Cost Reduction, Competition, and Industry Performance
Author(s): Michael Spence
Source: *Econometrica*, Vol. 52, No. 1 (Jan., 1984), pp. 101-122
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/1911463

Econometrica, Vol. 52, No. 1 (January, 1984)

COST REDUCTION, COMPETITION, AND INDUSTRY PERFORMANCE[1]

BY MICHAEL SPENCE

1. INTRODUCTION

IN MANY MARKETS, firms compete over time by expending resources with the purpose of reducing their costs. Sometimes the cost reducing investments operate directly on costs. In many instances, they take the form of developing new products that deliver what customers need more cheaply. Therefore product development can have the same ultimate effect as direct cost reduction. In fact if one thinks of the product as the services it delivers to the customer (in the way that Lancaster pioneered), then product development often is just cost reduction.[2]

There are at least three sorts of problems associated with industry performance. They occur simultaneously, making the problem of overall assessment of performance quite complicated. The problems are these. Cost reducing expenditures are largely fixed costs. In a market system, the criterion for determining the value of cost reducing R&D is profitability, or revenues. Since revenues may understate the social benefits both in the aggregate, and at the margin, there is no a priori reason to expect a market to result in optimal results. Second and related, because R&D represents a fixed cost, and depending upon the technological environment, sometimes a large one, market structures are likely to be concentrated and imperfectly competitive, with consequences for prices, margins, and allocative efficiency.[3]

These two problems are not unique to R&D. They would and do characterize markets with product differentiation and fixed costs associated with the sale of a differentiated product. The fact that the relevant scale economies are dynamic is of interest, and worthy of exploration. But it is not unique to R&D.[4]

What is distinctive about R&D is that to these differentiation and scale economies problems is added what is often referred to as appropriability problems. They are sometimes referred to as externalities problems. These are really

[1] This research was supported by the National Science Foundation. This paper is an abbreviated version of a working paper with the same title, available at the Harvard Institute for Economic Research.

[2] Suppose that products deliver services to consumers. Let s be the services and $P(s)$ be the inverse demand. Services are delivered through goods. Let x be the quantity of goods, and $c(x)$ be the cost function. Let $f(q)$ be the quantity of services per unit of the good. Then $s = f(q)x$, and the cost of delivering services s is $c(s/f(q))$. If $f'(q) > 0$, and q is raised through R&D of the product development kind, then the effect is to reduce the costs of the service. Thus formally this kind of product development is equivalent to cost reduction.

[3] Allocative inefficiency refers to losses associated with prices in excess of or below marginal costs.

[4] For discussions of the welfare economics of product differentiation, see Dixit and Stiglitz **[4]** and Spence **[10]**.

two versions of the same problem. If the R&D for the single firm is not appropriable, the initial incentives to do the R&D are reduced. On the other hand, the price of the results of the R&D, namely zero, is close to or at the correct price, namely marginal cost. The marginal cost is the cost of transmitting it to other firms. Restoring appropriability is sometimes regarded as a second best solution to the incentives problem because it creates monopoly or monopoly power. It may do that, but it is important to note that it also incorrectly prices the good that the R&D has created, and that by itself has its social costs. An alternative effect of near perfect appropriability (whether created by circumstances or policy) is the creation of redundant and hence excessive levels of R&D at the industry level. That is to say, the levels of cost reduction that obtain may be achieved at an excessively high cost. Thus there appears to be an unpleasant tradeoff between incentives on the one hand and the efficiency with which the industry achieves the levels of cost reduction it actually does achieve, on the other.

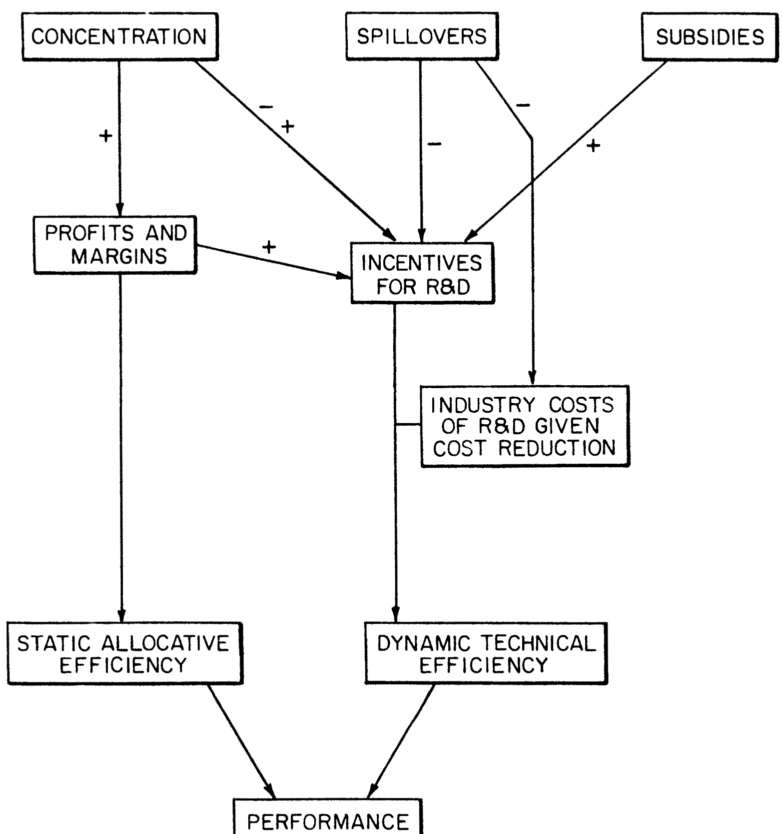


FIGURE 1—R&D market structure relationships.

Figure 1 summarizes the relevant structural characteristics of the market and their impact on performance in various dimensions. For reasons of limited space, I shall not review all of these interactions.[5] Many of them will be familiar to the reader, but the one that deserves emphasis is the negative effect of spillovers on the industry's R&D investment cost of achieving a given level of operating cost reduction over time. Since spillovers reduce those costs, the partial effect on dynamic performance is positive. Of course there is also the adverse negative effect of spillovers on incentives, but if it is important (and I should like to argue that it often is), then there are other ways of restoring incentives.

The analysis that follows is an attempt to describe and evaluate the performance of markets with varying structures. Structure includes concentration, spillovers, and the technology of cost reduction. I will distinguish actual and potential performance, where potential performance refers principally to performance with subsidies. The principal conclusions are two. There are structural environments (which I shall characterize) in which actual performance is poor regardless of the levels of concentration and spillovers. The second and more important is that potential performance is significantly better with high spillovers (or low appropriability). The reason is that the output R&D is essentially a public good; if it is implicitly priced to the potential consumers of it as if it were a private good, the performance of the system will suffer.[6]

2. A MODEL OF COMPETITION AND INDUSTRY PERFORMANCE

There are n firms indexed by i. At time t the ith firm's costs of production and marketing are $c_i(t)$ per unit sold. These are assumed not to depend on output, though with minor changes that assumption can be altered. The vector c is $(c_1, \ldots, c_n)$ with time arguments suppressed. Unit costs $c_i(t)$ depend on accumulated effects of the investment by the firm and possibly by other firms in R&D. Specifically,

$$c_i(t) = F(z_i(t)), \tag{2.1}$$

where $F(z_i)$ is a declining function of z_i, and z_i is the accumulated knowledge obtained by firm i, with respect to cost reduction. Let $m_i(t)$ be the current expenditures by firm i on R&D. Then it is assumed that

$$\dot{z}_i(t) = m_i(t) + \theta \sum_{j \neq i} m_j(t). \tag{2.2}$$

The parameter θ is intended to capture spillovers. If $\theta = 0$ there are no spillovers or externalities; if $\theta = 1$ the benefits of each firm's R&D are shared completely.

[5] One of the casualties of abbreviation is the references to the historical literature where these interactions are discussed. For a fuller treatment, please see Spence [**11**].

[6] Much of the analysis of the paper could be couched in the language of providing public goods competitively with private suppliers incentives influenced by suitable interventions in the price system.

For $0 < \theta < 1$, the spillovers are imperfect. A dot over a variable denotes its time derivative.

For future use we let $M_i = \int_0^t m_i(\tau)\,d\tau$, the accumulated investment in R&D by firm i to date t. Then with that notation

$$z_i = M_i + \theta \sum_{j \neq i} M_j . \tag{2.3}$$

I should say at this point that equation (2.2) embodies the assumption of no diminishing returns to current R&D expenditure. That will make the model essentially static as noted earlier.[7]

The goods produced by each firm are the same (the product is homogeneous). This again is an easily altered assumption.[8]

The benefits in dollars from the sale of x units of the good are $B(x)$. The inverse demand is $B'(x)$. The output by firm i is x_i and $x = \sum x_i$. The profits of firm i are

$$E^i = x_i B'(x) - c_i x_i . \tag{2.4}$$

It is assumed that there is an equilibrium at each point of time in the market, that depends on the costs $c = (F(z_1), \ldots, F(z_n))$, or on $z = (z_1, \ldots, z_n)$. It could be a Nash equilibrium in quantities x_i, or some other equilibrium. All that we require is that it is unique, given c or z. Let $x_i(z)$ and $x(z) = \sum_i x_i(z)$, be the equilibrium. The consumer surplus is just $B(x(z)) - x(x)B'(x(z)) = H(z)$. The earnings gross of R&D expenditures for firm i are

$$E^i(z) = x_i(z)B'(x(z)) - c_i(z_i)x_i(z). \tag{2.6}$$

We turn now to the R&D decision. It can be shown that in this environment, firms will do all their R&D at the outset in a lump. Hence while we could introduce the appropriate intertemporal notation, it would serve no purpose. Thus think of $E^i(z)$ as the present value of firm i's earnings gross of R&D investment. There is a subsidy of s for R&D so that each dollar of R&D costs the firm $(1 - s)$ dollars. Clearly $s = 0$ is a possibility. The present value of its earnings net of R&D investment is

$$V^i = E^i(z) - (1 - s)M_i ,$$

[7] For a treatment of the dynamic, diminishing-returns case, see Spence [**11**]. The diminishing returns case is more realistic. But the qualitative properties of the static model here and the dynamic model are the same.

[8] For example, one can let the inverse demand for the ith firm's differentiated product be $\partial G(\sum \phi(x_i))/\partial x_i$. Then if we let $y_i = \phi(x_i)$, and express costs in terms of y_i, as $c_i \phi^{-1}(y_i)$, we have something that is formally equivalent to a homogeneous product model, but with convex costs. The latter have little or no effect on the models that follow.

where

$$z_i = M_i + \theta \sum_{j \neq i} M_j$$

and

$$z_i(0) = 0.$$

The firm takes the M_j of its rivals as given (and presumed optimal). It maximizes V^i with respect to M_i, by setting

$$E_i^i + \theta \sum_{j \neq i} E_j^i = (1 - s).$$

The solution to these n equations is the market equilibrium. Here E_j^i is the derivative of E^i with respect to z_j. Given the equilibrium values of $M = (M_1, \ldots, M_n)$, the performance of the market can be evaluated by calculating the total surplus

$$T(M) = H(z(M)) + \sum_i V^i(M) - x \sum M_i .$$

The last term reflects the costs to the public sector of the subsidies.

3. THE SYMMETRIC CASE

I should like to focus for the most part on the symmetric case.[9] Except for possible entry deterrence issues arising from negative profits or strategic behavior, the symmetric case will result naturally from symmetry in the costs facing firms. We have symmetry if

$$E^i(z) \equiv E^j(z + (z_j - z_i)e_i + (z_i - z_j)e_j), \tag{3.1}$$

where e_j is a vector with a one in the jth place and zeroes elsewhere. When there is symmetry, it is easy to establish that

$$E_j^i(ve) + \theta \sum_{k \neq i} E_k^i(ve) = E_j^j(ve) + \theta \sum_{k \neq j} E_k^i(ve), \tag{3.2}$$

where v is a scalar and e a vector of ones.

Recall that the conditions for a market equilibrium are

$$E_i^i(z) + \theta \sum_{k \neq i} E_k^i(z) = (1 - s). \tag{3.3}$$

[9] Asymmetries are more likely and more interesting in the case in which $h(m)$ is concave. In that case, firms may fall behind, or find it optimal to stop investing and allow their relative costs to rise.

If for some i, we solve

$$(3.4)\qquad E_i^i(ve) + \theta \sum_{k \neq i} E_k^i(ve) = (1 - s)$$

for v, then ve is a symmetric equilibrium from equation (2.13). Define a new function

$$R(v) \equiv \int_0^v \left[E_i^i(\phi e) + \theta \sum_{k \neq i} E_k^i(\phi e) \right] d\phi.$$

From the preceding remarks, $R(v)$ is independent of i. The equation summarizing the symmetric equilibrium is

$$(3.5)\qquad R'(v) = (1 - s).$$

Thus the market acts as if it were maximizing[10]

$$(3.6)\qquad R(v) - (1 - s)v$$

with respect to v. In the symmetric case $M_i = M$, and hence

$$(3.7)\qquad v = \left[1 + \theta(n - 1)\right]M = K(\theta, n)M,$$

where

$$(3.8)\qquad K(\theta, n) = 1 + \theta(n - 1).$$

To summarize, in the symmetric case, the level of z_i will be the same for all firms. Call it z. The market result in z is the maximum of

$$(3.9)\qquad R(z) - (1 - s)z,$$

R&D expenditures per firm are $M = z/K$, where $K = 1 + \theta(n - 1)$. It should be noted that R depends on n and θ, as well as z, a fact which will be important to us later. The total surplus is

$$(3.10)\qquad T(z, n, \theta) = H(z, n) + nE(z, n) - nM$$

[10]The reader deserves some comments on second order conditions. Let L_i be the second derivative of V^i with respect to M_i. For an equilibrium we require $L_i < 0$. There are interesting cases in which this may fail to hold, creating strategic investment opportunities. I do not have the space to deal with those here. It is straightforward to establish that

$$R''(z) - L_i = (1 - \theta)\left[\sum_{j \neq i} E_{ij}^i + \theta \sum_{j,k \neq i} E_{jk}^i \right].$$

If $\theta = 1$, this difference is zero. If $\theta = 0$, the sign of the difference is the sign of $\sum_{j \neq i} E_{ij}^i$. I shall assume $E_{ij}^i < 0$, so that other firms' investment in cost reduction reduces the return to the ith firm's investment. The terms E_{jk}^i are difficult to sign. Therefore I cannot exclude the possibility that in the intermediate ranges of θ the sign of $R'' - L_i$ is reversed. Even that would not cause $R'' < 0$.

It is clear that if $R(z) - (1 - s)z$ has two local maxima, then at least as far as first order conditions go, the market has two symmetric equilibria. As noted above, that is one of several interesting possibilities. But for this analysis, I am going to assume it does not occur.

where $H(z,n)$ is the consumer surplus and $E(z,n)$ is the earnings per firm gross of R&D investment. Each function depends on the common z, and the number of firms.

4. PROPERTIES OF THE MARKET EQUILIBRIA

The function $R(z,n,\theta)$ captures the market incentives with respect to R&D investment. From the definition of R and assuming that $E_j^i < 0$, $i \neq j$, one can see that $R_\theta < 0$ and $R_{z\theta} < 0$. Thus an increase in the spillovers reduces the incentives for R&D and cost reduction, and will reduce the amount of cost reduction in the market equilibrium.

The dependence of R on n is somewhat complicated: I should like to defer a discussion of that until a specific example is introduced.

The most important relationship is the one expressing the industry's total investment in R&D as a function of z. For a given level of z and $n > 1$, the R&D costs of the achieved amount of cost reduction decline as θ increases. To see this, note that R&D costs at the industry level are

$$\text{(4.1)} \qquad \text{R\&D} = zn/K$$
$$= zn/\left[1 + \theta(n-1)\right].$$

If $\theta = 0$, the costs are proportional to the number of firms. With $\theta > 0$, the unit costs have an upper limit of $1/\theta$ as n increases. For example if $\theta = 0.5$ the unit costs (per unit of z) cannot exceed 2, what they would be with two firms and no spillovers. And if $\theta = 1$, the unit costs are independent of the number of firms and are equal to one.

Thus while spillovers reduce the incentives for cost reduction, they also reduce the costs at the industry level of achieving a given level of cost reduction. The incentives can be restored through subsidies. It can be shown that $R_z > 0$, provided that $dc/dz = F'(z) < 0$, so that subsidies are sufficient to determine the industry costs of cost reduction. It is therefore possible to maximize the surplus, $T(z,n,\theta)$, with respect to z for a given level of n and θ, by setting the subsidy s so as to induce the optimal z, in the relationship $R'(z) = 1 - s$.

From the remarks above concerning industry R&D costs, given z, one can see that spillovers improve the performance of the market with the incentive appropriately restored. Or to put it another way, with appropriability, the achievable surplus is lower because a high rate of cost reduction can only be achieved with a large R&D investment.

5. AN EXAMPLE

It may be clearer if at this point we explore the model in the context of an example. Suppose then that the demand is of the constant elasticity variety so that $x = Ap^{-b}$. And assume further that the static equilibrium is a Nash equilibrium in quantity of output, given unit costs. Let $c = q_0 + c_0 e^{-fz}$ be the unit costs in the symmetric case. Let $w = 1 - 1/bn$, where n is the number of firms and $b > 1$ the price elasticity of demand. In the constant elasticity case

$$(5.1) \qquad R(z,n,z) = \{A/[n(b-1)]\}w^{b-1} \times [2w + (K/n)((b-1)(1-w) - 2w)]c^{1-b}$$

where $K = 1 + \theta(n-1)$.[11] The earnings for the single firms are

$$(5.2) \qquad E(z,n) = Ab[w^{b-1}/n](1-w)c^{1-b}.$$

The consumer surplus is

$$(5.3) \qquad H(z,n) = [A/b - 1]w^{b-1}c^{1-b}.$$

The total surplus is

$$(5.4) \qquad T(z,n,\theta) = Aw^{b-1}[(1/(b-1)) + 1 - w]c^{1-b} - (n/K)z.$$

The market acts so as to minimize

$$(5.5) \qquad Q = R(z) - (1-s)z.$$

These two functions $Q(z,n,\theta)$ and $T(z,n,\theta)$ provide us with a complete summary of symmetric equilibria and market performance. In some instances, a profitability constraint, which would take the form

$$(5.6) \qquad Ab[w^{b-1}/n](1-w)c^{1-b} - [(1-s)/K]z \geq 0,$$

might be binding. But we will not dwell on those cases.

The coefficient of (K/n) in $R(z,n,\theta)$ is negative. Thus an increase in θ increases K for $n > 1$, and hence reduces R and R_z. Other things equal, that increase in spillovers reduces cost reduction, because $R(z,\theta,n)/(1-s)$ is the

[11] The argument is as follows. Given the assumptions, firm i maximizes its profits with respect to x_i so as to satisfy the equation

$$s_i = (1/b)(1 - c_i/p)$$

where s_i is its market share. As a result, since $\sum s_i = 1$, $p = c/w$, where $w = 1 - 1/(bn)$, and $c = \sum c_i/n$. The profits of firm i are thus

$$E^i = (1/b)(1 - wc_i/c)^2 aw^{b-1}c^{1-b},$$

where $c_i = q_0 + c_0 e^{-fz_i}$, and $z_i = M_i + \theta\sum_{j \neq i} M_j$. If one differentiates E^i with respect to c_i and c_j, $j \neq i$, then sets $c_j = c$ and $z_j = z$ for all j, the result is

$$\partial c/\partial z\left(E_i^i + \theta\sum_{j\neq i} E_j^i\right) = [Aw^{b-1}c^{-b}/n][-2w - (K/n)(b-1)(1-w) - 2w]fc_0e^{-fz}.$$

This is $R_z(z,n,\theta)$. Integrating with respect to z gives

$$R(z,n,\theta) = [Aw^{b-1}c^{1-b}/(n(b-1))][2w + (K/n)((b-1)(1-w) - 2w)].$$

function that is implicitly maximized as the market equilibrates. The spillovers unambiguously reduce the amount of cost reduction in an equilibrium.

I should like at this point to describe some calculations I did in order to provide a more quantitative picture of the incentives under various market structures, and of the consequent performance. The example is chosen so that there are significant cost reduction possibilities, but they require significant R&D investments to achieve them. The example has the following parameters: $A = 50$, $b = 2.0$, $q_0 = 1$, $c_0 = 1$, and $f = 0.5$. In this model, what matters is c_0/q_0 and the magnitude of Aq_0^{1-b}. Given c_0/q_0, any combination of A and q_0 that keeps Aq_0^{1-b} constant will give the same results. In this case, $Aq_0^{1-b} = 50$.

The example was not randomly selected, but rather to illustrate certain effects. To place it in perspective, one might imagine varying the parameter f. If f is very small then R&D has little effect on costs and the market will behave as if there were constant unit costs of $q_0 + c_0$. If f is very large, a small amount of R&D will reduce unit costs close to q_0. Every firm will do it and the market will again act like a constant unit cost case. In neither case are there performance problems. The performance problems arise for intermediate cases, in which significant cost reduction is neither prohibitively costly nor very cheap. The example outlined above is an intermediate case. The cost reduction is expensive, but is worth doing from the social standpoint.

In this model if $f = 0$ so that there were no cost reduction possibilities, then the optimal surplus (with price equal to marginal cost) is 25. With $f = 0.5$, the optimal surplus (with price equal to marginal cost) is 41.6456. R&D expenditures are 6.25 and costs after R&D is done are 1.0439.

Now consider market equilibria with $s = 0$ so that there is no intervention. The market maximizes $R(z,n,\theta) - z$. The optimal surplus is the maximum of

$$(5.7) \qquad [A/(b-1)]c^{1-b} - z.$$

Therefore the market incentives for cost reduction relative to the optimum are summarized by the ratio

$$(5.8) \qquad I(\theta,n) = [(b-1)R]/[Ac^{1-b}].$$

This ratio does not depend on c^{1-b}, and hence does not depend on z. For the numerical values above, Table I gives this ratio as a percentage for various values

TABLE I
INCENTIVES FOR R&D: $I(\theta, n)$

	Spillover Parameter				
N	0.00	0.25	0.50	0.75	1.00
1	25.00	25.00	25.00	25.00	25.00
2	32.80	26.90	21.10	15.20	9.40
3	32.40	25.40	18.50	11.60	4.60
4	29.40	22.70	16.10	9.40	2.70
5	26.30	20.20	14.00	7.90	1.80
6	23.60	18.00	12.40	6.80	1.30

of θ and n. The following points should be noted. For $\theta = 0$ (no spillovers) $I(\theta, n)$ rises as n goes from 1 to 2 and then falls. The same is true for $\theta = .25$. For $\theta > .25$, $I(n, \theta)$ declines with n. If you fix n and read across rows, $I(\theta, n)$ declines with θ for each n. Thus spillovers reduce R & D and cost reduction. When spillovers are small, some competition increases the amount of cost reduction. But then the fragmentation of the market overcomes the incentive provided by the downward price pressure and R & D falls.

What is striking about Table I however, is the smallness of the numbers. The largest value for $I(\theta, n)$ is 32.8, which means that the implicit objective function of the market attaches benefits to R & D that are less than one third of the appropriate ones.

It is not therefore surprising that the performance of the market is not particularly good under any market structure. For the same combinations of θ and n, Table II reports the ratio of the surplus actually achieved in the market, to the first best optimal surplus, as a percentage. Underlying these calculations is the fact that allocative efficiency is increasing with n, because margins are falling. The highest value for the performance ratio is 80.3. It occurs with $n = 2$ and $\theta = .25$. The modest spillovers reduce industry R & D costs without excessively removing incentives. But even here, there is a 19.7 per cent loss of potential surplus. For monopoly the loss is 34 per cent. When n is large, there is no cost reduction. And while allocative efficiency is high, 40 per cent of the surplus is lost because of the failure to reduce costs.

The appropriate conclusion seems to be that a market with this type of underlying structure will have performance problems.

Since spillovers have a pronounced negative effect on cost reduction and performance, it may seem appropriate to conclude that where possible they should be eliminated. That would be a mistake. First, the absence of spillovers does not eliminate performance problems as these calculations illustrate. Second, as discussed earlier, appropriability raises industry R & D costs associated with a given level of cost reduction. Thus if one wants to operate on incentives through subsidizing R & D, then it is better to do it in the lower cost (i.e. higher spillover) environment.

Let me illustrate these points with the same numerical example. Given n and θ, the surplus is

$$T = Aw^{b-1}\left[(1/(b-1)) + 1 - w\right]c^{1-b} - (n/K)z. \tag{5.9}$$

TABLE II

PERFORMANCE WITH NO SUBSIDIES

	Spillover Parameter				
N	0.00	0.25	0.50	0.75	1.00
1	65.40	65.40	65.40	65.40	65.40
2	79.30	80.60	77.00	56.30	56.30
3	74.30	80.30	74.40	58.40	58.40
4	67.50	77.90	61.80	59.10	59.10
5	61.40	74.70	59.40	59.40	59.40
6	56.30	70.50	59.60	59.70	59.70

Let $L = T + (n/K)z$. The market maximizes $R - (1 - s)z$. Therefore if the subsidy is set so that

$$(5.10) \quad \left[R/(1-s)\right] = LK/n,$$

then the market will maximize $T(z, n, \theta)$ with respect to z. Note that R/L does not depend on z. Increasing n raises w and hence lowers price-cost margins. But it also raises n/K except when $\theta = 1.0$. Moreover the rate of increase of n/K with n is a declining function of θ. Thus if the optimal subsidies are in place we should observe the following. Performance will increase with θ. As θ rises, the desirable number of firms will increase, since with high spillovers, adding firms raises industry R&D costs less. We could of course introduce a fixed cost to having a firm (doing R&D) and then actually do the optimization with respect to n, given θ. With the data in the tables that follow, you can do that by eye for any choice of the fixed cost.

Table III is the analogue of Table I. It is the ratio of the benefits implicitly recognized by the market to the optimal surplus. But here the market is, for each n and θ, provided with the optimal subsidy. As the reader can see, incentives rise with θ, reading across rows. They rise and fall with n (reading down columns). When $\theta = 1$, they just rise with n, because of the absence of the redundancy problem. The incentives are significantly higher, except when $\theta = 0$. When $\theta = 0$, the subsidies are essentially impotent. They do raise incentives for $n = 2$ (see, for example, Table I). Beyond that the redundancy costs overwhelm the competitive effect on margins. The optimal subsidies for each case are in Table IIIA.

Table IV provides the figures for the performance relative to the first best optimum in percentage terms. The figures in Table IVA are the amounts by which unit costs in the equilibrium exceed the optimally reduced unit costs.

With $\theta = 0$, subsidies have little power to alter performance. With $\theta = .5$ and $n = 3$, performance is over 90 per cent. With $\theta \geq .75$ performance is over 95 per cent and you can tolerate relatively large numbers of firms. But increases in benefits at the margin through adding firms are relatively small for $n \geq 5$ or 6.

These results are of course perfectly consistent with what the theory led us to expect. What may be new is the magnitude of the performance problems in the markets without some type of intervention.

One could repeat the calculations for many examples. The conclusions don't change. If you make f small, of course, the relative importance of allocative efficiency versus dynamic technical efficiency rises and those elements of structure that influence the latter become less critical as determinants of performance. Similarly, if f is larger so that the R&D investment required to achieve substantial cost reduction is small, then the qualitative effects are the same, but most market structures perform reasonably well on the cost reduction dimension. Thus their differentials are more closely related to margins and allocative efficiency.

None of this is surprising. If R&D is either ineffective in reducing costs or very effective and hence relatively cheap, markets perform well. But in cases where the opportunities for cost reduction are substantial, and the costs of achieving them are also substantial, but not prohibitive, there are potential

TABLE III

Incentives with Optimal Subsidies

	Spillover Parameter				
N	0.00	0.25	0.50	0.75	1.00
1	75.00	75.00	75.00	75.00	75.00
2	46.90	58.60	70.30	82.00	93.80
3	32.40	48.60	64.80	81.00	92.70
4	24.60	43.10	61.50	79.90	98.40
5	19.80	39.60	59.40	79.20	99.00
6	16.60	37.20	57.90	78.60	99.30

TABLE IIIA

Optimal Subsidies in Percentages

	Spillover Parameter				
N	0.00	0.25	0.50	0.75	1.00
1	66.70	66.70	66.70	66.70	66.70
2	30.00	54.00	70.00	81.40	90.00
3	0.00	47.50	71.40	85.70	95.20
4	− 19.40	47.20	73.90	88.20	97.20
5	− 32.70	49.10	76.40	90.00	98.20
6	− 42.30	51.70	78.60	91.30	98.70

TABLE IV

Performance with Optimal Subsidies

	Spillover Parameter				
N	0.00	0.25	0.50	0.75	1.00
1	71.40	71.40	71.40	71.40	71.40
2	80.20	84.80	88.20	90.80	92.80
3	74.20	84.00	89.90	93.90	96.80
4	67.90	82.30	90.00	94.90	98.20
5	62.90	80.70	89.80	95.30	98.80
6	59.90	79.40	89.40	95.40	99.20

TABLE IVA

Cost Reduction with Optimal Subsidies

	Spillover Parameter				
N	0.00	0.25	0.50	0.75	1.00
1	1.60	1.60	1.60	1.60	1.60
2	5.80	3.40	2.00	1.00	0.30
3	12.00	5.30	2.60	1.10	0.12
4	20.50	6.90	3.00	1.10	0.05
5	33.30	8.20	3.30	1.20	0.03
6	63.00	9.10	3.50	1.30	0.01

performance problems of considerable quantitative significance. The numerical example was selected to illustrate this last case.

It is worth noting that the optimal subsidies do not depend on f in this model. Hence one does not need to have a view in advance about the magnitudes just discussed in order to be able to proceed with the problem of approximating reasonable policy environments.

One might ask how profitable are these markets. With no spillovers, profits go negative at $n = 4$. With $n = 2$ and $\theta = 0$, the return on the initial investment in R&D is 22.4. With $n = 3$ and $\theta = 0$, the return is 11.1 per cent. With $\theta = .25$ and $n = 2$, the return is 31.8 per cent. With $\theta = .25$ and $n = 3$, the rate is 19.7 per cent. When there are spillovers, profits do not go to zero and entry is not blocked, at least not right away.

How would these markets perform with a subsidy of 70 per cent to R&D? The motivation for this question lies in the fact that the optimal subsidy varies with n, θ, and the price elasticity. Neither b nor θ are likely to be directly observable. A 70 per cent subsidy policy is likely to improve performance in most cases, except when θ is zero or small. Table V gives the performance in an equilibrium as a percentage of the first best outcome for the flat 70 per cent subsidy policy. Performance under this policy is quite good relative to market outcomes except at the extremes for θ. When $\theta = 0$, there is far too much R&D: the cost reduction is too expensive. With $\theta = 1$, and for $n \geq 3$, the disincentives created by the public good character of the R&D are too great for the 70 per cent subsidy to overcome.

The use of the example here is intended to illustrate propositions that are already I hope intuitively clear from the theory and not exhaustively to explore the mapping from parameters to market outcomes. It is also intended to show that there are market structures with performance problems of sizeable dimensions.

The reader may have concluded that in circumstances such as these, where redundancy is a problem, cooperative R&D might be useful. This idea may be reinforced by the fact that while cooperative R&D is not common in the U.S., it is used in other countries. Cooperative R&D can be analyzed in this framework. Fully cooperative R&D with n firms produces results identical to that of a monopolist with price-cost margins constrained to $p/c = 1/w$. The reasons are (i) that margins are set by competitive interaction and (ii) each firm's profits

TABLE V

PERFORMANCE WITH A 70% SUBSIDY

	Spillover Parameter				
N	0.00	0.25	0.50	0.75	1.00
1	71.40	71.40	71.40	71.40	71.40
2	77.30	84.10	88.10	89.90	87.30
3	65.80	82.60	89.90	91.40	58.30
4	53.30	80.60	89.90	90.20	59.10
5	41.30	79.00	89.50	87.90	59.40
6	30.30	77.90	88.90	84.80	59.60

gross of R&D costs are $1/n$ of industry profits, and its R&D costs are $1/n$ of industry R&D costs. Therefore the firm wants to maximize $1/n$ of industry profits net of R&D costs. They all agree and maximize net industry profits. Hence they would act like a margin-constrained monopoly.

As we have seen, that won't produce very good performance without a subsidy. But with subsidies, the results are quite good. In the example, a monopolist with constrained margins has profits of $Aw^{b-1}(1-w)c^{1-b}$. With $\theta = 1$,

$$R(z,n,\theta) = (1/n)Aw^{b-1}(1-w)c^{1-b}. \tag{5.11}$$

Thus the objective the monopolist pursues is n times as large as the objective implicitly maximized. Therefore if s^* is the optimal subsidy for n firms and s is the optimal subsidy for the margin constrained monopolist,

$$[n/(1-s)] = 1 - s^* \tag{5.12}$$

or

$$s = 1 - n(1 - s^*). \tag{5.13}$$

Thus with s, a margin-constrained monopolist or a group of firms doing cooperative R&D will duplicate exactly the results of the market with $\theta = 1$ and the subsidy s^*. Table IV gives the performance for various values of n when $\theta = 1$. The subsidies s^* are those in Table IIIA. The corresponding required subsidies for the cooperative R&D case are in Table VI.

There are interesting further questions concerning the desirability of having cooperation on parts of the R&D and competition on the remainder. An adequate treatment would take us beyond the scope of this paper.

6. THE EFFECTS OF UNANTICIPATED SPILLOVERS

There are industries in which the spillovers are high enough that one might expect a problem with incentives for product development. Certain of the electronics industries have high spillovers yet perform apparently quite well in terms of dynamic technical efficiency. To be sure in some of them the subsidies

TABLE VI

OPTIMAL SUBSIDIES WITH COOPERATIVE R&D

N	Subsidy
1	66.70
2	80.00
3	85.70
4	88.80
5	90.90
6	92.30

(direct and indirect) have been substantial. But in others they have not. In view of the qualitative predictions of the preceding model, such industries are something of a puzzle. It occurred to me in this context that firms might imperfectly anticipate or even ignore the effects of their own R&D investments on the costs of other firms and/or industry prices. That is to say that they might believe that this sort of investment would give them a cost advantage, perhaps through being poorly informed about the spillover effects.

A failure to anticipate correctly the effect of one's own R&D on industry prices does not of course imply that firms underestimate the industry rate of price and cost reduction. They could be quite correct about the latter and still not fully take account of the spillovers.

The effect of underestimating or ignoring spillovers is to make the investment decisions of firms more aggressive, because the return is perceived to be higher than it actually is. In this note, my purpose is simply to investigate what the effects of this altered investment behavior is on entry barriers and market performance. They are substantial. Markets characterized by high spillovers but populated by firms that underestimate the spillovers or their effects on prices, perform much better than the same markets populated by fully informed firms. In this respect, underestimated spillovers act like subsidies. There is a difference however. Subsidies lower entry barriers and increase the number of viable competitors in an equilibrium. More aggressive R&D investments based on underestimated spillovers increase entry barriers and reduce the number of viable competitors. These points are intuitively clear from the theory. Ignoring the effects of spillovers is equivalent to removing a negative term in the expression

TABLE VIIA

PERFORMANCE WITH SPILLOVERS IGNORED

	Spillovers				
No. of Firms	0.00	0.25	0.50	0.75	1.00
1	70.90	70.90	70.90	70.90	70.90
2		84.80	87.98	90.26	91.96
3				92.63	94.79
4					94.54
5					93.18
6					91.14
7					

TABLE VIIB

PERFORMANCE WITH SPILLOVERS RECOGNIZED

	Spillovers				
No. of Firms	0.00	0.25	0.50	0.75	1.00
1	65.39	65.39	65.39	65.39	65.39
2	79.30	80.59	77.07	56.28	56.28
3			74.36	58.36	58.36
4			62.20	59.09	59.09
5			59.43	59.43	59.43
6			59.61	59.61	59.61
7			59.72	59.72	59.72

for the marginal return on R&D investment. The positive effect on performance comes from the increased efficiency of the R&D investments at the industry level, while the failure to anticipate spillovers partially solves the incentive problem created by spillovers under fully anticipated effects. Let me therefore simply use the previous example of Section 5 to illustrate the quantitative effect of unanticipated spillovers. In the tables below, I have reported the results on market performance for various values of n, the number of firms, and θ, the spillover parameter. The latter ranges from zero (no spillovers) to one (which corresponds to a situation in which all the output of R&D by any firm is a public good).

The model in the paper, that is used again here was chosen to illustrate poor performance. That is the R&D is expensive but not prohibitively so from a social standpoint. We know in advance that these are the problem cases. Either very inexpensive or prohibitively expense cost reduction will lead to limited R&D investments, relatively low entry barriers and acceptable performance. But that is not true in the intermediate cases.

There are three tables of results. Table VII reports market performance as a percentage of the first best. Table VIII reports the return on sales for the firms in the industry. Table IX reports the percentage by which unit costs at the equilibrium exceed unit costs in the first best optimum. In all three tables, I have given the results for (A) the cases where firms ignore the spillovers and take the industry price as given, and (B) the case in which firms correctly anticipate and respond to the spillover effects. Case B corresponds exactly to the paper, and is included here for ease of comparison. I have also reported results only for those cases in which firms earn positive profits. Thus for a given level of θ, the equilibrium number of firms is the number corresponding to the last entry in the column of the table. The only qualification is that for $\theta \geq 0.5$ in the tables, in case B there is not R&D investment for large numbers of firms, and no entry barriers. The only source of entry limitation here is the fixed cost character of the R&D investment.

The tables illustrate several points. In markets with high spillovers (above .25 in the tables), performance is improved significantly when firms ignore the spillovers. Really here you get the best of both worlds. The spillovers increase the efficiency of the cost reduction process at the industry level without damaging incentives because firms ignore the damaged incentives.

The second observation is that the equilibrium number of firms declines when the spillovers are ignored. The entry barriers are higher because firms invest more. The market share needed to cover the investments is therefore larger and the number of viable competitors correspondingly reduced.

The monopoly case may look odd. The reason the numbers differ between cases A and B is that in case A I have assumed that the firms ignore not just spillovers. They take the price as given. That obviously is not realistic for a monopolist. If you have firms ignore only the spillovers, the results are similar but less dramatic. They fall between the two cases reported here.

Whatever the appropriate assumptions are concerning the capacity of firms to recognize and respond to the effects of their own R&D on competitors costs, it is

TABLE VIIIA

RETURN ON SALES: SPILLOVERS IGNORED

No. of Firms	Spillovers 0.00	0.25	0.50	0.75	1.00
1	29.40	29.40	29.40	29.40	29.40
2	− 3.70	2.04	5.86	8.60	10.65
3		− 7.20	− 1.24	2.34	4.73
4				− 3.09	2.09
5					0.75
6					0.06
7					− 0.23

TABLE VIIIB

RETURN ON SALES: SPILLOVERS RECOGNIZED

No. of Firms	Spillovers 0.00	0.25	0.50	0.75	1.00
1	36.12	36.12	36.12	36.12	36.12
2	2.63	9.27	14.73	25.00	25.00
3	− 13.31	− 2.41	8.50	16.67	16.67
4			10.55	12.50	12.50
5			9.99	9.99	9.99
6			8.33	8.33	8.33
7			7.14	7.14	7.14

TABLE IXA

COST RELATIVE TO FIRST BEST: SPILLOVERS IGNORED (EXPRESSED AS THE PERCENTAGE BY WHICH THE MARKET OUTCOME EXCEEDS THE FIRST BEST)

No. of Firms	Spillovers 0.00	0.25	0.50	0.75	1.00
1	5.03	5.03	5.03	5.03	5.03
2		3.82	3.82	3.82	3.82
3				5.94	5.94
4					8.71
5					11.95
6					15.78
7					

TABLE IXB

COSTS RELATIVE TO FIRST BEST: SPILLOVERS RECOGNIZED

No. of Firms	Spillovers 0.00	0.25	0.50	0.75	1.00
1	19.75	19.75	19.75	19.75	19.75
2	11.67	17.00	28.45	91.70	91.70
3			39.90	91.70	91.70
4			78.93	91.70	91.70
5			91.70	91.70	91.70
6			91.70	91.70	91.70
7			91.70	91.70	91.70

clear that whether or not they do so respond has a significant effect on market performance, except in cases where the spillovers are minimal. In some respects it is ideal when firms in a high spillover environment ignore the feedback of spillovers on industry costs and prices. The performance of the market is improved because the disincentive effect of spillovers has been assumed away. In fact performance is similar to that achievable with subsidies. The principal difference between spillovers overcome by subsidies and spillovers overcome by a failure to recognize the feedback effect is that in the latter case, entry barriers remain high because the R&D investments are large, and have the character of fixed costs.

Here I have indicated the impact of spillover effects that are ignored completely. There are obvious intermediate cases in which they are imperfectly perceived and partially ignored. The results are correspondingly positioned between the extremes. And the remarks above apply with the prefix "to the extent that" modifying each assertion.

In the end, it is an empirical question whether there are important instances of imperfect perception of spillovers. I would only conclude that ignoring spillovers for the firms in the market is not a disfunctional policy, in the sense that it leads to disastrously unprofitable results. At least for moderate spillovers, the results are high entry barriers, high rates of return, and dynamic performance that is not likely to attract negative critical comment from observers.

7. CONCLUSIONS

R&D has proved a complex subject because there are several interacting simultaneous market failures. First, there will generally be dynamic returns to scale, which result in entry barriers and therefore imperfect price competition. Second, because profits (before R&D investment costs) and social benefits differ both absolutely and at the margin, there is no a priori theoretical assurance that R&D at the industry level will be at the desired level. Third, there are the appropriability problems. It is commonly argued that imperfect appropriability dilutes incentives for R&D; and it does. But the failure is appropriability itself, not its absence. Once acquired, the marginal cost of the knowledge which is the output of the R&D, is its transmission cost. For the remainder of the discussion, I shall assume that cost is near zero, though in certain cases, it may not be. An industry organization that places a non-zero price on R&D has the potential of performing poorly.

In an unregulated market, the incentives for R&D are suboptimally low. The incentives deteriorate with spillovers (the absence of appropriability). And the incentives rise and then fall as concentration declines, at least in some cases. If the spillovers are large, incentives may simply decline with concentration as we have seen in examples. This seems to suggest that reasonably concentrated industries (an outcome the market will produce anyway because of the entry barriers that the fixed R&D costs erect) combined with as high a level of appropriability as is achievable, will produce the best feasible results.

I hope the preceding discussion casts doubt on this view. Appropriability involves implicitly an incorrect pricing decision, and thus concentration plus appropriability will not solve the problem. Market performance is not adequate.

The theory tells us to price the output of R&D at its marginal cost once done, that is zero, and hence to view spillovers as a positive attribute. The result is an incentive problem. The most direct way to deal with that problem is to subsidize the activity for which the market provides suboptimally low incentives. That has the added benefit of lowering entry barriers, increasing competition, lowering margins and improving allocative efficiency. But mainly it makes the transformation from the input, R&D expenditures, to the output, cost reduction, efficient at the industry level. Therefore it is not surprising that in the examples cited here, and in any others that one might examine, high performance of the market occurs in the context of high spillovers and appropriate subsidies. If high spillovers or low appropriability are hard to achieve, cooperative R&D with appropriate subsidies will also lead to better performance.

It is interesting that spillovers and appropriability are the source of the problem, though as noted in Section 6, they are a much reduced problem if the spillovers are not fully recognized by competitors. The fact that the producer of knowledge does not appropriate all the social benefits leads to the conclusion that it should at least be rewarded for the benefits it confers on other firms. And even that falls short of the social benefits. But it does not follow that other firms should pay for it. If they do pay for it, they are paying more than its marginal cost. There is a direct analogue with public goods. The output of R&D has the character of a public good. The incentives are weak for individuals to supply it. But we do not generally approach the solution to public goods problems by contriving to have the beneficiaries pay for it where possible, because that leads to underconsumption and suboptimal use. It is preferable to supply the public good publicly or subsidize the private supplier without paying for the subsidy by charging the users on the basis of use. The R&D problem is essentially the same. The mistake is to attempt to solve the problem by having the price paid to the supplier equal the price paid by the recipient of the benefits.

Harvard University

Manuscript received February, 1982; final revision received December, 1982.

APPENDIX

A REGULATED SINGLE FIRM

An unregulated monopoly performs relatively poorly because its margins are too high and because it lacks the incentives to do R&D (Tables I, II). On the other hand, a single firm does not duplicate R&D and hence produces cost reduction at the industry level efficiently. A monopoly that is subsidized performs better (Tables III, IV), but it still has high margins.

Regulating margins without subsidizing R&D is counterproductive. With $p/c = 1/w$, the profits of the single firm are

$$\pi = [Aw^{b-1}(1-w)]c^{1-b} - z. \tag{A.1}$$

The term in square brackets has a maximum at $\bar{w} = 1 - 1/b$. That is the profit maximizing price cost margin and it provides the maximum incentives for R&D. As w increases so that margins fall, the investment in R&D falls. The effect of constraining margins is to further reduce dynamic technical efficiency. A margin constrained single firm will underinvest in R&D unless the latter is subsidized at the levels shown in Table VI for the example in the paper.

Regulating price has a quite different effect. The profits of the single firm are

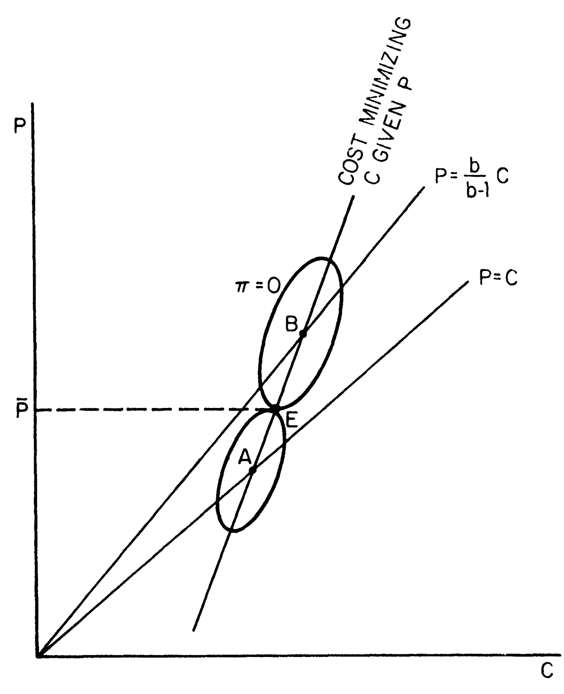
$$\pi = A(p - c)p^{-b} - z. \tag{A.2}$$

If the price is regulated, and if the firm is required to meet the demand at that price, it will set z so as to minimize total costs:

$$c = Ap^{-b}c + z. \tag{A.3}$$

Given p, that is the optimal level of z. Thus a price regulated single firm will set R&D optimally given that price. Now if the price corresponds to the optimum, then the optimal $p = c(z)$. Under these conditions, the firm would have profits of $\pi = -z < 0$.

Figure 2 provides a picture of the incentive structure of this situation. The line $p = c$ is the optimal price given cost, while $p = (b/(b-1))c$ is the profit maximizing price given cost. The line MN is the total cost minimizing c given p; that is both the surplus and the profit maximizing c given p. Point B is the monopoly outcome. Point A is the optimum. Isoprofit contours and iso-surplus contours are vertical and hence tangent to each other along MN. Therefore at a point like C, $\pi = 0$ and $p = \bar{p}$, the



A = FIRST BEST OUTCOME

B = UNREGULATED MONOPOLY

E = SECOND BEST WITH π = 0

AND

MARKET OUTCOME WITH REGULATED PRICE P̄

FIGURE 2—Price regulated monopoly.

outcome has two properties. It is the maximum surplus subject to $\pi \geq 0$. And it is the point the firm would choose if $p = \bar{p}$ is set. Thus a monopoly confronted with the price that emerges from the second best optimum calculation, will invest the second best optimum amount of R & D.

REFERENCES

[1] ARROW, K. J.: "Economic Welfare and the Allocation of Resources for Invention," in *Essays in the Theory of Risk Bearing*, ed. by K. J. Arrow. Chicago: Markham, 1971.

[2] DASGUPTA, P., AND J. STIGLITZ: "Industrial Structure and the Nature of Innovative Activity," *Economic Journal*, 80(1980), 266–293.

[3] ———: "Uncertainty, Industrial Structure and the Speed of R & D," *Bell Journal of Economics*, 11(1980), 1–28.

[4] DIXIT, A., AND J. STIGLITZ: "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(1977), 297–308.

[5] FISHER, F. M., AND P. TEMIN: "Returns to Scale in Research and Development: What Does the Schumpeterian Hypothesis Imply?" *Journal of Political Economy*, 81(1973), 56–70.

[6] KAMIEN, MORTON I., AND N. L. SCHWARTZ: "Market Structure and Innovation: A Survey," *Journal of Economic Literature*, 13(1975), 1–37.

[7] NELSON, R., AND S. WINTER: *An Evolutionary Theory of Economic Change*. Cambridge: Harvard University Press, 1982.

[8] SCHERER, F. M.: "Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions," *American Economic Review*, (1975),

[9] SCHUMPTER, J. A.: *Capitalism, Socialism, and Democracy*. New York: Harper and Row, 1950.

[10] SPENCE, M.: "Product Selection, Fixed Costs and Monopolistic Competition," *Review of Economic Studies*, 43(1976), 217–235.

[11] ———: "Cost Reduction, Competition and Industry Performance," Harvard Institute of Economic Research, Discussion Paper, 1982.