain by secret price-cutting is 28.9 per cent (Table 4). With 20 firms and 200 customers, a single firm can gain 3.6 per cent

from each rival, or about 30 per cent from 9 rivals; if these rivals merge, the corresponding figure falls to 14.0 per cent. The pooling of information therefore reduces substantially the scope for secret price-cutting.

This table exaggerates the effect of inequality of firm size because it fails to take account of the fact that the number of customers varies with firm size, on our argument that only customers above a certain size relative to the seller are a feasible group for secret price-cutting. The small firm can find it attractive to cut prices to buyers which are not large enough to be potential customers by price-cutting for the large seller.

The temporal pattern of buyers' behavior provides another kind of information: What is possibly due to random fluctuation in the short run cannot with equal probability be due to chance if repeated. Thus the maximum expected loss of old customers to a rival in one round of transactions is (at the 1σ level)

$\frac{n_o}{(n_s - 1)n_s}(1 - p)(1 + \theta)$, but for T consecutive periods the maximum expected loss is (over T periods) $\frac{T}{n_s - 1}(1 - p) \frac{n_o}{n_s} [1 + \theta\sqrt{T}]$, with a variance of $\sigma_3^2 = \frac{(n_s - 2)}{(n_s - 1)^2} T(1 - p) \frac{n_o}{n_s} [1 + \theta\sqrt{T}]$. This source of information is of minor efficacy in detecting price-cutting unless the rounds of successive transactions are numerous—that is, unless buyers purchase (enter contracts) frequently.

Our approach has certain implications for the measurement of concentration, if we wish concentration to measure likelihood of effective collusion. In the case of new customers, for example, let the probability of attracting a customer be proportional to the firm's share of industry output (s). Then the variance of the firm's share of sales to new customers will be $n_n s(1 - s)$, and the aggregate for the industry will be $C = n_n \sum_1^r s(1 - s)$ for r firms. This expression equals $n_n(1 - H)$, where $H = \sum s^2$ is the Herfindahl index of concentration. The same index holds, as an approximation, for potential price-cutting to attract old customers.¹²

The foregoing analysis can be extended to non-price variables, subject to two modifications. The first modification is that there be a definite joint profit-maximizing policy upon which the rivals can agree. Here we may expect to encounter a spectrum of possibilities, ranging from a clearly defined optimum policy (say, on favorable legislation) to a nebulous set of alternatives (say, directions of research).¹³ Collusion is less feasible, the less clear the basis on which it should proceed. The second modification is that the competitive moves of any one firm will differ widely among non-price variables in their detectability by rivals. Some forms of non-price competition will be easier to detect than price-cutting because they leave visible traces (advertising, product quality, servicing, etc.) but some variants will be elusive (reciprocity in purchasing, patent licensing arrangements). The common belief that non-price competition is more common than price competition is therefore not wholly in keeping with the present theory.

Those forms that are suitable areas for collusion will have less competition; those which are not suitable will have more competition.

IV. Some Fragments of Evidence

Before we seek empirical evidence on our theory, it is useful to report two investigations of the influence of numbers of sellers on price. These investigations have an intrinsic interest because, so far as I know, no systematic analysis of the effect of numbers has hitherto been made.

The first investigation was of newspaper advertising rates, as a function of the number of evening newspapers in a city. Advertising rates on a milline basis are closely (and negatively) related to circulation, so a regression of rates on circulation was made for fifty-three cities in 1939. The residuals (in logarithmic form) from this regression equation are tabulated in Table 5.

Table 5
Residuals from Regression of Advertising Rates on Circulation*

No. of Evening Papers	n	Mean Residual (Logarithm)	Standard Deviation of Mean
One	23	0.0211	0.0210
With morning paper	10	-.0174	.0324
Without morning paper	13	.0507	.0233
Two	30	-0.0213	0.0135

*The regression equation is

$$\log R = 5.194 - 1.688 \log c + .139(\log c)^2,$$

(.620) (.063)

where R is the 5 M milline rate and c is circulation.

Source: American Association of Advertising Agencies, *Market and Newspaper Statistics*, Vol. VIIIa (1939).

It will be observed that rates are 5 per cent above the average in one-newspaper towns and 5 per cent below the average in two-newspaper towns, and the towns with one evening paper but also an independent morning paper fall nearly midway between these points. Unfortunately there were too few cities with more than two evening newspapers to yield results for larger numbers of firms.

The second investigation is of spot commercial rates on AM radio stations in the four states of Ohio, Indiana, Michigan, and Illinois. The basic equation introduces, along with number of rivals, a series of other factors (power of station, population of the county in which the station is located, etc.). Unfortunately the number of stations is rather closely correlated with population ($r^2 = .796$ in the logarithms). The general result, shown in Table 6, is similar to that for newspapers: the elasticity of price with respect to number of rivals is quite small (-.07). Here the range of stations in a county was from 1 to 13.

Table 6

Regression of AM
Spot Commercial
Rates (26 Times)
and Station
Characteristics,
1961 ($n = 345$)

Independent Variables*	Regression Coefficient	Standard Error
1. Logarithm of population of county, 1960	.238	0.026
2. Logarithm of kilowatt power of station	.206	.015
3. Dummy variables of period of broadcasting:		
a) Sunrise to sunset	-.114	.025
b) More than (a), less than 18 hours	-.086	.027
c) 18-21 hours	-.053	.028
4. Logarithm of number of stations in county	-.074	0.046
	$R^2 = .743$	

*Dependent variable: logarithm of average rate, May 1, 1961 (dollars).

Source: "Spot Radio Rates and Data," *Standard Rate and Data Service, Inc.*, Vol. XLIII, No. 5 (May 1961).

Both studies suggest that the level of prices is not very responsive to the actual number of rivals. This is in keeping with the expectations based upon our model, for that model argues that the number of buyers, the proportion of new buyers, and the relative sizes of firms are as important as the number of rivals.

To turn to the present theory, the only test covering numerous industries so far devised has been one based upon profitability. This necessarily rests upon company data, and it has led to the exclusion of a large number of industries for which the companies do not operate in a well-defined industry. For example, the larger steel and chemical firms operate in a series of markets in which their position ranges from monopolistic to competitive. We have required of each industry that the earnings of a substantial fraction of the companies in the industry (measured by output) be determined by the profitability of that industry's products, that is, that we have a fair share of the industry and the industry's product is the dominant product of the firms.

Three measures of profitability are given in Table 7: (1) the rate of return on all capital (including debt), [sic](2) the rate of return on net worth (stockholders' equity); (3) the ratio of market value to book value of the common stock.

In addition, two measures of concentration are presented: (1) the conventional measure, the share of output produced by the four leading firms; and (2) the Herfindahl index, H .

Table 7

Profitability and Concentration Data

Industry*	Concentration (1954)		Average Rate of Return (1953–57)		Ratio of Market Value to Book Value (1953–57)
	Share of Top 4	H [†]	All Assets	Net Worth	
Sulfur mining (4)	98	0.407	19.03	23.85	3.02
Automobiles (3)	98	.369	11.71	20.26	2.30
Flat glass (3)	90	.296	11.79	16.17	2.22
Gypsum products (2)	90	.280	12.16	20.26	1.83
Primary aluminum (4)	98	.277	6.87	13.46	2.48
Metal cans (4)	80	.260	7.27	13.90	1.60
Chewing gum (2)	86	.254	13.50	17.06	2.46
Hard-surface floor coverings (3)	87	.233	6.56	7.59	0.98
Cigarettes (5)	83	.213	7.23	11.18	1.29
Industrial gases (3)	84	.202	8.25	11.53	1.33
Corn wet milling (3)	75	.201	9.17	11.55	1.48
Typewriters (3)	83	.198	3.55	5.39	0.84
Domestic laundry equipment (2)	68	.174	9.97	17.76	1.66
Rubber tires (9)	79	.171	7.86	14.02	1.70
Rayon fiber (4)	76	.169	5.64	6.62	0.84
Carbon black (2)	73	.152	8.29	9.97	1.40
Distilled liquors (6)	64	0.118	6.94	7.55	0.77

*The number of firms is given in parentheses after the industry title. Only those industries are included for which a substantial share (35 per cent or more) of the industry's sales is accounted for by the firms in the sample, and these firms derive their chief revenues (50 per cent or more) from the industry in question.

[†]H is Herfindahl index.

The various rank correlation measures are given in Table 8. The various concentration measures, on the one hand, and the various measures of profitability, on the other hand, are tolerably well correlated.¹⁴ All show the expected positive relationship. In general the data suggest that there is no relationship between profitability and concentration if *H* is less than 0.250 or the share of the four largest firms is less than about 80 per cent. These data, like those on advertising rates, confirm our theory only in the sense that they support theories which assert that competition increases with number of firms.

Table 8

Rank Correlations of Measures of Profitability and Measures of Concentration

Measure of Concentration	Measure of Profitability		
	Rate of Return on All Assets	Rate of Return on Net Worth	Ratio of Market Value to Book Value
Share of output produced by four largest firms	.322	.507	.642
Herfindahl index (<i>H</i>)	.524	.692	.730

Our last evidence is a study of the prices paid by buyers of steel products in 1939, measured relative to the quoted prices (Table 9). The figure of 8.3 for hot-rolled sheets for example, represents an average of 8.3 per cent reduction from quoted prices, *paid by buyers*, with a standard deviation of 7.3 per cent of quoted prices. The rate of price-cutting is almost perfectly correlated with the standard deviation of transaction prices, as we should expect: the less perfect the market knowledge, the more extensive the price-cutting.

Table 9

Prices of Steel Products, 1939, and Industry Structure, 1938

Product Class	Prices, 2d Quarter, 1939 (Per Cent)		Herfindahl Index	Output in 1939 Relative to 1937
	Average Discount from List Price	Standard Deviation		
Hot-rolled sheets	8.3	7.3	0.0902	1.14
Merchant bars	1.2	4.5	.1517	0.84
Hot-rolled strip	8.5	8.3	.1069	0.56
Plates	2.6	4.8	.1740	0.85
Structural shapes	3.2	4.3	.3280	0.92
Cold-rolled strip	8.8	9.8	.0549	0.88
Cold-rolled sheets	5.8	5.0	.0963	1.14
Cold-finished bars	0.9	3.4	0.0964	0.83

Source: "Labor Department Examines Consumers' Prices of Steel Products," *Iron Age*, April 25, 1946; industry structure: 1938 capacity data from *Directory of Iron and Steel Works of the United States and Canada*; output: *Annual Statistical Report, American Iron and Steel Institute* (New York, 1938, 1942).

In general, the more concentrated the industry structure (measured by the Herfindahl index), the larger were the price reductions. Although there were no extreme departures from this relationship, structural shapes and hot-rolled strip had prices somewhat lower than the average relationship, and cold finished bars prices somewhat higher than expected, and the deviations are not accounted for by the level of demand (measured by 1939 sales relative to 1937 sales). The number of buyers could be taken into account, but the BLS study states:

"The extent of price concessions shown by this study is probably understated because certain very large consumers in the automobile and container industries were excluded from the survey. This omission was at the request of the OPA which contemplated obtaining this information in connection with other studies. Since a small percentage of steel consumers, including

these companies, accounts for a large percentage of steel purchased, prices paid by a relatively few large consumers have an important influence upon the entire steel price structure. Very large steel consumers get greater reductions from published prices than smaller consumers, often the result of competitive bidding by the mills for the large volume of steel involved. One very large steel consumer, a firm that purchased over 2 pct. of the total consumption of hot and cold-rolled sheets in 1940, refused to give purchase prices. This firm wished to protect its suppliers, fearing that "certain transactions might be revealed which would break confidence" with the steel mills. However, this company did furnish percent changes of prices paid for several steel products which showed that for some products prices advanced markedly, and in one case, nearly 50 pct. The great price advances for this company indicate that it was receiving much larger concessions than smaller buyers."¹⁵

These various bits of evidence are fairly favorable to the theory, but they do not constitute strong support. More powerful tests will be feasible when the electrical equipment triple-damage suits are tried.¹⁶ The great merit of our theory, in fact, is that it has numerous testable hypotheses, unlike the immortal theories that have been traditional in this area.

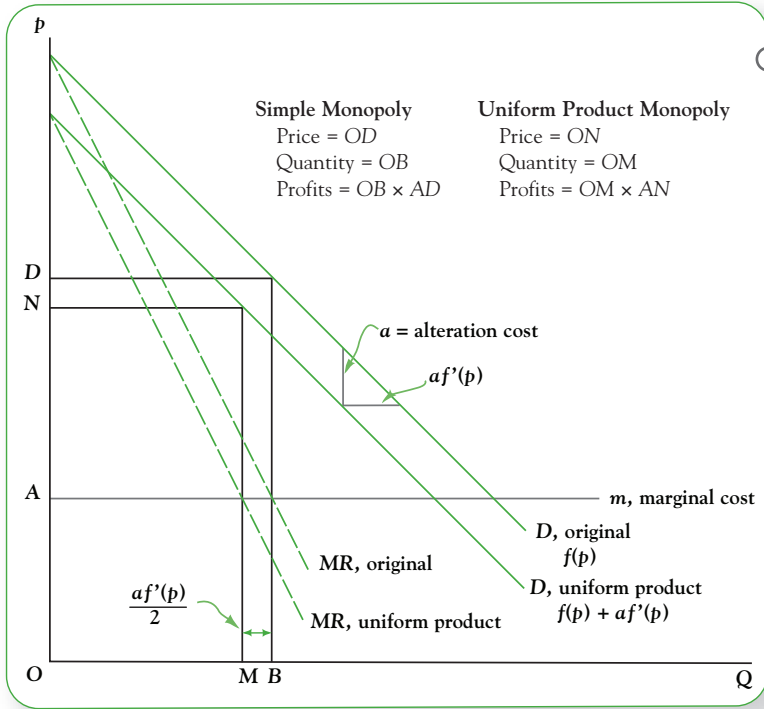
Appendix

The importance of product heterogeneity for profit-maximizing behavior cannot well be established by an a priori argument. Nevertheless, the following simple exposition of the implications for profitability of disregarding heterogeneity may have some heuristic value. The analysis, it will be observed, is formally equivalent to that of the effects of an excise tax on a monopolist.

Assume that a monopolist makes men's suits, and that he makes only one size of suit. This is absurd behavior but the picture of the sadistic monopolist who disregards consumer desires has often made fugitive appearances in the literature so that problem has some interest of its own. The demand curve of a consumer for suits that fit, $f(p)$, would now be reduced because he would have to incur some alteration cost a in order to wear the suit. His effective demand would therefore decline to $f(p + a)$. Assume further that the marginal cost of suits is constant (m), and that it would be the same if the monopolist were to make suits of various sizes.

The effect on profits of a uniform product—uniform is an especially appropriate word here—can be shown graphically.

Figure



The decrease in quantity sold, with a linear demand curve, is $MB = 1/2af'(p)$. The decrease in the price received by the monopolist is $DN = \frac{MB}{f'(p)} - a = -\frac{a}{2}$, so if π is profit per unit, and q is output, the relative decline in total profit is approximately $\frac{\Delta\pi}{\pi} + \frac{\Delta q}{q}$, or $\frac{MB}{OB} + \frac{ND}{AD}$. Since $OB = \frac{f(m)}{2}$ $AD = \frac{p}{\eta}$, where η is the elasticity of demand, the relative decline of profits with a uniform product is $\frac{af'(p)}{f(m)} + \frac{a\eta}{2p} = \frac{a\eta}{2p} + \frac{a\eta}{2p} = \frac{a\eta}{p}$.

The loss from imposed uniformity is therefore proportional to the ratio of alteration costs to price.

Our example is sufficiently unrealistic to make any quantitative estimate uninteresting. In general one would expect an upper limit to the ratio a/p , because it becomes cheaper to resort to other goods (customer tailoring in our example), or to abandon the attempt to find appropriate goods. The loss of profits of the monopolist will be proportional to the average value of a/p , and this will be smaller, the smaller the variation in buyers' circumstances.

Still, monopolists are lucky if their long-run demand curves have an elasticity only as large as -5 , and then even a ratio of a to p if $1/40$ will reduce their prof-

its by 12 per cent. The general conclusion I wish to draw is that a monopolist who does not cater to the diversities of his buyers' desires will suffer a substantial decline in his profits. ▼

-
- 1 Unless one defines heterogeneity of transactions to include also differences in luck in finding low price sellers; see my "Economics of Information," *Journal of Political Economy*, June, 1961.
 - 2 If the firms are multiproduct, with different product structures, the diseconomies of merger are not strictly those of scale (in any output) but of firm size measured either absolutely or in terms of variety of products.
 - 3 If price is above marginal cost, marginal revenue will be only slightly less than price (and hence above marginal cost) for price cuts by this one seller.
 - 4 More precisely, he is moving along a demand curve which is a fixed share of the industry demand, and hence has the same elasticity as the industry curve at every price.
 - 5 When the relative outputs of the firms change, the minimum cost condition of equal marginal costs for all sellers is likely to be violated. Hence industry profits are not maximized.
 - 6 The literature and cases on "open-price associations" contain numerous references to the collection of prices from sellers (see Federal Trade Commission, *Open-Price Trade Associations* [Washington, 1929], and cases cited).
 - 7 This argument applies to size of buyer relative to the individual seller. One can also explain the absence of higgling [sic] in small transactions because of the costs of bargaining, but this latter argument turns on the absolute size of the typical transaction, not its size relative to the buyer.
 - 8 The problem implicitly raised by these remarks is why all sales to the government are not at collusive prices. Part of the answer is that the government is usually not a sufficiently large buyer of a commodity to remunerate the costs of collection.
 - 9 For example, take $p = .95$. The entry for 10 customers per seller is 13.8 in Table 2—this is the maximum percentage of average sales that can be obtained by price reductions to old customers. The corresponding entries in Table 1 are 6.9 (2 sellers, 20 buyers), 8.9 (3 and 30), 7.4 (4 and 40), 6.4 (5 and 50), 4.2 (10 and 100), etc. Multiplying each entry in Table 1 by $(n_s - 1)$, we get the maximum gain in sales (without detection) by attracting customers of rivals, and beyond 2 sellers the gains are larger by this latter route. Since Table 1 is based upon a 10 per cent probability level, strict comparability requires that we use 1.6σ , instead of 2σ , in table 2, which would reduce the entries by one-fifth.
 - 10 Unlike old customers, whose behavior is better studied in a round of transactions, the new customers are a flow whose magnitude depends more crucially on the time period considered. The annual flow of new customers is here taken (relative to the number of old customers) as the unit.
 - 11 And slowly with the number of sellers if customers per seller are held constant.
 - 12 A similar argument leads to a measure of concentration appropriate to potential price-cutting for old customers. Firm i will lose $(1 - p)n_o s_i$ old customers, and firm j will gain $(1 - p)n_o \frac{s_j s_i}{1 - s_i}$ of them, with a variance of $(1 - p)n_o \frac{s_j s_i}{1 - s_i} \left(1 - \frac{s_j}{1 - s_i}\right)$. If we sum over all $i (\neq j)$, we obtain the variance of firm j 's sales to old customers of rivals $(1 - p)n_o s_j (1 + H - 2s_j)$, to an approximation, and summing over all j , we have the concentration measure, $(1 - p)n_o (1 - H)$. The agreement of this measure with

that for new customers is superficial: that for new customers implicitly assumes pooling of information and that for old customers does not.

- 13 Of course, price itself usually falls somewhere in this range rather than at the pole. The traditional assumption of stationary conditions conceals this fact.
- 14 The concentration measures have a rank correlation of .903. The profitability measures have the following rank correlations:

	Return on All Assets	Ratio of Market to Book Value
Return on net worth	.866	.872
Ratio of market to book value	.733	—

- 15 See "Labor Department Examines Consumers' Prices of Steel Products," *op. cit.* p. 133.
- 16 For example, it will be possible to test the prediction that prices will be higher and less dispersed in sales on public bids than in privately negotiated sales, and the prediction that price-cutting increases as the number of buyers diminishes.