Demand for variety may arise from a taste for diversity in individual consumption and/or from diversity in tastes even when each consumer chooses a single variant. The full degree of variety potentially demanded will not, in general, be supplied because scale economies (even to a small degree) mean that the potential welfare or revenue gain from greater variety must be balanced against the lower unit production costs with fewer variants. The economics of product variety consists essentially in analysing the effect of this balance in different situations, and comparing the degree of product variety for different market structures with each other and with the optimum. The survey commences with the work on market structures with single product firms (generalized monopolistic competition), tracing modern developments in both the Chamberlinian and Hotelling traditions. The latter has been particularly fruitful, due to the expansion of the original locational ideas into virtual spaces in product characteristics. The emphasis in recent work on product variety has been on multiproduct firms in both monopolistic and oligopolistic structures, including strategic market preemption. Although most work has been in a full information context, there have been advances in product variety under imperfect information (either by consumers to properties of the firms' products or firms as to consumers' demands).

(Product Differentiation; Product Variety; Monopolistic Competition; Oligopoly)

1. Introduction

This paper attempts to survey the problem of product variety from the economist's point of view. The term product variety is being used here to refer to the number of variants within a specific product group, corresponding broadly to the number of "brands" as the term is used in the marketing literature or the number of "models" in consumer durable markets. Pure conglomeration, in which firms expand the number of product groups but not the variety within a group, is not considered.

Implicit or explicit theories of product variety are to be found at various points in economics, management theory, and marketing. In each context, the view of product variety and the kind of questions being asked about it are rather different from those in others. In economic theory, for example, the emphasis placed by a general equilibrium theorist is on the socially optimal degree of product variety, given some welfare criterion, and on the relationship of the degree of variety generated by the market to this. Economists primarily interested in market structure theory will tend to emphasize competitive relationships, product differentiation and product variety as decision variables for the firm, and the types of market equilibria that result.

An attempt is made in this survey to do two main things. One is to give a broad overall
picture of how analytical economists have approached the study of product variety, the
other is to touch base with the large number of contributions that have been made to
specific problems or to the development of particular models. The study is confined to
models in which product variety is the outcome of an optimizing, strategic, or competitive
process, determined by the structure of the system and its parameters. Purely descriptive
studies as to the actual degree of product variety in specific markets are not considered,
although some reference is made to empirical studies relevant to particular analytical
models. There is a considerable amount of ongoing work (especially in relation to oligopoly
models), but discussion has been confined to work already published. With the exception
of certain classic papers, almost all the work discussed has been published since 1975.

2. Overview

In a market economy, it is clear that variety within a product group will persist only
if one or more of the following is true:

• Each individual consumer seeks variety in his own consumption.
• Different consumers want different variants because tastes vary.
• Individual firms can increase profits by producing a variety of models.
• Firms can increase profits by differentiating their products from those of their com-

Note that none of the above (nor any combination of them) is sufficient to guarantee
variety in the market.

Thus there are four contexts in which to consider questions concerning the degree of
product variety:

1. The individual consumer. How many of the available variants within a single product
group will the individual choose? What determines the choice?
2. The individual firm. What degree of product variety is it most profitable for the
firm to offer in a given competitive situation?
3. Market equilibrium. What degree of product variety will result from the operation
of the market within a particular competitive structure?
4. The social optimum. What degree of product variety is optimal for society on some
criterion? How is this related to the market equilibrium?

Since the analysis of product variety often involves considering several of these questions
together, it seems useful to give a brief overview of each before proceeding to more
detailed discussion.

The Individual Consumer

The economist’s traditional model of consumer choice, based on strictly quasi-concave
preferences (smooth indifference contours strictly convex to the origin) and infinitely
divisible products, was really devised to describe broad choices between product groups
rather than within groups. It poses some problems when used to describe intra-group
choices.

Strict convexity implies an inherent preference for variety, since it asserts that, for
some range of prices, there is a combination of \( n \) goods that is preferred to any combination
of fewer goods which costs no more. This is reasonable in describing broad choice over
bundles of aggregate goods such as food and clothing. When applied literally to choice
over a group of similar but differentiated goods, it predicts that the individual would
consume every product in the group and every brand or variant of that product for some
range of relative prices.¹

¹ If the indifference contours are asymptotic to the origin, as often modelled, the individual will always
consume all available products.
The traditional preference structure is preserved in models using the "representative consumer", a hypothetical single consumer whose behavior becomes that of the market when magnified sufficiently. Such a consumer necessarily consumes all varieties actually sold. Although widely used in empirical demand studies and in one class of studies in product variety, such models have no well-defined place for product positioning and provide no basis for a theory of product choice and product design.

At the other extreme are the locational and locational analog models in which the individual consumer buys only a single product in the group, choosing that variant which best fits his purposes. The consumer is assumed to have a most preferred or "ideal" location of a good in a space of product characteristics, his choice between available goods being based on a balance of prices versus distance from the ideal. In many such models, ad hoc assumptions, such as a fixed inelastic demand for one unit of the chosen product, are made, but others use fully developed demand functions. The reasons why the consumer must choose only one of the goods is not always explicit, but can be taken to be a property of the goods themselves (such as indivisibility) rather than of preferences.

Because there are few, if any, real markets in which there are individuals who consume all available brands but there are many in which individuals do buy more than one, there have been some recent attempts to combine features of the two types of models described above, with each consumer choosing a subset of goods rather than a single good.

**The Individual Firm**

Serious analysis of product variety in multiproduct firms, in the sense of attempts to explain the motives for producing a particular number of products, is quite recent in the economics literature. The approaches to the subject so far can be divided into three types:

1. Those centered on the production side, emphasizing cost advantages from joint production or economies of scope.
2. Those centered on the demand side, emphasizing the balance between the increased revenue possible from a more varied product line and the loss of scale economies in the production of each variant.
3. Those centered on strategic considerations, especially preemption in the product space as a deterrent to entry.

**Market Equilibrium**

The potential degree of product variety in the market is limited to the smaller of (a) the maximum number of variants which consumers in the aggregate are willing to buy, or (b) the number of variants which the suppliers in the aggregate are willing and able to produce, given the technology.

While a finite number of market segments is commonly assumed in the marketing literature, economists' models of product variety have almost all assumed that the aggregate potential demand for variety was unlimited (or not limiting in the analysis), either because each individual was interested in unlimited variety (representative con-

---

2 Spence (1976), Dixit and Stiglitz (1977), for example.
3 Hotelling (1929), Salop (1979).
5 In a pure location model it is because one cannot be in two places at once.
6 Lancaster (1966) draws the distinction between "combinable" and "noncombinable" goods as properties of the consumption technology.
8 This material has been surveyed by Bailey and Friedlander (1982).
9 The work in this area has been primarily concerned with the optimal choice of product variety for a monopolist. Examples include Swan (1970), Lancaster (1979), Itoh (1983), Moorthy (1984).
sumer models) or because there was a sufficiently wide variation of individual tastes when each preferred a single ideal good.

Thus the degree of product variety has been perceived as limited only on the supply side. Almost all analysis has been of groups made up of single-product firms, so that the degree of product variety is determined simply by the number of firms, provided no two firms produce products perceived to be identical. This is true of both of the main classes of market variety models, representative consumer (neo-Chamberlinian) and location-analog or characteristics (neo-Hotelling) models, and of almost all other kinds.

The analysis of variety in the case of a multi-product monopoly, the most fully examined case of market equilibrium with multi-product firms, has strong analytical similarities to the problem of determining the socially optimum, discussed below.

**Optimal Variety**

The fundamental structure of all optimal variety problems, for the individual firm as well as society, is the interplay of two elements in the economy—the existence of a gain from variety and the existence of scale economies of some kind. If there are no economies of scale associated with individual product variants (in distribution as well as in production), then it is optimal to custom produce to everyone’s chosen specification. If there is no gain from variety and there are scale economies, then it is clearly optimal to produce only a single variant if those economies are unlimited, or only such variety as uses scale economies to the limit (all products at minimum average cost output). Most cases involve a balance of some variety against some scale economies, the solution depending on the preference properties of consumers, the scale properties in production and distribution, and the way in which the social welfare criterion is derived from individual preferences.

Different criteria and assumptions can lead to quite different conclusions. Chamberlin (1933) concluded, for example, that monopolistic competition would always lead to more products than socially optimal (a view that was almost universally held until the 1970’s), yet Dixit and Stiglitz (1977), using a simplified general equilibrium model that was wholly in the spirit of Chamberlin, concluded that the market would always give too few products. In other analyses, such as those of Spence (1976) and Lancaster (1979), the relationship could go either way depending on circumstances.

The optimal choice of product variety for a monopolist is based on a decision process rather similar to that in determining the social optimum, since a potential social gain from variety can usually be translated into higher monopoly revenue from a more varied product line, to be balanced against higher costs from smaller outputs of each variant.

3. **Monopolistic Competition**

The theory of product variety in economics began as an incidental byproduct of analysis primarily concerned with deviations from the competitive model in prices and numbers of firms. It has a dual lineage, with one branch descending from Chamberlin’s work on monopolistic competition (Chamberlin 1933), which has been a textbook staple. The other, tracing its ancestry to Hotelling’s model of spatial competition (Hotelling 1929), is somewhat less well known and is discussed in a later section.

Each Chamberlinian firm produces and sells a single product which is taken to be unique to itself, but the firm is nevertheless a member of a well-defined group or industry and not a monopolist in the classical sense. The group exists because its members share similar cost and production conditions and because there is a relatively high degree of demand substitutability among the products of its members. A key element in the Cham-

---

11 An example of the interplay in its simplest form is given in Meade (1974).
berlin model is that demand substitutability is evenly spread over the group, so that there are no close rivalries between particular pairs of firms or within small subgroups which would give rise to oligopolistic behavior. For a given number of firms, the equilibrium is Nash in prices with every firm assuming zero reaction by other firms to its own price changes. The number of firms, and thus the degree of product variety within the group, is determined by the result of free entry. Each entry introduces a new product into the group and, since this is assumed to be an equally good substitute for all existing products, the demand for each existing product falls a little and profits fall. Entry ceases when the demand curves for all products have fallen to the point where it is tangent to the average cost curve, with price equal to average cost and marginal revenue equal to marginal cost. Since the demand curve for the firm’s specific product is downward sloping, average cost is falling at the equilibrium and so the production of each firm is less than its minimum cost output. In the strictly symmetrical case used in the traditional exposition, with firms having identical cost structures and identically structured market shares, the degree of product variety (= number of firms) is primarily determined by:

1. The extent of scale economies in output—the smaller these, the smaller the minimum average cost output, the larger the equilibrium number of firms and the greater the number of products.

2. The “intensity” of product differentiation, in the sense of the substitutability of products within the group as seen by consumers. The more intense the differentiation (the smaller the elasticity of substitution between the goods), the steeper the slope of the firms’ demand curves, the higher the margin of price over marginal cost, the smaller the equilibrium output relative to minimum average cost, and thus the greater the number of firms and products.

Since Chamberlin and his immediate followers took the optimal degree of product variety to be the number of products that would enable output per firm to achieve the minimum average cost level (thus considering only the scale economies and ignoring the gain from variety), the market equilibrium was perceived as necessarily leading to more variety than was optimal. The more intense the differentiation, the greater the deviation from optimality would be. Chamberlin’s writings are much richer than suggested by the simple textbook exposition, but he was not able to give a satisfactory derivation of the demand for a newly introduced brand from a consumer choice process.

Modern Neo-Chamberlinian Models

The feature that distinguishes a model as “Neo-Chamberlinian” is that of dispersed product rivalry, with any price change or product innovation viewed as impacting more or less evenly over all products in the group with no close oligopolistic rivalry between any small subset of firms.

The analysis of Dixit and Stiglitz (1977) is in the spirit of the original Chamberlin analysis but is a fully defined general equilibrium model with demand properly derived from maximization of a defined utility function. It uses the representative consumer approach and simple cost functions, but the simplicity permits derivation of demand from underlying utility and explicit determination of how particular parameters affect both the market equilibrium and the socially optimal degree of product variety.

The utility function is taken to be separable between the group and all other goods. Within the group, all goods (and potential goods) are equal substitutes for each other in the simplest version, with constant elasticity of substitution.\(^{12}\) This has the property that, for given prices and income, the consumer is always better off spending \(1/n\)th of his group budget on each of \(n\) goods, than spending \(1/(n-1)\)th of the budget on each of \(n-1\) goods, implying an insatiable taste for variety.

\(^{12}\) If \(V\) is the subutility from the group, \(V(s - 1)/s = (\sum_{i=1}^{n} q_i^{(s-1)/s})\), where \(s > 1\) is the elasticity of substitution.
Firms are assumed to be identical except as to their products (only one per firm), with the same simple cost functions determined by a constant marginal cost plus a fixed cost. An important limitation on the Dixit-Stiglitz and other neo-Chamberlinian models is that firms make no product choice—it is as though each firm, as it enters the group, is assigned a product by random choice (without replacement) from an urn containing blueprints for all possible products. In particular, no mechanism is provided for firms to contemplate and then accept or reject the possibility of producing the same product as an existing firm.

As in the original Chamberlin analysis, each firm is assumed to act as though it has zero effect on any other single firm, operating with a perceived demand curve rather than the true one. Entry will occur until all firms have zero profit, determining the equilibrium value of the number of firms and thus of product varieties.

This equilibrium degree of product variety will be greater

1. The smaller the economies of scale (due here to the existence of fixed costs).
2. The less effective are goods in the group as substitutes for each other
3. The larger the market, measured by aggregate income, and the more important the group.

Note that, as the size of the market increases without limit, the number of goods does also, but the monopoly markup of price over marginal cost does not change. That is, the market does not converge to perfect competition but preserves its monopolistic competition structure.

Since consumers are identical and preferences are homothetic, the social welfare function can be treated as simply a magnified version of the individual utility. On this basis the model can be shown to imply that the socially optimal degree of product variety is greater than the variety generated by the working of the market, precisely the opposite result from the traditional one.

A neo-Chamberlinian model with the opposite welfare result from the Chamberlinian one is less paradoxical than it may seem—the reason is that the original Chamberlin analysis assumed there was no welfare gain from variety as such, while the Dixit-Stiglitz version implies unbounded gains from variety for every individual. Thus Chamberlin understated the benefits of variety and Dixit-Stiglitz overstates them.

**Neo-Chamberlinian Variations**

In Hart (1985a, b) and Perloff and Salop (1985), Chamberlin-type structures are generated without representative consumers or the necessity of anyone buying everything. Neither model is fully developed from a standard consumer choice mechanism, and so both contain important ad hoc assumptions.

Perloff and Salop assume that each individual has a vector of relative values (expressed in dollars per unit) which he places on each of the available brands, relative values that are invariant with respect to the quantities consumed. Given the prices, the consumer chooses his “best buy”—the brand for which the net surplus of his valuation over the price is greatest—and purchases only that. The vectors of relative values are randomly distributed over a large population so that, in aggregate, the demands for all brands are the same when prices are the same. As the price of one good rises, it becomes the best buy for fewer and fewer individuals, giving the demand function. The more “intense”

---

13 In this model, the perceived demand elasticity is \(-s\), so the firm will choose price at \(s/(s - 1)\) times marginal cost.

14 See Archibald et al. (1986) for some specific criticisms.
the preferences (the larger the gap between the value of the most preferred good and the next), the less elastic the demand.

The structure is Chamberlinian in the sense that every good competes symmetrically with every other. Firms produce a unique single product, so the degree of variety is determined by the outcome of the free entry process. Perloff and Salop do not relate the market equilibrium degree of variety to any optimal concept, so the results cannot be compared with those of, say, Dixit and Stiglitz.

Perhaps the most interesting result in this analysis is the attempt to examine the effects of imperfect information. If each consumer is aware of only $k$ brands, for example, then the equilibrium price is that for a $k$-brand market, even if the true number of goods is $n > k$. Although the authors do not make the point, it is obvious that a free entry equilibrium will then result in a greater overall degree of variety than in the full information case.

The Hart model has similarities to the Perloff-Salop model, but is somewhat more elaborate. The basic assumptions concerning consumer preferences are that
- in the overall group of $N$ potential goods there is a subgroup of $m$ goods which are relevant to a particular individual,
- all other goods in the group are useless to the individual (give zero utility),
- different individuals have different relevant subgroups, and all possible subgroups of $m$ goods are represented equally in the aggregate population.

In general, each consumer will buy up to $m$ different brands. If the number of firms is $n < N$ (as assumed always to be the case), some individuals will not have all their particular $m$ goods actually available. When an $(n+1)$th good is introduced, it will be a desired but hitherto missing good for consumers having all possible combinations of other goods in their sets, and thus the new good will substitute uniformly for all other goods in the aggregate, including goods not yet available. Thus there is a Chamberlinian demand structure in the aggregate, although individuals consume only a few of the brands.

Hart does not derive any propositions concerning the degree of variety from this general model, being primarily concerned with the existence of equilibrium and whether the system preserves its monopolistic competitive character as the size of the economy becomes very large (yes). In a more specialized version (Hart 1985b) in which the individual regards all brands in his subgroup of $m$ as perfect substitutes for each other and thus chooses only the best buy, a comparison is made between the market equilibrium degree of variety and the optimal degree. The conclusion is that the market might generate more brands than optimal, or it might generate less, depending on values of the various parameters.

Although these models can generate uniformly distributed substitutability without multiple brand consumption by individuals, they carry a built-in presumption that a new entrant might just as well choose its product at random, provided it was not being produced by an existing firm. Such models cannot provide a basis for the analysis of problems of brand positioning and product design, or of the degree of differentiation as opposed to the number of brands. The solution of problems like these requires an analysis with some concept of location (perhaps in an abstract space)—models that can be classified as neo-Hotelling rather than neo-Chamberlinian.

### 4. Using Locational Concepts

Hotelling's classic paper (Hotelling 1929) introduced the idea of firms competing on more than one level—on both price and location. The model introduced was that of a one dimensional space (Main Street in his basic example) in which firms could locate and sell products that were identical except as to the location of the sales outlet.

Potential buyers were assumed to be uniformly distributed in the space but identical otherwise, to be intending to buy a fixed quantity of the good, and to have a transport
cost that was a linear function of the distance between their location and that of the sale. Each buyer would assess the full cost of buying from each outlet (price plus cost of transport) and choose the least cost one. For a consumer located between two outlets, the choice would depend on both the relative locations and the relative prices. Since the consumers were identical except as to location, if any consumer chose outlet A, then so would all consumers located between him and the outlet. Thus the market for each firm or outlet was a connected segment of the space with the firm’s location somewhere in it. The edge of the market in the direction of another firm was defined by the location of the dividing customer, to whom the delivered cost from each firm was the same. If a firm raised its price, customers on and close to the fringes of the market would shift to the next firm in that direction. Since individual demands were inelastic, the total demand for a firm’s product was directly proportional to the width of its market. Firms were assumed to have identical costs, to have free entry, and to be able to choose and vary their location and price at will.

Hotelling examined the possible equilibrium for the duopoly case in the above structure. He concluded that the two firms would locate close to each other near the center (principle of minimum differentiation), but that conclusion, due essentially to boundary effects (as shown later), is much less important than the framework created for product differentiation theory using locational ideas.

In a large market without boundary effects (often depicted as a circular road as in Salop 1979), a symmetric equilibrium for the Hotelling model exists and has the property that market equilibrium gives more than the optimal number of firms, the optimality criterion being to minimize production plus transport cost while supplying everyone with his one unit.

The Characteristics Approach

While a theory of real spatial location is a topic of importance in itself, the expansion of the locational framework to cover “location” in non-spatial contexts has been largely responsible for the recent revival of interest in models of the Hotelling type. Although Hotelling himself suggested that the location results might be applied to products differentiated in nonspatial respects, there was no full formal structure for doing so prior to the development of the characteristics approach.

In this analysis, goods are perceived as bundles of “characteristics”. These have similarities to, but are somewhat more specific than, the “attributes” of marketing and psychology. The characteristics are taken to be objective and measurable in the basic version. Consumers’ preferences are assumed to be over collections of characteristics and not over collections of goods per se, the role of goods being analogous to inputs in a consumption process with the demand for goods derived from the demand for characteristics. Within this framework a large number of different situations can be modeled according to whether goods can or cannot be combined in consumption to give characteristics combinations different from those of either good separately, whether such combining is linear or not, whether the number of relevant characteristics is larger or smaller than the number of available goods containing them, and so on.

---

15 For a firm located between two other firms at distances \(d_1, d_2\) each selling at price \(\bar{p}\), demand is given by:

\[ q(p, d) = (\bar{p} - p)/t + (d_1 + d_2)/2 \]

where \(t\) is the unit transport cost.


The strict neo-Hotelling models\(^{18}\) of monopolistic competition are built, implicitly or explicitly, on the special case in which consumption is not combinable within the group of goods being considered and the consumer chooses only one of the available brands. In the simplest versions the goods group is defined on the basis of two relevant characteristics so that the relative quantities of the two characteristics can be mapped into a one-dimensional spectrum analogous to the Hotelling “street”. The “location” of the firm is then given by the specification (characteristics mix) of the good it chooses to produce. The “location” of the consumer is a rather more complex concept, which can be expressed as his “ideal good” or “most preferred specification” and is the specification of the good that he would prefer if goods of all possible specifications were available on equal terms. The analog of “transport cost” is the diminution in the consumer’s valuation of a good as the “distance” between its specification and that of the consumer’s ideal good increases.

Although it is common for authors to assert that their location analysis can be directly converted into an analog in characteristics space,\(^{19}\) there are some considerable methodological problems involved. These are discussed in Lancaster (1979). Note that the simple symmetric Hotelling model described earlier, when converted by direct analog into characteristics space,\(^{20}\) gives the following results:

1. The market equilibrium number of products is larger, the larger the market, the less the degree of substitutability between products in the group (the analog of transport cost), and the smaller the fixed costs in production.
2. The market equilibrium generates more than the optimum number of products.

Qualitatively (although not necessarily quantitatively) these are essentially the same results as those of the basic Chamberlin model. Before proceeding to the more general neo-Hotelling models, however, it is necessary to comment on the fate of the principle of minimum differentiation since it has important implications for product differentiation theory.

**Product Clustering**

The basic Hotelling proposition was that the equilibrium competitive location for two firms in a linear market with distinct boundaries at each end and uniformly distributed customers would be as close together as possible without actually coinciding, and in the center of the market area. The basic argument was that, if they were close together but not at the center then the market would be larger for the firm on one side (East, for example) than on the other, since each firm’s market would extend out to the boundary on its own side. Then it would pay for the West firm to jump over to the East of its rival, continuing the process until the markets were equal on both sides. If the firms were not close together, then one firm might increase its market by moving closer to the other. When extended beyond pure geographical location, the proposition is concerned with the degree of differentiation rather than the number of goods, answering the question: if there are two firms in the market, will their best strategy be to choose products which are very similar or very different?\(^{21}\)

The Hotelling result is a fragile one and was the subject of debate in the 1930’s and 1940’s. Revival of the scrutiny in recent years\(^{22}\) has given the definitive answer to ap-

\(^{18}\) Note that not all location analog models could be called neo-Hotelling. See Capozza and Van Order (1982), for example.

\(^{19}\) Salop (1979).

\(^{20}\) But see Lancaster (1979) or Economides (1984) on the special problems created by inelastic demand and linear transport costs in this simple analog.

\(^{21}\) The idea was popular with political scientists, as explaining why both political parties move towards the center in a two-party system. See Downs (1957).

\(^{22}\) See, for example, Eaton and Lipsey (1975), Shaked (1975), D’Aspremont et al. (1979), Lancaster (1979), Salop (1979), Novshek (1980), Graitson (1982), Economides (1984), de Palma et al. (1985), Neven (1985).
plicability of the result: minimum differentiation is not the general outcome to be expected from spatial oligopoly but is a special result.

It was shown early (Lerner and Singer 1937) that arguments that may hold for two firms do not necessarily hold for three. If a third firm enters, it will be optimal for it to move close to one of the existing firms. But then there will be a firm flanked by two others with almost no market. It will move to the outside, leaving a new inside firm, and so on, with no equilibrium. Eaton and Lipsey (1975) examined the $n$-firm case in some detail and showed that there could be no equilibrium with three firms, but there could be equilibria with more than three, typically involving pairings of firms at certain locations but with separation between the pairs. Expansion of the space dimension from one to two does not remove the three-firm equilibrium problem (Shaked 1975).

In D’Aspremont et al. (1979), it was shown that a two-firm Nash equilibrium in prices might not exist if the firms were located close together, since either firm could attempt to gain the other’s market by undercutting. Closeness increases market width, but at the expense of greater sensitivity to price competition. Thus the attempt to find a locational equilibrium by moving close together would destabilize an existing price equilibrium. Shaked and Sutton (1982) argue that a three-stage process (entry, product quality, then price) will result in differentiation, not clustering.

Hotelling assumed inelastic individual demands, so that a firm’s sales were determined only by the width of its market area and were independent of its location within that area—hence the location very close to one end of the area as implied by minimum differentiation causes no revenue loss. If the quantity purchased is sensitive to distance in real or characteristics space, the outcome can be very much changed, and there is then an incentive for the firms to separate so that each is closer to the center of its own market area. See Salop (1979), Lancaster (1979), Graitson (1982), Economides (1984), for this and also for the case in which there are no boundary effects and the centrist forces vanish.

Recent location or location-analog models have tended to embody assumptions which differ from Hotelling’s, particularly in removing boundary effects (by circular or infinitely long markets) and in taking demand to be sensitive to distance between the location of the good and the consumer, in either real space or some goods spectrum. Such models generate equilibria in which firms are evenly spaced and not clustered at all. For examples see Lancaster (1979), Salop (1979).

It is still possible to have clustering effects without the highly special Hotelling assumptions. The case so obvious that it is often forgotten is that in which consumers are clustered geographically or consumer preferences are clustered in characteristics space, instead of being uniformly distributed as most of the theoretical models assume. Clustering may occur in geographical space if there is also differentiation between the goods in characteristics at some other level (de Palma, et al. 1985), and there are a variety of search cost and other imperfect information arguments that can explain many types of clustering, particularly the geographical clustering of stores. Studies designed to test for clustering in characteristics space, notably Shaw (1982) and Swann (1985) have tended to find it.

**Locational Analog Models**

These are models in which locational or geographical space is replaced by a virtual space of goods or their characteristics. They are variously known as “locational analog models”, “Neo-Hotelling models”, “characteristics models”, “Lancastrian models” or “address models”.24

---

23 Hotelling assumed firms chose prices first, then locations.

24 The term “address” is an attempt to generalize the idea of location to nongeographical contexts. See Archibald et al. (1986).
The most elaborate of the earlier models in this category is that in Lancaster (1979), and this is representative of the approach. In this analysis, characteristics are taken to be measurable and a good is defined by its "specification", the bundle of characteristics contained in one unit of the good. It is assumed that all goods in the group under consideration have the same characteristics but in different proportions, and that the set of characteristics relevant for both manufacturing and consumer choices is a small subset of all the actual physical characteristics. The sharing of common characteristics is what defines the group. For the basic model it is assumed that a subset of two characteristics ("speed" and "size", for example, or "power" and "fidelity") is sufficient, so that potential combinations of characteristics can be represented along a line spectrum, giving the analog to the Hotelling analysis.

Each consumer is presumed to have a "most preferred" (or "ideal") goods specification. If his most preferred good is not available, it requires $h(x)$ units of a good at distance $x$ from the ideal to be equivalent to a unit of ideal good, where $h(0) = 1$. The function $h(x)$ is referred to as the "compensating function", since it shows how distance from the ideal can be compensated for by increased quantity. It is assumed throughout that all individuals have the same compensating function, individual differences being confined to their choice of ideal goods.

The model is based on a two-sector economy with a differentiated sector and a single undifferentiated outside good. Within the differentiated sector, products are taken to be noncombinable. Combinable products are those that can readily be mixed or consumed together to give characteristics proportions between those of the components (sugar and lemon juice). Half a Mercedes combined with half a Hyundai does not, however, give the characteristics of a mid-level automobile—automobiles are noncombinable. Due to the noncombinability, each consumer chooses only one good in the group, the choice based on the relationship between the price and the divergence from ideal specification of each good actually available.

The utility function is modeled as being of constant elasticity of substitution (CES) form in two subutilities, one from the outside good, the other from the group, with substitution elasticity $\sigma$. The individual demand function which is derived has the expected properties, demand decreasing in price and in distance from the ideal (provided $\sigma > 1$, as assumed). All consumers are considered to have the same form of utility function, but express their individuality in their choice of ideal specification.

Aggregate demand for a good is the sum of all individual demands from all consumers within the market area. Its properties depend on those of the compensating function (how the utility of a good varies with its deviation from the consumer's ideal specification), the individual demand functions, and the distribution of consumers over the space. No simple expression can be written for the aggregate demand function even when consumers are identical and uniformly distributed over the space. It can be shown that the elasticity of demand is never less than $\sigma$ and increases with increasing proximity to other goods in characteristics space, approaching infinite (perfect substitutability) as goods become very close.

If there are no boundary effects, firms have identical cost structures, and consumers are uniformly distributed, it can be shown that there will be a symmetrical monopolistic competition equilibrium in which firms will choose to produce products different from those of other firms and in which the firm's product choice is fully explained as an optimal choice. The equilibrium degree of product variety generated by the market is then, as in all monopolistic competition models, determined by the number of firms at which excess profit per firm drops to zero.

Not all treatments of monopolistic competition in characteristics space assume noncombinable consumption or a neo-Hotelling structure. See Archibald and Rosenbluth (1975), Leland (1977).
The equilibrium product variety will be greater
1. The smaller the economies of scale.
2. The lower the substitutability between group goods and outside goods.
3. The more important the group in the economy (share of total expenditure).

These are somewhat similar to the properties of the neo-Chamberlinian models except for one very important point. Whereas a larger market, measured by total consumer expenditure, unambiguously predicts more variety in Chamberlinian models, it does not necessarily do so here. This is because there is a distinction in this case between the width of the market (the degree of dispersion of preferences) and the depth of the market (density of consumer purchasing power at each location). It is possible for the market with the larger consumer expenditure to have less variety because it may have consumers with less diverse preferences but more of them.

The relationship between the market equilibrium and the optimum is more complex and more difficult to establish in this model than in simpler ones. In terms of the criterion most similar to general usage in the monopolistic competition literature, the market will generate more than the optimal degree of variety under most circumstances. It is possible, however, that the market might produce too few products if the degree of substitutability between group goods and outside goods is relatively high.

Although the market in question is that of a multi-product oligopoly, we might note Scherer’s (1979) attempt to determine empirically whether the degree of product variety in the ready-to-eat breakfast cereals industry was more than socially optimal or not. His approach was to estimate the welfare gain from each new variety (capitalized into present value form) and compare it with the launching costs. Scherer’s conclusion was that the number of varieties was more than socially optimal, although the problems in reaching any firm conclusion were made clear. Wildman (1984) criticized Scherer’s method and showed that it understated the socially optimal number of varieties, leaving the final results more uncertain.

**Dimensionality of the Product Space**

The structure of the market equilibrium and the optimum choice of product variety for the firm depend, in the locational analog models, on the dimensions of the product space on which the market model is based. This fact is often obscured because the best known almost all such models have assumed that there were only two relevant characteristics in the group, so that all goods could be represented along a line measuring the ratio of the two characteristics, thus following Hotelling’s original location model. But the two-characteristic single-dimension model possesses special features that one cannot generalize to higher dimensions. The most important of these is the number of neighbors for any good. In the two-characteristic case this is always just two (one to the left and one to the right in the usual line diagram), and the firm’s market area is always a simple line segment with two boundary points. If there are many characteristics, however, the number of potential neighbors escalates. In the three characteristics case, a product may have many neighbors, each in a different direction. In particular, one product could have any number of equidistant neighbors if the latter were positioned around a circle with center on the former or, with three products only, all could be equidistant from each other if arranged as the vertices of an equilateral triangle. Indeed, the locational model and the Chamberlinian model can coincide when there are at least \( n - 1 \) relevant characteristics for \( n \) different products. Because two dimensions (three characteristics) correspond to real geographic space, the results of pure locational analyses, as in Eaton and Lipsey (1976) can be drawn upon with suitable modifications. The most important result from the two-dimensional location models is that there may be multiple market equilibria with different arrangements of the firms in the space, different market area shapes (rectangles versus hexagons for example), different numbers of immediate neighbors (rivals).
for the firms, and different levels of profit per firm. It is possible for the same demand and cost conditions to generate different free entry ("zero profit") equilibria with different levels of product variety. A discussion of such cases together with other aspects of many-characteristic models is given in Lancaster (1982). As the number of relevant characteristics increases the configurations become very complex, with no locational models to draw upon.

**Variety in Product Quality**

The discussion up to this point has been concerned almost entirely with "horizontal" product differentiation, that is, with products which could be considered to be broadly of equivalent quality, even though different consumers prefer different variants because of their specific characteristics proportions. If one product has more of all characteristics than another, or is universally ranked as better (like a Rolls Royce relative to a cheap subcompact), then there is "vertical" product differentiation with products ranked in a universally accepted order of preference by product quality. The latter is generally similar to single-attribute analysis in marketing, while the former requires at least two attributes.

Since all consumers have the same preference ranking for the products in vertical differentiation, something other than different preferences is required to provide markets for different qualities. The most usual is to assume variations in income, as in Gabszewicz and Thisse (1979).

One problem to be solved is the relationship between the quality level that will be produced under competition in the single-product case, compared with the optimum or the competitive outcome, as in Mussa and Rosen (1978). The problem of interest here is the number of different qualities that will be offered by a multiproduct monopolist, a problem that has been tackled by Itoh (1983) and Moorthy (1984). As might be expected, the outcome depends very heavily on the way in which the demands for products of different quality are distributed over the population and on the cost of incorporating higher quality in a product.

Shaked and Sutton (1982) and others (unpublished) have considered the case in which the production cost was the same for all qualities. A monopolist will sell either the highest quality alone, or the highest quality and a lower quality (at a lower price) if the range of consumer incomes is varied enough. If entry is possible, the first firm will sell at the high quality, the second entrant at lower, and there will be, in general, room at most for two firms (and thus two quality levels). Simple models tend to reinforce an old proposition, that intermediate quality classes vanish at market equilibrium.

### 5. Multiproduct Firms

As noted earlier, there are three main potential influences on the firm's choice of product variety:

(a) The existence of interproduct economies on the production side.

(b) The potential for increasing demand by offering more variety.

(c) Use of product variety for strategic purposes.

Recent work approaching the problem of product variety in the output of a firm from the production side has placed the emphasis more on the degree of market straddling by large firms, especially in regulated sectors, and less on product variety as such. It is notable for the considerable progress that has been made in the formal analysis of the potential economies from simultaneous production of two or more goods, as in Panzar and Willig (1981). This work has been thoroughly surveyed by Bailey and Friedlaender (1982) is discussed extensively by Baumol, Panzar, and Willig (1982), so it is not proposed to deal further with this material here. Approaches that emphasize organizational, man-
agement or financial economies rather than that technical production economies throw more light on conglomerates than on product variety in the sense being used here.

On the demand side, the most fully developed work on the demand side has been for a firm having an uncontested monopoly over the whole goods group. The incentive for a monopolist to provide variety, even if there are no economies of scope, occurs when revenue can be increased by additional products either because representative individual consumers gain utility from variety and thus are willing to pay more for a varied collection or because preferences are heterogeneous and increased variety taps consumers in new segments of the market. The profitability of increased variety, given the potential revenue gain, is diminished by the existence of scale economies in the production of each variant and reinforced by any economies of scope. In general, a monopolist cannot capture all the potential consumer surplus without perfect price discrimination and so the revenue gain from variety to an ordinary nondiscriminating monopoly understates the social gain.

Thus one might expect the monopolist to produce less than optimal variety. Some of the early papers in this area, such as Swan (1970) and Lancaster (1975), concluded that the monopolist would produce optimum variety, but these used restricted demand formulations. Using the more complete demand specification of the companion monopolistic competition model, Lancaster (1979) showed the earlier results to be false in general—rather, that a monopolist would produce less than the optimal variety under most conditions. Moorthy (1984) obtains a similar result using a finite number of market segments instead of a demand continuum.

In the simplest linear Hotelling type model with inelastic demand, a monopolist will have the whole market even if it sells at only one point—which it will do if there are unlimited scale economies associated with each point of sale. In this case the monopolist will certainly produce less variety than would emerge from monopolistic competition. It is not difficult to show that the general relationship continues to hold (although the monopolist now provides more than one variety) if demand is inelastic up to a reservation price (Salop 1979). When demand is fully responsive to price and distance in product space, as in Lancaster (1979), the relationship is more difficult to prove, but still holds.

There have been few attempts at a theory of optimal product variety and positioning for multiproduct oligopolists although the oligopoly case is very important in real business situations. Brander and Eaton (1984) tackles the problem by considering two possible types of outcome for the duopoly case, one in which each firm has a complete segment of the market and the rivals have close substitutes only at the boundary between the segments, and the other in which the firms’ products are interlaced so that each product is flanked by rival products.

Using a somewhat specialized demand structure, both outcomes are shown to be potential Nash equilibria if there is simultaneous product selection, but the market segmentation outcome (which gives rise to higher prices and profits) will be the Nash equilibrium if the product selection is sequential. If there is a threat of further entry, however, the interlaced pattern can be better for the two incumbents than the segmented pattern.

26 Such as Williamson (1975).
27 See White (1977).
28 The exception being rather special conditions involving comparison with what is really a second best optimum, rather than a first best situation, and require that the elasticity substitution between the group and outside goods be low.
29 The same will be true for the simplest Dixit-Stiglitz model, if the elasticity of substitution between the group and outside goods is unity.
30 Similar ideas (with terms “split” and “interleaved”) appear in analyzing rivalry between home goods and imports under conditions of product differentiation in Lancaster (1984).
Strategic Considerations

This last result of the Brander and Eaton model is related to the work on preemptive product differentiation, the third of the approaches outlined at the beginning of the section. This work commenced with pure locational models as in Hay (1976), Prescott and Visscher (1977), Eaton and Lipsey (1979). See the discussion in Dorward (1982).

The argument in the simple case of the one-dimensional locational model with inelastic demand and without boundary effects (the nonclustering case) is that if outlets are evenly spaced at distance $d$ apart and all sell at the same price, each outlet will have revenue proportional to $d$. A new entrant will have to locate between two existing ones and thus will have revenue proportional only to $d/2$. If monopoly, common ownership, or collusion permits the incumbent(s) to choose $d$, it can be made large enough so that a profit can be made on market width $d$ but not so large that a loss can be avoided on a market width of $d/2$, and thus entry is inhibited. Analysis along similar lines appears in Bonanno (1987), who shows that the incumbent will tend to produce more than the minimal entry-deterrent variety, measured by the number of stores in his particular model.

Schmalensee (1978) applies the locational reasoning to product space and makes a specific application to the ready-to-eat breakfast cereals group, in which he argues that such preemptive brand proliferation occurs and gives a welfare loss. Since a monopolist without threat of potential entrants will generally produce too little variety, a preemptive increase in variety under threat of entry may merely be a move towards the optimum. It is obvious that preemptive strategy will not result in fewer products than under guaranteed monopoly (and almost always more) and that there will be fewer products than under monopolistic competition (else the strategy has failed), but the relationship to the optimum is unclear and is probably sensitive to the specification of the model.

6. Other Aspects of Product Variety

There is a growing literature on the relationship between international trade and product variety. Representative works would include Krugman (1970), Lancaster (1980, 1982), Helpman (1981), Horn (1984), Eaton and Kierzkowski (1984). This is regarded as a specialized area and not reported here in detail. The general conclusion from this work is that the opening of trade between hitherto isolated product-differentiated markets may increase or decrease world product variety. However, the variety available to consumers in each country will generally increase, even though the variety produced within each country will generally decrease. Stokey (1988) has explored the relationship between economic growth and product variety, using a learning model in conjunction with a characteristics approach.

Virtually all the work on the economic theory of product variety has been within a perfect information framework and, in location and location-analog models, for a uniform distribution of consumers. Even with these simplifications, much still remains to be done at the core of the analysis.

There have, however, been a few scattered attempts to address the problem of imperfect information in relation to product variety. Stahl (1982), for example, considers a two-level product differentiation model (firms produce different varieties and can vary location) with imperfect information as to the firm’s product characteristics, leading to a geographic search. His conclusion is that there will be geographical clustering to more than the optimal degree. Wolinsky (1984) considers a location-analog model in product space, with consumers who are imperfectly informed about the specifications of the goods. The central argument of this paper is that the optimal degree of variety is less

31 There is, of course, a very large literature on imperfect information in relation to prices, including prices in product differentiated markets.
than would be the case under perfect information, and thus that the market equilibrium gives too much variety when the market is large and/or the economies of scale are small.

There is a section of Lancaster (1979) concerned with imperfect information as to specification: the conclusion is that this will result in a softening of rivalry between adjacent firms, a greater emphasis on price competition as compared with product differentiation, and a structure closer to Chamberlinian competition. Imperfect information also receives some consideration in a section of Perloff and Salop (1985), where each consumer is assumed to be aware of only a subset of all the available goods. In Ireland (1985), the uncertainty is in the other direction—firms are uncertain about the demands for their products. Such uncertainty can lead to an equilibrium with either more or less product variety than is optimal, according to the parameters of the situation.

Conclusion

Although there is wide variety in the supply of models of product variety, there appear to be certain conclusions on which almost all models agree:

1. The degree of scale economies (in the production of each variant) is a major determinant of the degree of product variety, and an increase in scale economies reduces product variety for both monopolistic competition and multiproduct monopoly. The socially optimal degree of product variety also falls when economies of scale increase.

2. The "intensity" with which consumers view the differences between similar products (in those models in which there is a demand parameter capable of this interpretation) is also important. In all structures, the degree of variety will be less when similar products are viewed as satisfactory substitutes.

3. The degree of product variety increases with the "competitiveness" of the market, in the sense that the variety is greater under monopolistic competition than under monopoly, and greater under monopoly threatened by potential entry than under protected monopoly.

There are other aspects of product variety, however, on which different models reach different conclusions. For example, a larger market will be associated with greater variety in models based on the idea of the representative consumer, but not necessarily in locational analog models where there is a distinction between the breadth of the market (width of the spectrum of diverse tastes) and depth (density of consumers at each point). There is much disagreement on an important policy issue—whether particular market structures produce more or less variety than is optimal. The conclusion in this regard varies from model to model, and in the more complex models, from situation to situation. A fair statement, however, is that most of the models predict that the monopolistic competition equilibrium will give more than optimal variety under most circumstances, and that protected monopoly will give less variety than is optimal. There seems to be no clear cut answer to such a question as whether an oligopolistic structure of multiproduct firms, or a monopolist attempting to deter entry, will result in more or less than the optimal degree of variety.\(^32\)

Acknowledgement. The assistance of the National Science Foundation, Grant 8408905, is acknowledged.

\(^32\) This paper was received in November 1985 and has been with the author 34 months for 2 revisions. Accepted by Subrata K. Sen.

References


ECONOMICS OF PRODUCT VARIETY


