The New Industrial Organization and
the Economic Analysis of Modern Markets*

Richard Schmalensee

WP[]33-80 July 1980

*Prepared for presentation as an Invited Symposium at the Econometric Society Fourth World Congress. Research support by the U.S. Federal Trade Commission and useful advise from Paul Joskow are gratefully acknowledged, though only I can be held responsible for this essay's flaws or the opinions it expresses.
1. The New Industrial Organization

It is generally accepted that the modern field of industrial organization began with the work of Edward Mason and others at Harvard in the 1930's. Lacking faith in the ability of available price theory to explain important aspects of industrial behavior, Mason called for detailed case studies of a wide variety of industries. It was hoped that relatively simple generalizations useful for antitrust policy, among other applications, would emerge from a sufficient number of careful studies. Perhaps because such generalizations were not actually uncovered very rapidly by case analysis, or perhaps because of easier access to data and computers, the case study approach was generally abandoned by the early 1960's. Most students of industrial organization followed Joe Bain (1951, 1956) and turned instead to cross-section studies, electing "to treat much of the rich detail as random noise, and to evaluate hypotheses by statistical tests of an inter-firm or inter-industry nature." The need to describe each firm or industry in the sample by a small number of more or less readily available measures effectively limited consideration to relatively simple hypotheses not involving "the rich detail" so important to students of particular industries. Thus the standard regression equation in this literature specified some measure of profitability as a linear function of a concentration ratio and, usually, other similar variables. Bain's (1968) major text similarly focused on simply-stated qualitative generalizations and contained almost no formal microeconomic theory.

Despite the original prescription of Edward Mason, practitioners in this area have moved from an early reliance on case studies and toward the use of econometric methods of analysis. To a large extent, therefore, a review of econometric studies of industrial organization is a review of much of the content of the field.

Recognizing that methods of scientific inquiry inevitably change, Weiss (1971, p. 398) opined at the end of his survey that "[p]erhaps the right next step is back to the industry study, but this time with regression in hand."

Some econometric work on individual industries had of course been done when Weiss wrote; MacAvoy's (1962, 1965) studies of natural gas price formation and nineteenth century railroad cartels are particularly noteworthy. A great deal more of this sort of analysis was done during the 1970's, however. The investigations of insurance markets by Joskow (1973) and Smallwood (1975) and of airline regulation by Douglas and Miller (1975) are good examples. In contrast to the earlier cross-section work and even to some of the still earlier case studies, the industry-specific econometric analyses of the 1970's seem to have been more concerned with understanding the particular industry at hand than with developing or testing simple propositions that might apply to all markets. This may have reflected a shift in scientific interest toward the fine structure of markets. In the U.S. at least, it likely also reflected a rise in the importance of industry-specific regulation relative to antitrust policy. All the industries mentioned above were regulated in the U.S. Regulation can at least in principle respond to an industry's idiosyncratic features in a way that is difficult for antitrust policy, which must ultimately be based on relatively simply-stated rules that apply to all markets.

Not only were scholars pulled toward industry-specific analysis, they were pushed away from cross-section regressions. After Weiss (1971) wrote, critics such as Demsetz (1973), Mancke (1974), and Phillips (1976) began to
demonstrate the extreme difficulty of drawing firm conclusions about causation from the sorts of cross-section regressions that began filling the journals in the 1960's. Those regressions now seem much less central to the field of industrial organization than they were a decade ago.

Along with a shift in the focus of econometric analyses away from cross sections and toward particular industries has come an important change in the role and status of formal theory in industrial organization. The first, Masonian wave of case studies were explicitly part of an inductive enterprise distrustful of received theory. One also finds very little explicit theorizing in the cross-section literature; a priori arguments are typically limited to verbal justifications for the inclusion or exclusion of particular variables on the right-hand side of a single linear equation. In the 1960's, however, students in good graduate programs were learning that one had to have a formal structural model, not just a list of plausible candidate independent variables, in order to do serious econometrics. Thus the empirical essays of Joskow (1973) and Smallwood (1975) mentioned above contain more explicit development and use of theory than most of the early, classic, book-length industry studies, and they are not atypical in this regard.

The tools of theoretical analysis available to well-trained economists today are much more powerful than those Mason and his contemporaries had. In recent years, these tools have been employed with increasing frequency to construct formal models that either attempt to do justice to "the rich detail" of particular industries or promise to be helpful in the analysis of classes of real markets. Indeed, a sizeable literature has lately grown up in what can only be called "the pure theory of industrial organization";
theory that is designed to help one analyze individual real markets correctly but that is not tied to or based upon any particular set of facts. To paraphrase Weiss, the rallying cry of many of those working in industrial organization in the 1970's seems to have been "back to the industry, but this time with the tools of modern economic theory in hand." Michael Spence has recently provided a revealing description of this approach:4

My instinct as an economist is to study industries on a case by case basis, applying and adapting models as appropriate. For those of us who do this kind of work, the differences among industries may seem more important or interesting than the similarities. And thus we are uncomfortable with general rules.

This new industrial organization of the 1970's differs from that of both classical industry-studiers and cross-section regression-runners in a number of respects. First, though the focus is on understanding the particular, formal theory is used intensively, and its power is appreciated. If nothing else, formal modeling serves as a check on the tendency of verbal argument to make any imaginable form of conduct sound plausible in small numbers situations, the same sort of check provided by close examination of actual conduct. Both checks are easily by-passed in the cross-section econometric approach. Second, in "applying and adapting models as appropriate," the investigator goes beyond mechanical use of textbook polar case analysis of competition and monopoly. Just as industrial organization economists began to become econometricians in the 1960's, many began to become theorists in the 1970's. Third, the systematic search for simple generalizations of the sort that Mason hoped to find in case studies, the same sort that cross-section regressions seek, is essentially abandoned. This is not inconsistent with the emphasis on development of tractable, and thus simple, formal models, since these are taken to be tools useful for understanding "the rich detail" of reality. In any case,
Spence's comments make it clear that faith in the adequacy of simple general rules for either market analysis or public policy is no longer universal.

Much of the interesting theoretical work in industrial organization deals with markets in which the offerings of rival sellers are essentially identical and buyers are very well informed. Considerable attention is paid to oligopolistic interaction and to the strategic use by established sellers of first-mover advantages and economies of scale to protect monopoly profits from outside entry. Studies by Spence (1977, 1979), Dixit (1979, 1980) and others go beyond the familiar criticisms of the Bain-Sylos limit-pricing model, which was developed in the 1950's, to the construction of more satisfactory models of entry deterrence in which all actors behave rationally. 5

In many real-world markets, however, buyers do not perceive all sellers' products as identical, and not all buyers are well informed. Markets with product differentiation and non-price competition were forcibly brought to economists' attention by Chamberlin (1933); they were not considered explicitly in the Marshallian price theory he inherited. Formal analysis of the consequences of imperfect buyer information about price seems to have begun with Stigler's (1961) seminal work. For many products, however, especially those sold in supermarkets and similar multi-brand outlets, information about quality is at least equally important and much less perfect. In markets where products may differ and buyers may be unsure of the exact differences among them, a central element of seller conduct is product selection. Markets of this sort are more visible and important in modern economies -- with cheap transportation, mass communication, and routine
commercial application of the scientific method -- than in the economies about which Adam Smith and Alfred Marshall wrote. They are thus referred to as "modern markets" here.

The remainder of this essay considers problems that must be faced in the economic analysis of modern markets and the development of analytical tools to cope with those problems. I do not attempt to be comprehensive but rather focus on three sets of issues that seem to me both interesting and important. Section 2 is concerned with two of Bain's (1968) three key dimensions of market structure: concentration and product differentiation. The form of product differentiation is shown to have important implications for the appropriate measurement of concentration. Standard measures, which implicitly assume product homogeneity, can easily lead one to incorrect inferences about the nature of market interaction. A new measure of concentration that deals with these problems is presented.

Section 3 deals with Bain's (1968) third key dimension, conditions of entry. We focus on what he (1956, p. 216) found to be "the most important barrier to entry discovered by detailed study": product differentiation advantages of established sellers. Some suggestive evidence on the nature of those advantages is discussed. A simple model of rational buyer behavior under imperfect quality information is sketched in which differentiation advantages arise naturally. Implications for patterns of competition are discussed.

Finally, advertising is important in many modern markets, and buyer behavior therein involves problem-solving in important respects. Section 4 concludes this essay with a few general remarks about the treatment of advertising and consumer behavior in industrial organization.
2. Market Concentration and Product Differentiation

Concentration is surely the most frequently quantified element of market structure. With no product differentiation, received doctrine holds that seller behavior will be more monopoly-like, at least in the short run before entry can occur, the fewer the sellers or the less equal their market shares. In markets with homogeneous products, it is thus sensible to define and measure "concentration" by means of some function that is decreasing in the number of sellers and increasing in some measure of the inequality of their shares. If concentration is to be used as a predictor of market conduct or performance, one would like to derive the exact form of this function from a generally accepted theory of oligopoly, but no such theory exists.

Two derivations of concentration indices from models of market behavior nonetheless deserve mention. Following Rader(1972, pp. 269-73), let us consider Cournot equilibrium with constant costs. Let \( c_i \) be the unit cost of firm \( i \), with \( i = 1, \ldots , N \), let \( q_i \) be firm \( i \)'s output, let \( Q \) be the sum of the \( q_i \), and let \( P(Q) \) be the market inverse demand curve. Profits are then given by

\[
i = [P(Q) - c_i]q_i, \quad i = 1, \ldots, N, \tag{2.1}\]

At Cournot equilibrium, firm \( i \) sets \( \delta \pi_i/\delta q_i = 0 \) assuming all other outputs fixed. If \( E \) is the absolute value of market demand elasticity, and \( s_i = q_i/Q \) is firm \( i \)'s market share, these equilibrium conditions can be written as

\[
(P - c_i)/P = s_i/E, \quad i = 1, \ldots, N. \tag{2.2}\]
Letting \( \Pi \) be total industry profit, the sum of the \( \pi_i \), one can multiply both sides of (2.2) by \( q_i/Q \) and sum over \( i \) to obtain

\[
\frac{\Pi}{PQ} = \frac{H}{E}
\]  

(2.3)

where the \( H \) index of concentration is defined by

\[
H = \sum_{i=1}^{N} (s_i^4) \quad (2.4)
\]

Proceeding in a very different fashion, Stigler (1964) derives this same index as a measure of the likelihood of collusive behavior in a market with imperfect seller information. In markets with no product differentiation, the \( H \) index thus seems a sensible measure of concentration. I have elsewhere [Schmalensee (1977a)] attempted to show that it can be well approximated using published official data on concentration ratios.

In the Cournot model above, an increase in any one firm's output affects all other firms by reducing the market price. All are affected in proportion to their market shares. With product differentiation, however, this kind of symmetric or generalized interaction need not be present. If it is not, the theoretical rationale for market-wide concentration measures like (2.4) is weakened, as the development below establishes.

Markets with differentiated products began to receive serious attention from theorists in the 1920's. Two polar case models of market demand and seller interaction emerged at the very start of this work. The spatial model of Hotelling (1929) stressed buyer diversity; additional brands made it more likely that any individual buyer would find one well-suited to his particular tastes. The symmetric model usually associated with Chamberlin (1933), on the other hand, involved a representative buyer who benefits from increased product variety. Both polar cases are still used extensively;
compare the spatial analyses of Salop (1979a) and Schmalensee (1978b) with the symmetric models of Spence (1976) and Dixit and Stiglitz (1977).

In the original Hotelling model, brands (of cider in his example) are located along a line in the space of potential products. Buyers' ideal products are spread out along this same line. Each brand competes only with its two nearest neighbors, no matter how long the line is or how many brands it holds. (End brands have only one nearest neighbor.) Because rivalry is thus localized, each firm faces only a small number of actual rivals, no matter what standard market-wide concentration measures imply. Oligopolistic behavior would be predicted in markets of this sort even with many sellers.

In the 1930's, Kaldor (1934, 1935) argued strongly for the spatial view of differentiated markets, and he recognized that it implied a world of overlapping oligopolies. By the early 1950's, Chamberlin (1951, 1953) himself accepted the spatial model as the more useful of the two polar cases. He (1951, p. 68) also clearly recognized that it implied ubiquitous oligopoly, not the large-numbers case with which he is usually associated.

There is no reason to suppose that either extreme model is universally appropriate. One might model the automobile market in spatial terms, for instance, while analyzing the restaurant market in some locality with a symmetric model. In some markets, interactions among rival sellers might have more structure than the symmetric model implies but less than in a one-dimensional Hotelling framework. Lancaster's (1966, 1971) model of demand, where a linear technology converts purchases of goods into consumptions of characteristics about which buyers care, might be able to shed light on these intermediate cases, but so far very little has been done in this
direction. As the analyses of Baumol (1967), Lancaster (1975), and Salop (1979a) have shown, the formal correspondence between Lancastrian models with two characteristics and one-dimensional spatial models is almost exact. In particular, the same localization of competition is preserved. Archibald and Rosenbluth (1975) have further shown that localization is preserved in models with three characteristics. In Lancastrian models with four or more characteristics, however, they demonstrate that in principle the average brand might have a large number of direct competitors. Conditions that would either guarantee or rule out this possibility are apparently not known, and no work within the Lancastrian framework has apparently sought useful summary statistics to describe intermediate degrees of localization.

The marketing literature contains both symmetric and spatial models, though the latter usually involve more than one dimension. A good deal of econometric work in marketing adopts the symmetric "us/us+them" specification. On the other hand, the construction of "perceptual maps" of brands' locations in product space, based on various sorts of questionnaire data, has become commonplace. These maps are somewhat hard to interpret in economic terms, however, since the meaning of distance is rarely clear. Attempts have also been made to capture the structure of brand interactions by analysis of brand switching data and, recently, by the estimation of nested multinomial logic models based on forced-choice experimental data. These structured approaches are designed to provide insight to those concerned with marketing actual existing brands or seeking profitable niches for new brands. They seem less helpful to an analyst concerned with the general nature of seller interaction in the market as a whole.

I now want to develop a measure of the overall extent to which rivalry
is localized in a particular market. This measure is in turn based on a
measure of concentration that reflects the structure of brands' interactions.
My approach is modeled on that leading to equation (2.3) above, except that
it is both easier and at least arguably more natural in the context of
differentiated products to work with non-price competition.

Thus suppose that the difference between price and unit cost is a constant,
m, for all firms, and assume that total market sales are fixed at Q. Let \( a_i \)
be firm i's effective advertising, with \( c_i \) the unit cost of that advertising.
(Per dollar spent, high quality brands may have more effective advertising.)
If the \( s_i \) are market shares, as above, profits can be written as follows:

\[
\pi_i = mQs_i - c_i a_i, \quad i = 1, \ldots, N. \tag{2.6}
\]

In Nash/Cournot noncooperative equilibrium, with each firm maximizing its own
profit taking the others' \( a_i \)'s as fixed, it is easy to see that the ratio of
actual profit to potential monopoly profit is given by

\[
\Pi/mQ = 1 - \sum_{i=1}^{N} a_i (s_i/\partial a_i). \tag{2.7}
\]

The most natural symmetric demand model in this framework is the following:

\[
s_i = a_i / \sum_{j=1}^{N} a_j, \quad i = 1, \ldots, N. \tag{2.8}
\]

Straightforward differentiation yields

\[
\partial s_i / \partial a_i = (1 - s_i) / \sum_{j=1}^{N} a_j, \quad i = 1, \ldots, N. \tag{2.9}
\]
Substitution into (2.7) then gives us

\[ \Pi/m\overline{Q} = 1 - \sum_{i=1}^{N} s_i(1-s_i) = H. \tag{2.10} \]

The H index thus emerges as a sensible concentration measure in differentiated markets in which competition is not localized at all. Note that (2.8) implies

\[ \frac{\partial s_j}{\partial a_i} = s_j / \sum_{k=1}^{N} a_k, \quad \text{if } j; i, j=1, \ldots, N. \tag{2.11} \]

That is, rivals are affected by any brand's actions in proportion to their own market shares, just as in the homogeneous product Cournot model above. Note also that if the \( c_i \) are nearly equal, so are the \( s_i \), and \( H \) is then approximately \( 1/N \).

Now consider a spatial setup in which \( N \) brands are distributed evenly around a circle with unit circumference, and buyers are distributed uniformly around the same circle. Suppose that brands are numbered consecutively and that each brand competes only with its two nearest neighbors. A tractable demand structure with this property involves the following share equation:

\[ s_i = \frac{3a_i}{N(a_{i-1} + a_i + a_{i+1})}, \quad i = 2, \ldots, N-1, \tag{2.12} \]

with the obvious modification for brands 1 and \( N \).\(^{16}\) If all the \( a_i \) are equal, this structure implies that all brands have shares of \( 1/N \), exactly as in the fully symmetric model, (2.8). Differentiation yields immediately

\[ a_i(3s_i/3a_i) = s_i(a_{i-1} + a_i + a_{i+1})/(a_{i-1} + a_{i} + a_{i+1}) \]
\[ = s_i[1 - (Ns_i/3)], \quad i = 2, \ldots, N-1. \tag{2.13} \]
Since the first and last expressions are equal for $i=1$ and $i=N$ as well, direct substitution into (2.7) produces

$$\Pi/mQ = 1 - \sum_{i=1}^{N} s_i[1 - (Ns_i/3)] = (N/3)H.$$  \hspace{1cm} (2.14)

If costs are roughly equal, the ratio of actual to maximum profit is on the order of $1/3$, no matter how large $N$ is. One can thus think of $(N/3)$ as measuring the extent of localization, or of $(3/N)$ as measuring the extent to which rivalry among the brands in this market is generalized.

In any real market, the investigator is likely to have very incomplete information about the structure of firms' demands. Given high quality estimates of the demand structure, of course, equation (2.7) can be used directly to make predictions about conduct and performance. Unless one has a great deal of confidence in the second-order, curvature properties of these estimates, however, this is likely to be a risky undertaking. Suppose, for instance, that one admits the possibility of shares being determined by a simple generalization of (2.8):\(^{17}\)

$$s_i = (a_i)^e/\sum_{i=1}^{N} (a_j)^e, \quad 0 < e \leq 1, \ i=1, \ldots, \ N.$$  \hspace{1cm} (2.8')

This is clearly a symmetric model, but (2.10) must now be replaced by

$$\Pi/mQ = 1 - \sum_{i=1}^{N} es_i(1-s_i) = (1-e) + eH.$$  \hspace{1cm} (2.10')

The concentration measure developed below avoids dependence on difficult-to-obtain second-order information, like the value of $e$, by essentially building in
It is crucial, however, to have first-order information on the relative values of demand cross-derivatives. It is convenient to deal with that information in the following form:

\[ \theta_{ij} = -\frac{(\partial s_j/\partial a_i)}{(\partial s_i/\partial a_i)} = k_{ij} s_j/(1-s_i), \quad i\neq j, i, j=1, \ldots, N. \quad (2.15) \]

The first equality defines the \( \theta_{ij} \); the second defines the \( k_{ij} \). It is reasonable to assume that all these quantities are non-negative. Because market share must sum to one,

\[ \sum_{j \neq i} \theta_{ij} = 1, \text{ or } \sum_{j \neq i} k_{ij} s_j = (1-s_i), \quad i = 1, \ldots, N. \quad (2.16) \]

The \( \theta_{ij} \) indicate at whose expense firm \( i \) can increase its sales. One can have a good idea of who loses how much if \( i \) gains share, without having any information about how rapidly the marginal product of \( i \)'s advertising is falling off. In the basic symmetric model, (2.8), all the \( k_{ij} \) are equal to unity. When demand has more structure, as in (2.12), they differ in value. All else equal, one would like a concentration measure that increased in response to this sort of departure from symmetry.

I now proceed to construct such a measure. The summation in (2.10) is a share-weighted average of the shares of the total market held by each firm's rivals. Similarly, in (2.14) if the \( s_i \) are approximately \( 1/N \), the summation gives the share-weighted average of quantities approximately equal to \( 2/3 \), the share of each firm's two rivals in the part of the market for which that
firm is competing. The larger any firm is relative to its direct competitors in either case, the smaller is the corresponding term in the summation, and the larger is the ultimate concentration measure. This makes sense, both in terms of the diminishing returns built into (2.8) and (2.12) and in terms of more general notions of oligopoly interaction.

Suppose one thinks of \( (\theta_{ij}/s_j) \) for \( j \neq i \) as a sort of estimate of the reciprocal of the share of the total market held by i's rivals. These estimates are exact if (2.8) holds for all \( j \neq i \), but in general, they will differ. In order to obtain a single number for each \( i \), let us weight these estimates by the \( \theta_{ij} \), as these weights reflect the relative importance of the corresponding rival firms to firm \( i \). This yields

\[
(1 - s_i) = \frac{1}{\sum_{j \neq i} \theta_{ij}^2/s_j}.
\]

On the reasoning above, the relevant quantity is not the share of the total market held by i's rivals, but their share of that part of the sub-market for which i competes. One can estimate this latter share simply as follows:

\[
(1 - s_i)^* = \frac{(1 - s_i)/[s_i + (1 - s_i)]}{s_i + (1 - s_i)}.
\]

Proceeding by analogy with (2.10) and (2.14), our measure of concentration becomes

\[
H^* = 1 - \sum_{i=1}^{N} s_i(1 - s_i)^* = \sum_{i=1}^{N} (s_i^2)^2/G_i,
\]

where substitution from (2.15) - (2.18) establishes...
\[ G_i = s_i + (1-s_i)^2 / \sum_{j \neq i} (k_{ij})^2 s_j \quad i = 1, \ldots, N. \]  

It is straightforward to show that any single \( G_i \) is maximized subject to (2.16) and the non-negativity of the \( k_{ij} \) if \( k_{ij} = 1 \) for all \( i \) and \( j \). In this fully symmetric case, \( G_i = 1 \) and \( H^* = H \), as one would hope. Since the summation in (2.20) is convex in the \( k_{ij} \), \( H^* \) increases with deviations from this symmetric situation in much the same way that \( H \) increases with deviations from market share equality. It is clear that \( G_i \geq s_i \) for all \( i \), so that \( H^* \) has natural bounds:

\[ H \leq H^* \leq 1, \]  

(2.20)

with the left inequality strict except in the symmetric case. These bounds induce limits on \( G^* \), the natural measure of the extent to which rivalry is generalized among competing brands:

\[ 0 \leq G^* = H/H^* \leq 1. \]  

(2.21)

Except in symmetric models like (2.8) or (2.8'), \( G^* \) is less than unity. In the spatial model (2.12) with equal costs, it is straightforward to show that \( H^* = 1/3 \) and \( G^* = 3/N \), as above. One can think of \( 1/G^* \) as a measure of the extent of localization of rivalry.

I do not claim that \( H^* \) and \( G^* \) are the only measures of their general type or the best ones; they are merely the only such measures I have found. They are not intended primarily for theorem-proving but for use in the analysis of actual modern markets. Two remarks about such applications thus seem in order. (I hope to have more to report in the reasonably near future.)
First, symmetric structures like (2.8) or (2.8') are much easier to write down and estimate than nonsymmetric structures of any generality. It would thus seem sensible in most applications to test for deviations from symmetry before attempting to estimate a general matrix of $k_{ij}$. One can base such tests on arguments that under the null hypothesis of symmetry, the cross-equation disturbance covariance matrix from a properly-specified system of estimated share equations should reflect the underlying symmetry in testable ways.

Second, closely related problems are posed by the existence of marketing instruments besides advertising and by firms selling multiple brands in the same market. In some cases, the pattern of cross-effects might be more or less invariant to the marketing instrument employed. That is, if an increase in firm 1's advertising would increase its sales mainly at the expense of firm 2, it might also be that firm 2 would be the main loser if firm 1 lowered price. In such cases, one could estimate the $k_{ij}$ separately for each important instrument of rivalry and average them in almost any straightforward way. If rival firms belong to different strategic groups, in the sense of Caves and Porter (1977) and Porter (1979), these instrument-specific estimates might differ substantially, however. An increase in advertising might affect rivals who advertise a lot, while a cut in price might affect those who charge a low price and advertise very little, for instance. In situations of this sort, it is not immediately obvious how substantially different sets of $k_{ij}$ should be combined.

Similarly, if a firm sells multiple brands, it can take a variety of different actions when attempting to increase its overall share. The analysis above used "firm" and "brand" interchangeably, but in fact a
single firm may sell several brands with very different patterns of \( k_{ij} \). It is again not immediately obvious how these should be combined if one wants a measure of concentration at the firm level, not the brand level. It should be possible to devise an appropriate generalization of \( H^* \) by beginning with models like (2.8) in which firms sell multiple brands, but I have not yet attempted this.

3. Product Differentiation Advantages

Standard usage defines product differentiation by the consequences of its absence: product differentiation is present whenever buyers do not treat the wares of competing sellers as perfect substitutes. It is also standard to follow Bain (1956, p. 3) and evaluate barriers to entry "... by the advantages of established sellers in an industry over potential entrant sellers, those advantages being reflected in the extent to which established sellers can persistently raise their prices above a competitive level without attracting new firms to enter the industry."\(^{19}\) Adopting these definitions, there is no obvious reason why these two dimensions of market structure should be causally related or even correlated in cross-section. Indeed, one can think of examples of markets with considerable apparent differentiation in which entry barriers appear negligible (restaurants) and in which they are at least arguably substantial (breakfast cereals).\(^{20}\)

In his seminal work on conditions of entry, Bain (1956, ch. 4) observed that differentiation might in some instances translate into a preference for established brands over new brands and that this sort of product differentiation advantage could constitute a barrier to entry. Bain found this source of entry barriers to be very important in some of
the markets he studied, and he sought (mainly through interviews) to understand the nature and origins of the corresponding buyer preferences in each case. Summarizing his work on product differentiation advantages, Bain (1956, p. 143) concluded:

All of these things might seem to suggest the existence of fundamental technical considerations, institutional developments, and more or less fundamental consumer traits which make possible or even very probable the development of strong and stable product-preference patterns. They may also suggest that advertising per se is not necessarily the main or most important key to the product-differentiation problem as it affects intra-industry competition and the condition of entry. Although instances are found in which it is, we may need in general to look past advertising to other things to get to the heart of the problem.

Despite this conclusion, most of the relevant cross-section econometric work on the determinants of profitability has followed Comanor and Wilson (1967) and used the ratio of advertising to sales as a proxy variable for both the importance of product differentiation and the significance of product differentiation advantages of established firms over potential entrants. Most of these studies have found a strong positive statistical relation between the advertising/sales ratio and various measures of profitability across industries. This has most commonly been interpreted as reflecting advertising's ability to enhance and protect monopoly profits of established firms. Since the theoretical case for such a causal relation between these variables is less than airtight, there are measurement problems associated with the standard practice of expensing advertising instead of capitalizing it, and the observed statistical relation must reflect firms' advertising budgeting rules as well as
advertising's effect on market structure, the interpretation of the cross-section econometric results has been hotly debated. Whatever the eventual outcome of that debate, these results cannot by their nature refute Bain's conclusion that advertising may not be "the main or most important key" to understanding the relation between product differentiation and conditions of entry, since advertising has generally been the only differentiation-related variable considered in cross-section econometric work.

Bain (1956, ch. 4) does not explicitly assert the nature of this "key", but a number of his remarks suggest that he thinks that buyer uncertainty about product quality is at the heart of the mechanism involved. Thus he notes early in his discussion (p. 116), "There is a good a priori possibility, moreover, that most buyers will on balance prefer established and known products to new and unknown ones." Similarly, he states (p. 130) that within his sample, "the allegiance of consumers to established products in areas in which they are ignorant or uncertain concerning the actual properties of products is quite important." Finally, in summarizing (p. 142) the most common "strategic underlying considerations in strong product differentiation," he begins with "durability and complexity of the product (and corresponding infrequency of purchase by the individual consumer), generally associated with poor consumer knowledge or ability to appraise products, and thus with dependence on 'product reputation', and also with dependence on customer-service organizations." All of this suggests that established firms' advantages in differentiated markets might often depend heavily on buyers' uncertainty about the attributes of new brands, so that being first in a market might often be much more important
than merely spending a lot of money on advertising. Conventional wisdom in marketing and scattered recent empirical work both support the notion that there are important advantages of being the first entrant in some sorts of markets. Runyon (1977, p. 214) states the conventional wisdom clearly: "If the product is virtually identical with [established] competitive products, it has little chance of marketing success." Bond and Lean (1977, 1979) find that important and long-lived advantages are enjoyed by pioneering brands of prescription drugs, advantages that can only be overcome by late entrants if they offer distinct therapeutic benefits. Whittin's (1979) study of cigarette market segments points in this same direction, as does the cross-section analysis of marketing costs by Buzzell and Farris (1976). Urban, Johnson, and Brudnick (1979) use pre-test market analyses of buyer preferences to adjust for quality differences between first brands and later entrants. Based on data for 42 products in 16 consumer goods markets, they conclude that the second brand on average attains less than 60% of the first brand's share, the third brand obtains less than 40%, and so on. The problems faced by an actual entrant into the U.S. reconstituted lemon juice industry [Schmalensee (1979)] are at least consistent with this sort of disadvantage. Finally, experiments reported in the marketing literature by Tucker (1964), McConnell (1968) and others reveal that consumers are willing to pay a premium to continue purchasing brands with which they have acquired experience, even when all "brands" are identical in appearance and in fact.

A simple model involving rational buyer behavior serves to support the idea that product differentiation advantages can be built on differential
information and to shed some light on the mechanism that might be involved. Consider a narrowly-defined product class, like bottled lemon juice, such that individual consumers can be sensibly modeled as using at most one brand in the class at a time. It is assumed that brands either "work" or "don't work"; they either perform as a brand in this class should, or they fail to perform acceptably. Whenever a new brand is introduced, all consumers are naturally uncertain about whether or not it will work. Assume that we have what Nelson (1970) christened "experience goods": the only way a consumer can resolve this uncertainty is to buy a brand and try it. One trial suffices to determine whether or not any brand works.

Let the function $F(v)$, $0 \leq v \leq V$ give the number of consumers willing to pay at least $v$ for a unit of a brand in this class that works. Suppose that prior to the introduction of the first brand, all consumers have subjective probability $\pi$ that it will not work, and all value a unit that doesn't work at $(-\phi v)$, with $\phi \geq 0$. (One might have $\phi > 0$ for a bleach that could ruin clothes, for instance.) Suppose that the time between purchases is constant, call it one period, and let the corresponding one-period discount rate, assumed common to all consumers, be $r$. Given market interest rates, more frequent purchase implies a smaller value of $r$. The assumption that all consumers have the same values of $\pi$, $\phi$, and $r$ is not as restrictive as it might seem, since the development below can be interpreted as applying only to a subset of consumers with the same values of these parameters, and the demand functions obtained can then be summed across all such subsets.
To simplify the analysis, let us assume that the prices of individual brands must be held constant over time. In Schmalensee (1980b) it is shown that the main conclusions developed below go through if price changes are allowed, but the details of the argument are too space-consuming for inclusion here. Let us also assume that consumers are risk-neutral, have infinite horizons, and behave perfectly rationally. Risk-neutrality and infinite longevity merely simplify, while perfect rationality could perhaps even be defended to a non-economist in this context by noting that the consumer's decision problem is relatively straightforward.\(^{25}\)

The solution to that problem is the heart of the model. Suppose that in order to try a new brand, a consumer ceases (for one period) to use a substitute that yields a non-negative surplus (demand price minus purchase price) of \(s\). Then it is rational to try a new brand selling at price \(p\) if and only if the following inequality is satisfied:

\[
\pi[(-\phi - p) + (s/r)] + (1 - \pi)[(v - p)(1+r)/r] \geq s(1+r)/r \quad (3.1)
\]

The first bracketed term on the left gives discounted surplus if the new brand is tried, doesn't work, and the consumer switches back to the substitute. (It is assumed that one must use the entire unit of the new brand in order to evaluate it, so that its failure to work does not trigger early purchase of the substitute. This is not crucial to the argument.) Since the new brand need not be chosen once and for all, the consumer is willing to pay more than the expected gain vis à vis the substitute to try it. The second term on the left of (3.1) capitalizes the stream of surplus associated with a brand that works, and the term on the right gives the benefit associated with continuing to purchase the substitute.
Inequality (3.1) can be re-written as

\[ p \leq (v - s) - \tau v, \quad (3.2) \]

where the quantity \( \tau \) is defined by

\[ \tau = \frac{r(1+\phi)}{(1+r-\pi)}. \quad (3.3) \]

This quantity captures the premium that must be paid to induce trial. If \( \tau = 0 \), condition (3.2) indicates that the new brand will be purchased if and only if its net surplus, \( v - p \), exceeds \( s \). As one would expect, \( \tau \) is increasing in \( \phi \) and \( \pi \). It is also increasing in \( r \). Larger values of \( r \) reflect lower purchase frequency, this serves to increase the importance of any single purchase relative to the entire future stream of purchases, and this makes the risk associated with trying the new brand loom larger relative to the alternative of sticking with the substitute. As noted above, Bain (1956, p. 142) attaches some importance to low purchase frequency in this context.

For the first brand in some particular product class, it is reasonable to take \( s = 0 \). Without loss of generality, we can set the marginal utility of income for every consumer to unity. This means that before the appearance of the product category, every consumer receives zero surplus (on the definition used here) from the marginal unit of every commodity purchased. (We are assuming away indivisibilities.) If only a small fraction of total spending is devoted to the substitute product relevant here, so that income effects can be neglected, it follows that the foregone surplus from reducing spending on the substitute is zero. Alternatively, one can construct
a mechanical (and somewhat strained) normalization argument for setting $s = 0$.

If the first brand is priced at $p_1$ and announced to consumers who have $s=0$, it follows immediately from (3.2) that it will be tried by those with

$$v \geq \frac{p_1}{1-\tau}.$$  

(3.4)

If we suppose for simplicity that the first brand in fact always works, then all who try it stick with it, and its demand function is given by

$$q_1 = F\left[\frac{p_1}{1-\tau}\right].$$  

(3.5)

Note that the assumption that a brand always works is perfectly consistent with consumers' ex ante uncertainty about its quality. Pre-trial advertising may be able to convey a great deal of product information, but as Nelson (1974) has noted, sophisticated consumers must be skeptical about quality assertions that they can directly verify only after purchase. 26 If one sets $V = 1$ by choice of units and if $F(v) = 1-v$ for $0 \leq v \leq 1$, so that consumers' valuations are uniformly distributed over the unit interval, equation (3.5) becomes

$$q_1 = 1 - \frac{p_1}{1-\tau}, \quad 0 \leq p_1 \leq 1 - \tau.$$  

(3.6)

Now suppose a second brand appears that is objectively identical to the first. Again, consumers know that it is worth $v$ if it works and $(-\phi v)$ if it doesn't. Even though the second brand is assumed always to work, it
is again reasonable for imperfectly informed consumers to attach some probability to its not working. If consumers do not talk to each other about this product class, as has been in effect assumed so far, those who did not try brand one have no reason to assign any value but \( \pi \) to the probability that brand two doesn't work. Those who have tried brand one and found it to work might assign a somewhat lower probability to brand two's not working, since they know that it is at least possible to produce a brand in this class that works. For algebraic simplicity, I assume instead that they also have subjective probability \( \pi \) that brand two does not work. It is shown below that greater optimism on the part of this group improves brand two's prospects in a continuous fashion. (That is, group members do not all rush to buy brand two if they are only a little bit more optimistic about its chances of working.)

The second brand's sales of course depend on the first brand's price. In order to highlight the problem of late entry, suppose that the first brand does not alter its price in response to new competition. This is a much more passive response than any considered in the recent literature on entry deterrence cited in Section 1, above. Empirically, it is easy to find cases in which price is reduced in response to attempted entry; see Schmalensee (1979) for an example. Further, unless one assumes that collusion is without cost or difficulty, successful entry that transforms a market from monopoly to duopoly must lower at least the expected value of price under any reasonable model of firm behavior. The assumption that \( p_1 \) will be held constant under all circumstances is thus surely the most optimistic assumption a potential entrant could plausibly hold.

Under this assumption, suppose that brand two enters and sets its
price, \( p_2 \), equal to \( p_1 \). Those consumers who did not find it optimal to try brand one then find it optimal not to try brand two either. If \( p_2 \) were slightly below \( p_1 \), a small number of this group would be induced to try brand two and, since brand two always works, they would stay with it indefinitely. Those consumers who did try brand one are now enjoying a surplus of \((v - p_1)\) on each purchase occasion. For any such person to try brand two would mean giving up this positive quantity, while trying brand one involved giving up a zero surplus. It is easy to see that nobody currently buying brand one will rationally try brand two unless it is discretely cheaper, since otherwise there is no point to bearing the risk of trial. Thus if brand two enters with a price equal to brand one's, it will sell nothing at all, and if it uncercuts brand one a tiny bit, it will have correspondingly tiny sales. The demand conditions facing brand two are not the same as those that faced brand one, because some consumers have made irreversible investments in learning about brand one. In order to do persuade them to make the same sort of investment in learning about brand two, that brand must offer some advantage over brand one. The only way to do that here is to charge a lower price.

In general, using (3.2) and (3.4), brand two is tried and used by non-users of brand one for whom

\[
p_2 \leq (1-\tau)v \quad \text{and} \quad v \leq p_1 / (1-\tau), \tag{3.7a}
\]

and it is tried and used thereafter by customers of brand one for whom

\[
p_2 \leq p_1 - \tau v \quad \text{and} \quad v \geq p_1 / (1-\tau). \tag{3.7b}
\]
Note that consumers with large values of \( v \) are least likely to try brand two, even though they were most likely to try brand one. This switch occurs because their high valuation of brand one gives them a high opportunity cost of trying brand two once they've learned that brand one works.

Conditions (3.7) are depicted in Figure 1. Before brand two's entry, brand one has sales of \( F[p_1/(1-\tau)] \), from (3.5). Upon entry, brand two is tried by those consumers with \( v \)'s located between the intersections of the \( p = p_2 \) line and the solid kinked schedule. As the Figure is drawn, brand two has sales of \( \{F[p_2/(1-\tau)] - F[p(1-p_2)/\tau]\} \), so that brand one has lost sales of \( \{F[p_1/(1-\tau)] - F[(p_1-p_2)/\tau]\} \). Note that from (3.7b), brand two does not capture any of brand one's customers unless \( p_2 < p_1(1-2\tau)/(1-\tau) \).

If experience with brand one causes consumers to be more optimistic about brand two, the first inequality in (3.7b) would involve some parameter \( \tau' \) less than \( \tau \) in place of \( \tau \). In terms of Figure 1, this would mean replacing the \( (p = p_1-\tau v) \) locus with a flatter line having the same intercept and slope \((-\tau')\). Small changes in \( \tau' \) would then clearly give rise to small changes in brand two's prospects, as asserted above. As long as \( \tau' \) is positive, brand two must charge a discrete amount less than brand one in order to capture any of the latter's customers.\(^{27}\)

As they stand, conditions (3.7) imply the following demand function for brand two:

\[
q_2 = F[p_2/(1-\tau)] - F[p_1/(1-\tau)], \quad (1-2\tau)p_1/(1-\tau) \leq p_2 \leq p_1, \\
(3.8a) \\
q_2 = F[p_2/(1-\tau)] - F[(p_1-p_2)/\tau], \quad p_1-\tau v \leq p_2 \leq (1-2\tau)p_1/(1-\tau), \\
(3.8b)
\]
Figure 1 Demand for Brand Two from Conditions (3.7)
\[ q_2 = F[p_2/(1-\tau)], \quad p_2 \leq p_1 - \tau V. \] \hspace{1cm} (3.8c)

Only if \( p_2 \) is low enough to induce all of brand one's customers, along with some non-users, to invest in trial does brand two face the same demand curve as brand one did. Otherwise, brand two's demand curve is below brand one's, no matter what price brand one changes.

In order to show this difference, the demand function (3.8) is graphed in Figure 2 under the uniformity assumptions that gave rise to (3.6). The right-most dotted line is (3.6), the demand function that initially faced brand one. Since brand two can make positive sales if it undercuts brand one by a tiny amount, its entry cannot be deterred under our assumptions if unit costs are constant and brand one is generating excess profit. If there are economies of scale, however, so that the long-run average cost curve looks like the curve labeled AC in Figure 2, brand one can enjoy excess profits and not worry about entry. In general, if pre-entry excess profits are possible for brand one and if there are any scale economies, the geometry of Figure 2 makes it clear that there is a barrier to entry in Bain's sense after brand one is established. This holds \textit{a fortiori} if potential entrants are less optimistic than assumed here.

A number of comments about this barrier are in order. First, the second brand's handicap vanishes if \( \tau = 0 \) and, roughly, increases with \( \tau \). Since \( \tau \) increases with \( \phi \) and \( \pi \), which measure the risk of trying a new brand, and with \( r \), which varies inversely with purchase frequency, this model is broadly consistent with Bain's observations about risk and purchase frequency quoted above. Second, in this model consumers acquire information about new brands only by trying them or, perhaps, by exposure to advertising
Figure 2  Demand for Brands One and Two under Uniformity
that is not fully credible. If information about new brands is valuable enough, however, one can expect consumers to use word-of-mouth and other channels to seek it. In this model, it is easy to show that if the expected value of perfect information about a new brand is positive, multiplying \( p \), \( s \), and \( v \) by any positive scalar \( k \) also multiplies that value by \( k \). This means that information is worth more for products with a higher unit value, all else equal. This in turn suggests that the barrier modeled here may be most important, all else equal, for low-priced products, since consumer information-seeking may be least important there.

Third, it is interesting to note that a barrier to entry arises in this framework only through the interaction of late entrants' demand disadvantage and brand-specific scale economies. Bain (1956, ch. 4) was concerned about economies of scale in marketing in this context, but the present analysis implies that scale economies in production are also relevant. Fourth, it is worth noting explicitly that advertising plays no role in creating the entry barrier modeled here. This does not prove that advertising never creates product differentiation advantages that established sellers can use to protect monopoly profits from entry. It does suggest, at least, that Bain's (1956, p. 143) finding that "advertising per se is not necessarily the main or most important key to the product-differentiation problem" deserves to be taken more seriously than it generally has been.

The model of this section is consistent with marketers' conventional wisdom that so-called "me-too" brands, brands that promise quality identical to established brands but sell at lower prices, are unlikely to be very successful in a set of consumer markets. This lends some theoretical
respectability to the apparent fact that economists' most commonly-assumed form of market rivalry, price-cutting with no quality change, is very rare in some such markets. In situations where the model developed here captures important aspects of market operation, the most natural and common form of rivalry must involve developing products that differ from existing ones, so that one's wares look like brand one, not brand two. Models of market interaction that focus on price may be utterly misleading in these cases, even when there are no technical obstacles to duplicating rivals' products.

In the model considered here, all parties are rational. Brand one's advantage does not arise because its customers are lazy or stupid or because they are easily misled by advertising. This might lead one to suspect that an equilibrium in which brand one earns excess profits in perpetuity can have some sort of optimality properties. In fact, the brand one's of this world have taken risks and been innovative. I see nothing in this framework that might link the size of the rewards from early entry to the social value of the corresponding innovation, however. The dependence of the barrier's height on the importance of scale economies makes it especially doubtful that this kind of permanent monopoly situation has any particularly attractive optimality properties. On the other hand, it is not obvious what sort of public policy would reliably enhance efficiency, since consumers are generally rational to worry about the quality of products with which they lack experience. More work is needed, as usual.

4. Advertising and Consumer Behavior

When they participate as buyers in modern markets, consumers often
make purchase decisions under uncertainty about complex products. In many modern markets, sellers spend a substantial fraction of their revenue on advertising, attempting to affect those decisions. In such cases, it is reasonable to assume that on average, advertising spending has some impact on firm or brand sales. A deeper set of issues involve the effect of these expenditures on market structure. Hypotheses about these latter effects must be based on explicit or implicit hypotheses about consumer behavior, since these effects must derive from shifts in the structure of consumer demand. I hope to show in this concluding Section that widespread failure to recognize the implications of this straightforward observation and unwillingness to grapple with the rich detail of market operation has limited economists' understanding of advertising and thus impaired our ability to analyze modern markets.

In mainstream economic theory, households are almost always assumed to solve dynamic optimization problems under uncertainty correctly, no matter how hard the solutions are to characterize or compute. If the problem involved is relatively straightforward, as in the previous Section or in Butters' (1977) interesting model of informative advertising, this may be empirically plausible. When consumers must solve complex dynamic programming problems correctly in order to behave rationally, however, most non-economists and even some economists would doubt the value of imposing perfect rationality. This does not imply that rational behavior models cannot yield useful insights in complex situations, only that it may be misleading to take such models too seriously. One might, for instance, be willing to accept as a plausible hypothesis the prediction of Grossman, Kihlstrom, and Mirman (1977) that if larger purchases yield more useful information about product quality in a dynamic context,
consumers will buy more than they would if information had no value, without having much faith in the ability of their complex optimization model to make correct quantitative or detailed predictions about behavior. Those who have taught introductory statistics or have been persuaded by the experimental evidence reported by Kahneman and Tversky (1979) and others might suspect that consumer behavior could deviate systematically from optimality even in simple situations involving uncertainty. In any case, one cannot appeal to natural selection and talk about as if optimization for households as is commonly done for competitive firms.\footnote{28} I am not trying to argue that it is \textit{wrong} in any sense to make the assumption that consumers behave optimally at all times, just to suggest that this assumption lacks obvious empirical plausibility when real households are assumed to solve problems that economic theorists themselves find difficult to analyze.

Most discussions of advertising in industrial organization run no risk at all of taking the rationality assumption too seriously. Consumers are generally treated as responding more or less mechanically to advertising spending, not as intelligent agents consciously deciding how to react to the advertising they encounter. Sometimes, as in the work of Dixit and Norman (1979), it is explicitly assumed that advertising spending changes tastes in a predictable way, but it is more common to be less formal and less explicit about how consumers are being manipulated.\footnote{29} Again, I am not trying to argue that this view of consumer behavior is always wrong. But I would agree with its "Chicago School" critics that it misses important aspects of at least some situations.\footnote{30} Advertising sometimes does provide
information that clearly improves consumer decisions, if only by making them aware that new products exist. It seems odd to assert that consumers' tastes are changed by want ads or grocery store ads that stress price information, for instance. When the rational response to some particular advertisement or other market event is obvious to most economists, it seems a bit unreasonable to assume that most consumers will respond very differently.

Economists have generally left the detailed study of actual consumer decision-making to psychologists and marketers. In the marketing literature, especially, one encounters descriptions of consumer behavior intermediate between the two extremes just discussed. Consumers are treated as boundedly rational, in Herbert Simon's phrase. They are assumed to pursue reasonably stable objectives in a reasonably intelligent fashion, but not necessarily in the supremely rational fashion economists reflexively assume. As humans, consumers' effectiveness is sometimes limited by their finite abilities to receive, store, retrieve, and process information. Given their limits, they sometimes sensibly adopt rules of thumb to economize on decision-making time. Both introspection and considerable empirical work support this general view.

Williamson (1975, 1979) has emphasized the importance of recognizing bounded rationality in the study of institutional and contractual structure. He argues that it becomes important to economize on human rationality when complex uncertainty is an important feature of the relevant environment. Contracts and institutions can be expected to reflect this. In simple situations, where uncertainty is either of minor importance or easy to describe, one loses little by ignoring the boundedness of rationality. On very similar reasoning, one might expect the bounds on consumer rationality
to be of negligible importance in some market situations, while in others they might be the prime determinants of observed behavior. Thus if optimization is possible for reasonably intelligent people, optimal behavior is probably a good prediction. But if consumers face a problem for which the exact solution is either difficult or impossible to compute, it might be important to know what sort of heuristics and rules of thumb humans tend to employ in related contexts if one wishes to predict actual behavior. The non-economic literature on consumer behavior thus suggests strongly that the most appropriate shorthand description of consumer behavior in general, and of consumer reaction to advertising in particular, might depend crucially on the details of the decision problem posed for consumers by the market considered.

This in turn suggests that the effects of advertising on market structure and competition can be expected to vary considerably, depending on both the initial market situation and the nature of the advertising involved. There is some empirical support for this hypothesis in the industrial organization literature. Benham (1972), for instance, finds that restrictions on eyeglass sellers' advertising tend to raise the price of eyeglasses, while Scherer (1980, pp. 380-3) documents the extraordinary price premia commanded by heavily nationally advertised brands in some markets. Boyer (1974) finds advertising intensity to be negatively associated with profitability in a sample of trade and service industries. Porter (1974) finds no association between advertising and profitability across markets in which retailers serve as an important source of consumer information. He finds a strong positive association in a sample of what he calls "convenience goods", which are usually purchased without consulting
a retailer. In later work, Porter (1976) finds suggestions that different media have different competitive effects. Finally, Lambin's (1976) detailed econometric analysis of sixteen product classes in eight European countries supports the general impression of diversity.

Almost all the substantial literature recently surveyed by Comanor and Wilson (1979) is concerned with choosing between two simple generalizations: advertising is always pro-competitive, or advertising is always anti-competitive. If the nature of advertising and of consumer response to it varies across markets in response to more fundamental market attributes as the discussion above implies, both of these are almost surely wrong. Adequate analysis of the role of advertising in any individual modern market likely requires the development or deployment of a model reflecting the key features of that market. In many models that prove useful for market analysis, consumers may be assumed supremely rational as a good approximation. In at least some situations, however, it may be necessary to deal explicitly with bounds on consumer rationality in order to explain observed behavior adequately. Sufficient analysis of actual modern markets, using the modern theoretical and econometric tools of the new industrial organization, can be expected to teach us something about the conditions under which various forms of advertising have desirable or undesirable effects on market structure and operation.
REFERENCES


Demsetz, H., Accounting for Advertising as a Barrier to Entry. Journal of Business, 52 (July 1979), 345-360.


Eaton, B.C., and Lipsey, R.G., Exit Barriers are Entry Barriers: The Durability of Capital as a Barrier to Entry. Mimeographed, Yale University, February 1980.


Kaldor, N., Mrs. Robinson's "Economics of Imperfect Competition." *Economica*, 1 (August 1934), 335-341.


Runyon, K.E., Consumer Behavior and the Practice of Marketing. Columbus, Ohio: Merrill, 1977.

Salop, S.C., Monopolistic Competition with Outside Goods. *Bell Journal of Economics*, 10 (Spring 1979), 141-156. (a)

Salop, S.C., Strategic Entry Deterrence. *American Economic Review*, 69 (May 1979), 335-338. (b)


Schmalensee, R., Advertising and Profitability: Further Implications of the Null Hypothesis. *Journal of Industrial Economics*, 25 (September 1976), 45-54. (a)


Schmalensee, R., Economies of Scale and Barriers to Entry. Sloan School Working Paper No. 1130-80, M.I.T., June 1980. (a)

Schmalensee, R., Product Differentiation Advantages of Pioneering Brands. In preparation, 1980. (b)


FOOTNOTES

1. Grether (1970) and Weiss (1971) discuss this early history. Mason (1939) is still worth reading on problems and methods.

2. Weiss (1971, p. 362). This history doesn't describe the evolution of the Chicago School very well. Perhaps because of its early resistance to the Chamberlinian revolution in microeconomic theory, Chicago seems never to have lost faith in the ability of Marshallian price theory to explain the central features of observed market behavior. For whatever reason, formal theory is much more important and generalizations from cross-section evidence are much less important in Stigler (1968) than in Bain (1968).

3. For instance, most early cross-section studies find a positive correlation between concentration and profitability and interpret it as reflecting a positive causal relation between concentration and ease of collusion. In an important revisionist study, Peltzman (1977) argues that this correlation in fact mainly reflects a positive relation between cost-reducing innovative activity and increases in concentration. That is, Peltzman argues that profits are high mainly where costs of infra-marginal innovative firms with large market shares are relatively low. Careful readers of his study will come to appreciate the difficulty of using available data to discriminate among alternative models of these sorts.


5. See also Salop (1979b), Eaton and Lipsey (1980), and Schmalensee (1980a). Scherer (1980, ch. 8) provides a good discussion of the limit-pricing model and its critics. Very recent work by Kreps and Wilson (1980) and others they cite seems likely to deepen our understanding of predatory
pricing and other strategic responses to entry.

6. In general terms, Bain (1968, ch. 5) and Stigler (1968, pp. 29-36) agree on this point. Demsetz (1973) apparently does not join them.

7. Hannah and Kay (1977, ch. 4) provide a useful discussion of functions that have been employed to define and measure concentration; see also Scherer (1980, pp. 56-59) and the references he cites.

8. It is generally but not universally accepted that market concentration is of interest mainly for its predictive value. Dansby and Willig (1979) propose concentration measures that reflect the returns from government intervention under certain assumptions about the costs thereof, while Blackorby, Donaldson, and Weymark (1979) argue that concentration measures should be derived as part of a market performance index concerned directly with unequal firm shares.

9. Hannah and Kay (1977, pp. 11-12) have this same derivation. For other examples of the sort of approach taken here, see Dansby and Willig (1979) and Encaoua and Jacquemin (1980).

10. In retrospect, at least, some of Sraffa's (1926, pp. 182-197) language, which is cited approvingly by Hotelling (1929), is very suggestive of the spatial model.

11. Parsons and Schultz (1976, ch. 7) provide a useful survey.

12. Discussions of alternative techniques and lists of references are provided by Hauser and Koppelman (1979) and Huber and Holbrook (1979). Clarke (1978) has explored the use of perceptual mapping data in econometric work.

13. Urban, Johnson, and Brudnick (1979) discuss the literature on the first of these approaches and present the second technique in detail.

14. Bernhardt and Mackensie (1968) propose a rather different approach to this problem of structural measurement.
15. Bell, Keeney, and Little (1975) and Barnett (1976) present axiomatic derivations of this basic functional form. Models of this sort are analyzed in detail in Schmalensee (1976a, 1976b).

16. This is not the most natural spatial model of non-price competition. In the standard one-dimensional spatial model of price competition, each firm has only one rival for each half its market; see Salop (1979a) for a good discussion. Such a framework is a bit harder to work with here, however, and its use would not add much insight. Hannah and Kay (1977, pp. 12-15) have an interesting analysis of price-setting in the standard model from which a multiple of the H index of concentration emerges, much as it does below.

17. Models of this sort and more general implications of related curvature properties are explored in Schmalensee (1976b, 1977b, 1978a).

18. If one is very lucky about the distribution of the \(a_i\) in some historical sample, for instance, one might be able to obtain useable estimates of the relative magnitudes of these cross-derivatives from ordinary linear regression, even though the underlying structure must be nonlinear. White (1980) provides a very careful analysis of this approach.

19. Both Fisher (1979) and von Weiszacker (1979) would apparently call such an advantage an entry barrier only if it would be socially preferable to remove it. I prefer Bain's usage, mainly because I see little to be gained from a forced merger of the difficult positive and normative issues in this area.

20. An argument is given in Schmalensee (1978b).

21. That debate is surveyed by Comanor and Wilson (1979), Demsetz (1979), and Scherer (1980, chs. 9 and 14). It is perhaps worth mentioning that
the U.S. ready-to-eat breakfast cereal industry has generally had both very heavy advertising relative to sales and very high measured profitability. While entry of new firms has been rare, entry of new brands sold by established firms has been quite common. Since most cereal advertising deals with the attributes of individual brands, if consumer goods advertising generally serves to inhibit new entry, it should have served to inhibit the introduction of new brands of cereal. As it apparently did not do so, the meaning of at least this one sample point in cross-section studies would seem to be in doubt. [Schmalensee (1978b)]

22. Bain (1956, ch. 4) mentions a number of other factors, including economies of scale in advertising and pre-emption of scarce retail outlets, that are not directly addressed in the present analysis. On scale economies, see the references cited in footnote 21 and Schmalensee (1980a).

23. Peckham (1966) makes essentially the same point.

24. This extends the model in the Appendix to Schmalensee (1979) and is a special case of the model analyzed in Schmalensee (1980b). Bond and Lean (1979) present a model that has the same sort of implications as the one in the text, but they focus on advertising and do not treat consumers as (necessarily) rational. A broadly similar set up is considered by von Weizsacker (1979, ch. 5), but his model is basically competitive and the one in the text is not.

25. The general status of the rationality assumption in this sort of context is discussed briefly in Section 4, below.

26. On the limitations of Nelson's (1974) analysis, see Schmalensee (1978a) and Boyer, Kihlstrom, and Laffont (1978). Note that after the first period, all those buying the first brand are enjoying positive surplus. The first brand could raise its price to $p_1/(1-r)$ and lose no sales. The consequences
of allowing pricing of this sort, which is commonly observed in new brand introduction, are explored in Schmalensee (1980b).

27. In Schmalensee (1980b), brand two is allowed to charge a low price upon entry in order to induce trial and then to undercut brand one by a tiny amount. Brand two's disadvantage is smaller, roughly speaking, in this more complicated model, but it does not vanish.

28. Several of the points made here are developed more fully in Schmalensee (1975).

29. See, for instance, the discussions in Comanor and Wilson (1967, 1979), Mann (1974), and Scherer (1980, ch. 14). Kotowitz and Mathewson (1979b) develop a dynamic model of "persuasive" advertising, which is assumed mechanically to alter consumers' perceptions. See also their (1979a) model of "informative" advertising.

30. Telser (1964) provides an important early statement of this "advertising is information" viewpoint; see also Brozen (1974), Ferguson (1974), Nelson (1974), and Demsetz (1979).

31. Simon (1978) provides an excellent discussion of rationality in and out of economics that is directly relevant here.

32. This view of consumer behavior is presented most explicitly by Bettman (1979), who also surveys the relevant evidence. Other useful references in this literature include Howard (1977), Runyon (1977), and Engle, Blackwell, and Kollat (1978).

33. Howard (1977, p. 151) presents an interesting classification of products similar in spirit but apparently more general than that used by Porter (1974).