Chapter 12

PRODUCT DIFFERENTIATION

B. CURTIS EATON

Simon Fraser University

RICHARD G. LIPSEY

Queen's University and C.D. Howe Institute

Contents

Handbook of Industrial Organization, Volume I, Edited by R. Schmalensee and R.D. Willig

1. Introduction

In a loose sense, any set of commodities closely related in consumption and/or in production may be regarded as differentiated products. Close relation in consumption depends on consumers' tastes. Do consumers perceive two products to be close substitutes for each other? Close relation in production concerns economies of scope. Is there any cost economy in having two products produced by one firm rather than two? The same set of products may be closely related in both production and consumption, as with yellow and orange tennis balls, in the former but not the latter, as with size 12 and size 8 shoes, or in the latter but not the former, as with coffee and tea.

We follow most of the existing literature in dealing with industries producing a large number of products that are closely related in consumption, while ignoring the interesting issues that arise from the presence of scope economies. We also ignore issues arising out of consumption complimentarities, e.g. should IBM produce software as well as hardware?

1.1. The awkward facts

Elementary scientific methodology tells us that theories aspiring to empirical relevance must be consistent with the observed facts. For this reason, awkward facts are to be welcomed; indeed, the more awkward they are, the greater are the constraints that they place on our theorizing.

Seven of the most important awkward facts that are available to constrain theorizing about product differentiation are listed below.

(1) Many industries, including most that produce consumers' goods, produce a large number of similar but differentiated products. Observe, for example, the variety of cars and bicycles on the streets of any moderately sized city.

(2) The consumers' goods produced by different firms in the same industry are differentiated from each other so that two products produced by two different firms are rarely, if ever, identical. Consider, for example, the differences between the competing middle-priced cars produced by GM, Ford and Chrysler.

(3) The set of products made by the firms in any one industry is a small subset of the set of possible products. Consider, for example, all of the different cars that could be produced by marginally varying the characteristics of existing cars one way or the other – e.g. a bit more or less acceleration, or braking power, or fuel consumption.

(4) In most industries each firm produces a range of differentiated products; indeed, a typical pattern in consumers' goods industries is for a large number of differentiated products to be produced by quite a small number of firms. For example, most of the many soaps, cleansers and detergents on sale in the United States today are produced by two firms.

(5) Any one consumer purchases only a small subset of the products that are available from any one industry. For example, how many brands of toothpaste has the reader purchased in the recent past?

(6) Consumers perceive the differences among differentiated products to be real and there is often approximate agreement on which ones are, and are not, close substitutes. Consider how many loyal supporters there are for different brands of cigarettes, cars and cameras. Also consider how much agreement there is that different brands of low-tar cigarettes are closer substitutes for each other than for any high-tar cigarette, and that different types of subcompact cars are closer substitutes for each other than for any full-sized car.

(7) Tastes are revealed to vary among consumers because different consumers purchase different bundles of differentiated commodities and these differences cannot be fully accounted for by differences in their incomes. For example, look into different houses inhabited by people of roughly similar incomes and observe that while each has a car, a refrigerator, a TV, a hi-fi, a tape deck, a video recorder, a camera, a stove, and so on, each has a different mix of brands, styles and types of these generic products.

The literature on product differentiation can be seen as a search for answers to three basic questions that arise out of the above awkward facts. What are the processes that give rise to these facts? What are their positive implications? What can be said about their normative implications?

The models used to study these problems are standard in their broad outline, although not in many substantial details. Most employ equilibrium techniques and use comparative statics to derive their predictions. From among the possibilities open to them, firms are assumed to choose the alternatives that will maximize profits.

A complete model of product differentiation would specify (i) the set of possible products, (ii) the technology associated with each product, (iii) the tastes of consumers over the set of possible products, and (iv) an equilibrium concept. At any significant level of generality such a model seems intractable. Hence, most of the literature involves strong simplifying assumptions of one sort or another.

1.2. Technology

There is little debate over the cost aspects of relevant models. Of the assumptions that are typically made, some are needed to accommodate one or another of the awkward facts, while others are employed merely for analytical convenience. Awkward fact (3), that the number of produced products is a small subset of the number of possible products, would seem to have two main lines of possible explanation.

(i) *Demand side explanations*. Given linearly homogeneous production functions, the explanation must come from the demand side. The explanation would be that consumers wish to consume only the subset of products currently being produced. This could be understood in terms of a representative consumer, or in terms of consumers with tastes that differ but are concentrated on the subset of produced goods.

(ii) *Supply side explanations*. Now let there be a diversity of tastes such that every differentiated product would be demanded by some consumers if all were priced at minimum average cost. Given linearly homogeneous production functions, all possible products would then be produced. For example, if there were a continuum of consumers' preferences over some set of product characteristics, there would be a continuum of products produced to satisfy these tastes. In these circumstances the limitation on the number of produced products must come from the supply side. The explanation of awkward fact (3) is then provided by production non-convexities, which result from such things as product development costs, and indivisibilities of fixed capital, and which imply decreasing average costs over an initial range of output. This explanation is the one that we accept, along with most of the writers in the field.

These non-convexities are commonly captured by the assumption of a simple cost function with a fixed cost of entry and (for convenience) a constant marginal cost of production that may or may not be subject to a capacity constraint. (Of course the cost function may be product specific.)

1.3. Two approaches to consumers' preferences

A basic requirement on the demand side of the problem is to have a model that accommodates awkward fact (1) – many differentiated products are produced and consumed in the typical consumer good industry.

1.3.1. The address branch

One branch of the literature captures awkward fact (1) by positing a distribution of consumers' tastes over some continuous space of parameters describing the nature of products. Different consumers have different most preferred locations in this space and thus can be thought of as having different *addresses* in that space. Products are also defined by their addresses in the space, and this makes the set of all possible products infinite.

This approach follows Hotelling's (1929) seminal paper and we refer to it as the *address branch*. Tractability demands a relatively simple parameterization of tastes. A major issue here is whether the parameterization of preferences is sufficiently rich to approximate the real diversity of consumers' tastes.

1.3.2. The non-address branch

The other branch follows traditional value theory in assuming that consumers' preferences for differentiated goods are defined over a predetermined set of all possible goods, which set may be finite or countably infinite. For obvious reasons we call this the *goods-are-goods* or the *non-address* branch. Chamberlin's original vision of monopolistic competition lies in this branch of the literature. Once again, tractability requires parameterization of tastes.

Within this branch, there are two possible ways of accounting for the purchase of many differentiated commodities. The first assumes that the aggregated preferences for differentiated goods can be captured by the fiction of a representative consumer. Since the problem of aggregating tastes over diverse consumers is ignored, the approach is more flexible in some important respects than the address approach.

It should be clear, however, that this approach does not directly incorporate awkward facts (5), (6) and (7), which relate to differences among consumers. This raises the question: Can the representative consumer's utility function be derived from a model which does allow for these awkward facts?

The second approach is to assume differences in individual tastes and deduce aggregate behavior from individual motivations. All models that have taken this approach so far have used some variant of Chamberlin's symmetry assumption which, loosely interpreted, means that all products are in equal competition with all other products. (This is discussed in more detail below.)

In the next section we consider the non-address branch and in subsequent sections the address branch.

2. Non-address branch

2.1. The representative consumer approach

The seminal papers are Spence (1976a, 1976b) and Dixit and Stiglitz (1977). In these models there is a large number of possible products in the sector or industry of interest. Product demands arise from the utility maximizing decisions of a representative consumer with a strictly quasi-concave utility function:

$$u = U(y, x_1, \ldots, x_n),$$

where y is quantity of a composite commodity, which is produced under conditions of constant returns to scale, and x_i is quantity of the ith sectoral good. Cost functions, $C_i(x_i)$, are potentially product specific, and ordinarily take the following convenient form:

$$C_i(x_i) = K_i + c_i x_i,$$

where K_i is a fixed cost, associated with product development or indivisible capital, and c_i is a constant marginal cost. Prices are normalized so that the price of the composite commodity is \$1, and costs are denominated in dollars or units of the composite commodity.

In the basic models, each product is produced by at most one firm, and, when the number of possible products is sufficiently large, not all products will be produced. Given any set of produced products, the equilibrium is (ordinarily) a Cournot equilibrium – firms choose quantities to maximize profit. [Koenker and Perry (1981) use conjectural variations to generalize this aspect of the basic model.] In free-entry equilibrium all products which are produced earn non-negative profit, and entry of any additional product is not profitable. That is, there exists no non-produced product which could cover its costs in the Cournot equilibrium which would result if the product were produced.

This sort of model is obviously quite flexible and well adapted to welfare analysis. One can create a variety of tractable models by choosing appropriate functional forms for the utility function. Since profit can be measured in units of the composite commodity, welfare analysis is relatively easy. Supposing that profits accrue to the representative consumer, welfare comparisons merely involve comparisons of the representative consumer's utility in alternative situations. If utility is linear in the composite commodity [as in Spence (1976a)] an equivalent welfare criterion is maximization of total surplus.

Two sorts of normative questions have been addressed in these models. The first concerns possible biases in the set of produced products, and is discussed in the following section. The second concerns possible biases in the number of products produced, and is best discussed in the context of Chamberlin's vision of monopolistic competition.

Product selection bias

Given the number of products produced in equilibrium, are some products produced which should not be, and others not produced which should be? We can convey the basic insights regarding possible biases in product selection by adapting the model used in Spence (1976b). Let the utility function take the following form:

$$u = y + a_1x_1 + a_2x_2 - 1/2\left(b_1x_1^2 + 2dx_1x_2 + b_2x_2^2\right),$$

where the parameters a_i, b_i and d are positive and $d < b_i$, $i = 1,2$. To focus on product selection bias, we suppose that both products cannot profitably be produced in a Cournot equilibrium, and ask which one will be produced and which one should be produced. Since utility is linear in y we use the total surplus optimality criterion.

In the monopoly equilibrium the gross profit of firm i (total profit plus K_i) is

$$\Pi_i = \theta_i^2/4b_i,$$

and the gross surplus generated by product i (total surplus plus K_i) is

$$S_i = 3\theta_i^2/8b_i,$$

where $\theta_i = a_i - c_i$. Thus, firm i captures two-thirds of the gross surplus generated by its product. There is a selection bias if the product which produces the larger total surplus is not produced in equilibrium.

To illustrate the possibility of a selection bias, suppose that $K_1 > K_2$ and that $\Pi_2 - K_2 > 0$, or equivalently, that $S_2 > 3K_2/2$ (the second product is profitable). In this case there is inevitably a selection bias if product 1 generates the larger total surplus ($S_1 - K_1 > S_2 - K_2$) but does not generate positive profit ($S_1 < 3K_1/2$), and hence will not be produced. Note also that if $K_i > 3S_i/2$ for $i = 1,2$, either product is viable by itself, but, by hypothesis, both are not. Hence, multiple equilibria are possible.

To see the forces which tend to generate a selection bias, suppose initially that $S_i = 3K_i/2$, $i = 1,2$, so that either product would earn zero profit, and that $K_1 > K_2$. Either product is then viable, but the first generates a discretely larger surplus. Then a small increase in K_1 or b_1 or a small decrease in θ_1 (equal to $a_1 - c_1$) renders product 1 unprofitable, even though it still generates the larger surplus. Thus, among other things, large product development costs and price inelastic demand functions tend to produce a selection bias. This point is developed more fully in Spence (1976a, 1976b).

A standard result in welfare economics is that price discrimination may be welfare improving if there are significant product development costs or economies of scale. This is because a non-discriminating monopolist cannot capture all the surplus it creates. In the model developed here it is easy to show that if firms can perfectly discriminate (that is, capture all the surplus they create), there is no selection bias and the first best optimum is achieved. Spence (1976a) develops this result in a more general context.

It is also worth noting that when $d < 0$, making the products complements, it is possible that neither product by itself is profitable, but that, as a package, they are profitable. In this case, one expects *one* firm to produce both products.

2.2. Monopolistic competition and the representative consumer

A major contribution of the non-address branch is the formalization of Chamberlin's model of monopolistic competition. Hart (1985) is especially interesting in this regard. The representative consumer approach can be used to construct a Chamberlinian model of monopolistic competition as follows. Write the utility function as:

$$u = U(y, V(x_1, \ldots, x_n)),$$

and assume that $V(\cdot)$ is a symmetric function. [Dixit and Stiglitz (1977), for example, use a symmetric CES specification.] Similarly, assume that the cost functions are identical, so that $c_i = c$ and $K_i = K$ for all i. This generates a Chamberlin model because the demand functions inherit the symmetry of $V(x_1, \ldots, x_n)$.

The major issue in this case is whether there are too few or too many products in equilibrium. Dixit and Stiglitz (1977) provide an interesting analysis of this problem. In the unconstrained optimum, prices of all produced goods must equal c, an impossibility in the Chamberlin equilibrium. A constrained optimality criterion is therefore appropriate – the constraint being that all firms must cover their costs of production. They discover cases in which the equilibrium is a constrained optimum and cases in which there are too few and too many products in equilibrium, relative to the constrained optimum. Hence, there is no presumption that the number of products is optimal in Chamberlin's model of monopolistic competition.

An important question that arises with the model of the representative consumer is what lies behind the assumed utility function. Of course, one answer is that all individual's preferences are identical. This answer is, however, inconsistent with awkward facts (5), that each consumer buys only a small subset of the available commodities, and (7), that tastes are revealed to differ among individuals. Hence, a number of authors have investigated the micro foundations of the Chamberlin model.

2.3. Chamberlin's model and diversity of tastes

A number of papers derive the symmetric demand functions of Chamberlin's models from diverse consumer tastes. See, especially, Anderson et al. (1988), Ferguson (1983), Hart (1985), Perloff and Salop (1985) and Sattinger (1984). We will develop the Perloff and Salop model as a way of illustrating the type of preferences that are required for symmetry. Although we will not develop the point, these models are interesting for another reason as well: implicit in them is a way to develop models that are hybrids of the address and non-address approaches. See, especially, Deneckere and Rothschild (1986).

There is a large number, N, of consumers and an infinite number of possible products. Any consumer buys at most one product, and if he buys a product he buys exactly one unit of it. Given n produced goods, a consumer's preferences over the n goods are described by $(B\theta_1, \dots, B\theta_n)$. $B\theta_i$ is the value (in units of a composite commodity) that the consumer attaches to one unit of good i. For all consumers, θ_i is a random drawing from a differentiable density function, $f(\theta)$, with finite support, and B is a parameter that captures preference intensity.[1]

Consider two products with prices p_i and p_j. Product i is preferred to product j if $B\theta_i - p_i > B\theta_j - p_j$ or if $\theta_j < \theta_i + p_j/B - p_i/B$. Since θ_i and θ_j are independent drawings from $f(\theta)$, we can compute the probability that any consumer prefers product i to product j. Given p_i, p_j and (for the moment) θ_i, the probability that i is preferred to j is $F(\theta_i + p_j/B - p_i/B)$, where $F(\cdot)$ is the cumulative density function associated with $f(\cdot)$. But θ_i is also a random variable and hence the probability that any consumer prefers i to j is

$$\int F(\theta_i + p_j/B - p_i/B) f(\theta_i)\,\mathrm{d}\theta_i.$$

Given n produced goods, the probability that any consumer prefers i to all other produced products is

$$H_i(p_i, \bar{p}, n) = \prod_{j \neq i} \int F(\theta_i + p_j/B - p_i/B) f(\theta_i)\,\mathrm{d}\theta_i,$$

where $\bar{p}$ is the vector of prices, p_j, $j \neq i$. Assuming that there is at least one good for which $B\theta > p$ for each consumer, the (expected) demand function for good i is

$$Q_i(p_i, \bar{p}, n) = NH_i(p_i, \bar{p}, n).$$

These demand functions exhibit some obvious symmetry properties. First, $Q_i(p_i, \bar{p}, n)$ is symmetric in the prices p_j, $j \neq i$. For example, if $n = 3$, then

$$Q_1(p_1, p_2, p_3, 3) = Q_1(p_1, p_3, p_2, 3).$$

Furthermore, if all prices are identical, as they are in equilibrium, quantities demanded from each firm are identical and equal to N/n.

[1]It is not necessary for symmetry that different consumers' θ's are drawn from the *same* density function. What is necessary is that any consumer's θ's be *independent* drawings from some density function. The way to construct hybrid models is to fix the number of possible products and let any consumer's $(\theta_1, \dots, \theta_n)$ be a drawing from some *joint* density function.

Now let us characterize the symmetric equilibrium of this model. For concreteness, it is useful to consider the case in which $f(\theta)$ has uniform density on $[u, v]$. To characterize the equilibrium price given n, $p(n)$, set all prices but the ith price equal to a common value, p. Then the ith firm's profit is a function of p_i, p and n. Setting the partial derivative of its profit function with respect to p_i equal to zero, and all prices equal to $p(n)$, we obtain:

$$p(n) = c + B(v - u)/n.$$

As n increases without bound, $p(n)$ approaches c. As preference intensity, B, goes to zero, all goods become perfect substitutes and $p(n)$ approaches c. This reflects the fact that this model is just the Bertrand model in the limit where $B = 0$.

To characterize the free-entry equilibrium, we use the zero-profit condition. In free-entry equilibrium the equilibrium number of firms, $\bar{n}$, is such that each firm earns zero profit. The zero-profit condition implies that

$$\bar{n} = [BN(v - u)/K]^{1/2}$$

It is also easy to determine the optimal number of firms in free-entry equilibrium. Given n, the expected maximum value of $B\theta_i$, $M(n)$, is

$$M(n) = B(nv + u)/(n + 1).$$

The expected value of total surplus, when all goods are sold at a common price, is then $N[M(n) - c] - nK$.

The optimal number of products, n^*, maximizes total surplus:

$$n^* = [BN(v - u)/K]^{1/2} - 1.$$

Observe that n^* is $\bar{n} - 1$: in free-entry equilibrium, the number of firms is approximately the optimal number.

Because, by assumption, each firm produces at most one product, the Chamberlinian equilibrium is inconsistent with awkward fact (4). This raises an interesting question: If firms were allowed to produce more than one product in a Chamberlin model, would they choose to do so? That is to say, is the equilibrium of such a model consistent with awkward fact (4), when the one-firm, one-product assumption is relaxed?

Two key characteristics of this model are now apparent. First, the zero-profit condition that we used above is appropriate only because of symmetry. There is an integer problem (see Section 5 below) in that n firms might make small profits while $n + 1$ firms would make losses. But within the limits of this integer problem (which is trivial when n is at all large) zero profit is appropriate because a new entrant takes customers equally from all existing products. We shall see below that because symmetry is not a property of address models, the zero-profit condition for entry equilibrium does not apply to them.

Second, symmetry arises because the θ's for each consumer are independent random drawings from some density function. It follows that if one good i were removed from the choice set of n differentiated goods, all those consumers who were purchasing i would redistribute themselves uniformly over the other $n - 1$ goods. This property is an appealing one where differentiation is spurious. If all soaps were identical so that perceived differences were solely a product of brand-image advertising, the removal of one soap might lead to this symmetric redistribution of purchases.

This property would not be found, however, where consumers agree on what differentiated products are, and are not, close substitutes for each other, and, in such cases, it conflicts with awkward fact (6). For example, the removal of one low-tar brand of cigarette would lead mainly to increases in the demands for other low-tar products, and not to symmetric increases in the demands for low-tar, medium-tar and high-tar cigarettes. Since the properties of equilibrium in models which do allow for such agreement among consumers are significantly different for those that do not, the propositions that follow from the Chamberlin model must be suspect in those cases where agreement exists. [For further discussion, see Archibald, Eaton and Lipsey (1986).]

3. The address approach

A key aspect of the address approach is that it allows for diversity of consumers' tastes while making the closeness of substitutability among goods at least partially an objective phenomenon. In all consumers' minds, two low-tar cigarettes will be closer substitutes than a low- and a high-tar cigarette; two subcompact cars will be closer substitutes than a subcompact and a stretch limousine; two adjacent drugstores will be closer substitutes than two drugstores at the opposite ends of town.

3.1. Describing a good

In the address branch, a good is described by (θ, p), where p is its price and θ is its "address", some relevant physical description of the good. The address can either be a scalar or a vector and the descriptor, θ, can have many different interpretations. Here are the most common:

- θ may be the location of a firm in some physical space. On a line, it is one number; on a plane, two.

- θ may be the description of the good in some other "spectrum"; for example, the spectrum of color, or a "quality" spectrum.
- θ may be the time at which a service is delivered as, for example, with airline or TV scheduling.
- In Lancaster's characteristics model, a specific good embodies characteristics, z, in fixed proportions. Thus, in the two-characteristics case, we can describe goods by one number, for example, $\theta = z_1/z_2$, where z_1/z_2 is the fixed proportion of the quantity of characteristic one to the quantity of characteristic two.

3.2. Consumers' preferences

Consider first the usual model of spatial competition where firms and consumers are distributed over some geographic space. Any consumer's preferences can be described by a standard utility function, $U(x, y)$, where x is quantity of the good sold by spatially differentiated firms and y is quantity of a composite commodity. Let $\bar{\theta}$ be the consumer's address in the physical space, θ_i the address of firm i, and $T(\theta_i, \bar{\theta})$ the cost of transporting a unit of x from θ_i to $\bar{\theta}$, and assume that consumers bear transport costs. The consumer will, of course, buy x from the firm with the lowest delivered price, $p_i + T(\theta_i, \bar{\theta})$. Thus, awkward fact (5), that each consumer buys only a small number of the goods available, is a theorem in this address model. (As we will see it is also a theorem in Lancaster's characteristics model.)

To emphasize the common features of all address models, it is useful to represent the consumer's preferences in the model of spatial competition indirectly:

$$W(\theta, p, \bar{\theta}) = \max_{x,\, y} \left\{U(x, y) \text{ s.t. } \left[p + T(\theta, \bar{\theta})\right]x + y = 1\right\}.$$

The level surfaces of the indirect utility function $W(\theta, p, \bar{\theta})$ in a one-dimensional physical space when transport costs are a linear function of distance are illustrated in Figure 12.1. They are linear tents centered on the consumers location, $\bar{\theta}$.

In Lancaster's characteristics model any good embodies characteristics in fixed proportions, and quantities of the characteristics are arguments of consumer's utility functions. In the primal space with two characteristics the utility function is $U(g(z_1, z_2), y)$, where $g(z_1, z_2)$ is a utility aggregator, z_1 and z_2 are the quantities of characteristics embodied in group goods, and y is the quantity of a composite commodity. If goods are combinable, so that aggregate quantities of z_1 and z_2 obtained by a consumer are simply the sums of the quantities of characteristics embodied in the bundle of goods he purchases, the utility-maximizing consumer needs to purchase no more goods than there are character-

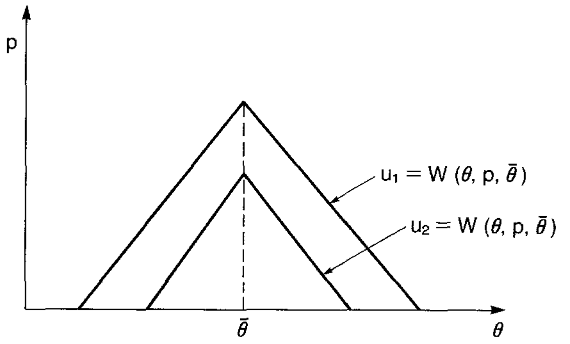

Figure 12.1

istics. Hence, when the number of characteristics embodied in a group of goods is small relative to the number of goods embodying them, awkward fact (5) is again a theorem. See Lancaster (1966).

The recent literature on Lancaster's characteristics model has assumed that each consumer purchases at most one out of a group of differentiated goods, arguing that for technological or other reasons, goods are not combinable. [See Lancaster (1979) and Archibald and Eaton (1987) for a consideration of the very thorny issues concerning combinability.] In this case it is again convenient to derive an indirect representation of the consumer's preferences. A good may be described by the quantities of characteristics, z_1 and z_2, embodied in a unit of that good. The unit in which we measure quantity of the good is, of course, arbitrary. If the good is some brand of cigarettes and the initial unit is a pack of cigarettes, then $(10z_1, 10z_2)$ describes the good when the unit is a carton containing 10 packs. Of course, in any description of the good, z_1/z_2 is fixed.

Thus, we have a degree of freedom that we can use to simplify the way in which goods are described. We illustrate by considering one convenient *units convention*. The good is described by the angle, θ, whose tangent is the (fixed) ratio of z_1 to z_2. The units convention is a quarter circle in the (z_1, z_2) space – the line $z_1^2 + z_2^2 = 1$. The parametric (on θ) representation of this convention is

$$z_1 = \cos\theta, \qquad z_2 = \sin\theta.$$

The indirect utility function is then defined as

$$W(\theta, p) = \max_{x,y} \{U(g(x\cos\theta, x\sin\theta), y) \text{ s.t. } px + y = 1\}.$$

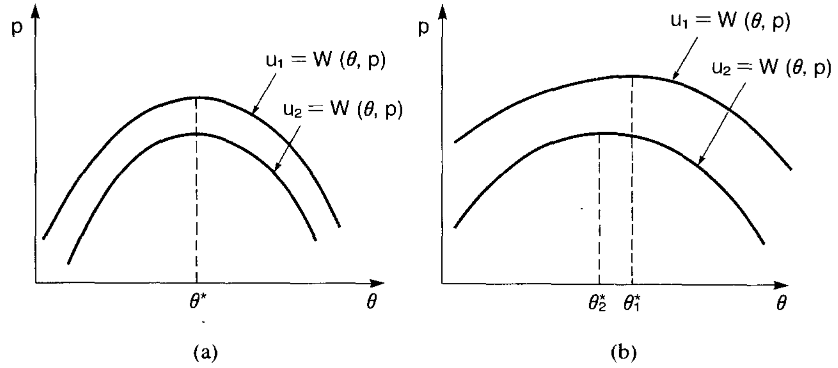

Figure 12.2

The level surfaces of the associated indirect utility function, when z_1 and z_2 are goods, are illustrated in Figure 12.2. When $g(z_1, z_2)$ is homothetic, as it is in Figure 12.2(a), the consumer has a well-defined address, θ^*: for all p, θ^* is the most preferred good. When this is not so, the consumer's address depends on his or her utility level, as illustrated in Figure 12.2(b).

With three characteristics, any units convention involves two θ's. For example, θ_1 might be the ratio z_1/z_2, and θ_2 the ratio z_1/z_3. In general, the choice of a units convention in an n characteristic model requires $n - 1$ θ's.

At a more general level, a consumer's preferences in an address model are described by an indirect utility function, $W(\theta, p)$, where θ is the description of a good in θ space. If θ is interpreted as quality – anything of which more is better – level surfaces will be upward-sloping. In the color-spectrum example shown in Figure 12.3, there are no apparent constraints on the shape of the level surfaces. A consumer could, for example, have strong preferences for the three primary colors and not be attracted to intermediate shades.

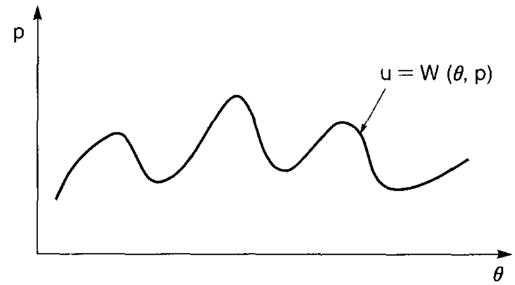

Figure 12.3

3.3. Aspects of an address model

The purpose of address models of product differentiation is to explain observed product diversity by reference to technology and diversity of tastes. Any such model requires the specification of diverse preferences, technology and an equilibrium concept.

Diversity of tastes is ordinarily captured by specifying a continuum of preferences in some appropriate space. In the one-dimensional model of spatial competition diversity is captured by specifying a density of consumers' addresses. Lane (1980) analyzes a two-characteristic model. He uses a Cobb–Douglas specification for $g(z_1, z_2)$, and captures diversity by assuming that the parameter of the Cobb–Douglas function is uniformly distributed on [0, 1].

Technology is ordinarily described by a cost function, $C(\theta, q)$, which specifies the cost of producing q units of a good with description θ. In most models the cost function takes the familiar form:

$$C(\theta, q) = K(\theta) + C(\theta)q.$$

The fixed cost, $K(\theta)$, is often associated with product development or with some other product-specific capital input, and is, therefore, a sunk cost.

Recent address models use subgame perfection in some form as the equilibrium concept. In the simplest case, to characterize free-entry equilibrium potential entrants are treated as players in a game of perfect information. In more complex models, the entire industry structure is generated by an extensive form game of sequential entry. Where sunk costs are an important aspect of the technology, any rational theory of product diversity would seem to require subgame perfection. Sunk costs imply that firms have something to lose if their profit expectations are incorrect. Subgame perfection forces these expectations to be correct.

Two important questions now arise. First (for all consumers) is $W(\theta, p)$ continuous in θ and p? We believe that for the vast majority of cases, the answer is "yes". Second, are the level surfaces of $W(\theta, p)$ single peaked? In many cases the answer is clearly "yes" – e.g. where goods are differentiated by location, by quality, or by the fixed ratio of one Lancastrian characteristic to another. In other cases it is clearly possible that the answer is "no" – e.g. when goods are differentiated by color.

If the answer to both questions is "yes", we are in a world that is quite different from the "goods-are-goods" world of the non-address branch. Given a large set of available products, n, suppose we identify all consumers who buy product A. Now eliminate A from the set of available products. The fact that level surfaces are single peaked implies that the second choices of the consumers who bought A will be distributed over a small subset of the remaining products.

Now suppose that we perturb the price of A slightly. Then quantities demanded of a small subset of products will change. That is, the price of A will be an argument of a small subset of the product demand functions. Hence, the Chamberlin symmetry assumptions are violated. Moreover, aggregate choice behavior cannot be captured by the fiction of a representative consumer with a strictly quasi-concave utility function. For any strictly quasi-concave utility function, there is a non-empty set of prices $(p_1, \ldots, p_n)$ such that all n prices enter all n demand functions in a non-trivial way.

If level surfaces of $W(\theta, p)$ have many peaks (and tastes differ among consumers), then it is conceivable that second choices will be (uniformly) distributed over the remaining products and that the price of A is an argument of all demand functions. In this case there is no necessary conflict between the representative consumer and the address approaches to product differentiation.

4. An illustrative address model

We begin by developing an extremely simple model in which addresses are the only endogenous variables. That is, the model focuses exclusively on competition in addresses. It illustrates a number of important and robust propositions that arise in a wide range of address models. Having developed the simple model, we consider the robustness of the results in Section 5.

4.1. Competition in addresses

The space in which goods are differentiated is a segment of the real line of unit length. We refer to the good with address θ_i as good i. In general, the space may be interpreted as a geographic space – a main street, for example – or as some more general charactistics space. Lancaster (1979) and Salop (1979) carefully develop the characteristic interpretation in this type of model.

The cost function, which is independent of the good's address, is

$$C(q) = K + cq,$$

where q is output, c is a constant marginal cost – which, without loss of generality, we take to be zero – and K is a sunk cost associated with an address-specific, and infinitely durable investment, I. Thus, $K = rI$, where r is the rate of interest reflecting the opportunity cost of investing in this industry. The cost I may be thought of either as a product development cost or as the cost of some type of physical capital of which only one indivisible unit is needed. To focus exclusively on competition in addresses, we assume that prices of all goods are exogenous and equal to one. To focus on non-collusive free-entry equilibrium, we assume that each good is produced by a separate firm and that there is an arbitrarily large number of potential entrants.

Let $\bar{\theta}$ be the address of some consumer in the characteristics space. Let good i, with address θ_i, be the good nearest to $\bar{\theta}$. The consumer's utility function is

$$U(x, y, \bar{\theta}, \theta_i) = \begin{cases} y, & \text{if } x < 1, \\ y + R - T(D), & \text{if } x \geq 1, \end{cases}$$

where D is the distance from $\bar{\theta}$ to θ_i (that is, $D = |\bar{\theta} - \theta_i|$), y is the quantity of a composite commodity, and x is the quantity of good i. We assume that $R - T(D)$, the consumer's reservation price for good i, exceeds 1 (the exogenous price of good i) for all θ_i and $\bar{\theta}$ in the unit interval, which implies that the consumer demands one unit of good i. We assume that $T'(D) > 0$, which allows us to interpret $\bar{\theta}$ as this consumer's most preferred good since the consumer's utility is maximal when $\theta_i = \bar{\theta}$. We capture diversity of consumers' tastes by assuming that $\bar{\theta}$ is uniformly distributed on the unit interval with unit density. Notice that consumers differ only with respect to their most preferred good.

In a model of spatial competition $T(D)$ is the cost of transporting one unit of the good a distance D. In a more general characteristics framework, $T(D)$ is a utility cost (measured in units of the composite commodity) associated with the distance from θ_i to $\bar{\theta}$, the consumer's most preferred good. For concreteness and clarity we will use the spatial competition interpretation of the model.

4.2. *Free-entry equilibrium*

It is a simple matter to characterize the free-entry equilibrium of this model. Let w denote the length of any interval between any two adjacent goods. We call any such interval an *interior interval*. We assume that potential entrants play a game of perfect information. Given our assumption that I is an address-specific, infinitely durable investment, the potential entrant knows that the addresses of the two goods at either end of the interval of length w are fixed, which, given our other assumptions, allows the entrant to calculate its post-entry profit. For any address in this interval, the entrant would attract the custom of half of the customers in the interval. Its gross profit, total profit plus K, would then be $w/2$, and it would *not* enter if and only if

$$w \leq 2K,$$

which is the no-entry condition for any interior interval. (We assume for convenience that zero profit does not induce entry.)

Let v denote the length of the interval between either boundary of the space and the good nearest that boundary, the peripheral good. The entrant's best address in either peripheral segment is adjacent to the peripheral good, where it obtains the purchases of all of the customers in the peripheral segment. So, the no-entry condition for either peripheral segment is

$$v \leq K.$$

These two conditions completely determine the free-entry equilibrium of this simple model. Interior segments can be no larger than $2K$, and the peripheral segments can be no larger than K. Some important, and we stress *robust*, features of free-entry equilibrium in this model deserve attention.

4.3. *Pure profit in free-entry equilibrium*

The profit any existing good earns is obviously directly proportional to its market. To maximize the profit of an interior good in free-entry equilibrium, denoted by Π^i, we simply maximize the size of the interior segments on either side of the good. Setting each of these segments equal to $2K$, we see that the good's market is $2K$ since it attracts half of the custom from each of these two segments. Maximal pure profit in free-entry equilibrium is then K, which implies that the maximum rate of return an interior good can earn on its specific investment, I, is $2r$. The maximal pure profit of a peripheral good in free-entry equilibrium is also K.

4.4. *Non-uniqueness of free-entry equilibrium*

Another obvious consequence of the no-entry conditions is that free-entry equilibrium is not unique. Any configuration of addresses satisfying these conditions is a free-entry equilibrium.

4.5. *No invisible hand*

Any configuration of addresses is Pareto optimal in this model because any relocation of any firm must make some people worse off. Hence, Pareto optimality is not a useful welfare criterion for the model. To consider optimality, some form of aggregation of gains and losses is essential in address models. In the spatial interpretation of the model, the obvious criterion is the minimization of the total resource cost of serving the market, which is equivalent to maximizing the sum of producers' and consumers' surplus. Since consumers' demands are perfectly price inelastic in this model, the only welfare issue concerns the optimal amount of diversity. That is to say, the failure of price to be equal to marginal cost is not indicative of market failure since demand is not responsive to price.

There are two resource costs in this model: costs associated with goods, K per good, and transport costs. If $T''(D) \geq 0$, it is intuitive and easily proven that the configuration of addresses which minimizes transport costs, given n goods, is $(1/2n, 3/2n, \ldots, 1 - 1/2n)$. In this configuration, the total resource cost, $R(n)$, is

$$R(n) = n\left[K + 2\int_0^{n/2} T(D)\,\mathrm{d}D\right].$$

The second term in the brackets is transport costs borne by a representative good's customers, and K is, of course, the cost associated with the good. Ignoring the obvious integer problem, the optimal number of goods, n^*, is characterized by $R'(n^*) = 0$.

It is now easy to see that there is no invisible hand at work in this model. To characterize free-entry equilibrium we needed to know nothing about $T(D)$, whereas we do need to know $T(D)$ to determine n^*. It follows that there may be too much, too little, or just the right amount of diversity in free-entry equilibrium.

There are two intuitive ways to see this. First, suppose that the optimal configuration, $(1/2n^*, 3/2n^*, \ldots, 1 - 1/2n^*)$, is a free-entry equilibrium. Since free-entry equilibrium is not unique, there are any number of non-optimal free-entry equilibria. But the optimal configuration is not necessarily a free-entry equilibrium. Suppose, for concreteness, that $T(D) = tD$. In this case n^* is an increasing function of t. Then if t is sufficiently small, n^* will be so small that the no-entry conditions will not be satisfied in the optimal configuration. In this case, there is too much diversity in any free-entry equilibrium. Furthermore, if t is sufficiently large, n^* will be so large that each good in the optimal configuration could not cover its costs. In this case, although the optimal configuration is a free-entry equilibrium, it is difficult to imagine an entry process that would give rise to it. That is to say, if we insist that goods cover their costs, there will be too little diversity in free-entry equilibrium when t is large.

4.6. Foresighted entry

Is any one of the many free-entry equilibria which are possible in this model a salient equilibrium – one we would expect to see? Prescott and Visscher (1977) ask this question in this model when entry is sequential. [Hay (1976) examines sequential entry in a model with endogenous prices. See Baumol (1967, n. 4. p. 679) for an extremely insightful early discussion of the issue of entry in an address model.] Each firm is constrained to own at most one good, the market is initially unserved, and the order in which firms confront the entry decision is predetermined. The sequential entry process is modelled as a game of perfect information and hence the equilibrium they derive is subgame perfect.

We know from our earlier discussion that $\underline{w} = K$ is the size of market which produces zero pure profit for one good. Let $\underline{n}$ be the smallest integer, n, such that $n \geq 1/2\underline{w}$, and consider the following configuration of $\underline{n}$ firms: $[1/2\underline{n}, 3/2\underline{n}, \dots, 1 - (1/2\underline{n})]$. As you can easily verify, this configuration deters entry, and any configuration with fewer than $\underline{n}$ firms does not. Thus, this configuration is a free-entry equilibrium that maximizes the pure profit that can be extracted from the market. As you can also verify, if $1/2\underline{w}$ is not an integer, there are other configurations of $\underline{n}$ firms that satisfy the no-entry conditions.

To see the nature of the Prescott–Visscher equilibrium, suppose that $1/2\underline{w}$ is an integer. In this case

$$[1/2\underline{n}, 3/2\underline{n}, \dots, 1 - (1/2\underline{n})]$$

is the only configuration of $\underline{n}$ firms which deters further entry, and in this configuration each firm earns pure profit K, the maximum possible profit in free-entry equilibrium. This configuration is the unique perfect equilibrium configuration of the sequential entry game, although any firm in the sequence can be located at any of the $\underline{n}$ locations in the configuration. To see that it is a perfect equilibrium, consider the sequence in which firm 1 locates at $1/2\underline{n}$, firm 2 at $3/2\underline{n}$, firm three at $5/2\underline{n}$, and so on. It is clear that the $\underline{n}$th firm would choose to locate at $1 - (1/2\underline{n})$ if the first $\underline{n} - 1$ chose to locate at $(1/2\underline{n}, \dots, 1 - 3/2\underline{n})$ since by choosing $1 - (1/2\underline{n})$, it gets the maximum possible profit in free-entry equilibrium, K, and since any other location would induce entry of an additional firm which would leave firm $\underline{n}$ with a profit less than K. Knowing this, firm $\underline{n} - 1$ will choose to locate at $1 - 3/2\underline{n}$ if firms 1 through $\underline{n} - 2$ locate at $(1/2\underline{n}, \dots, 1 - 5/2\underline{n})$, because only this location offers $\underline{n} - 1$ the maximum possible profit in free-entry equilibrium. Replicating this argument for firms $\underline{n} - 2, \dots,$ and 1, we see that this sequence is subgame perfect because it offers each firm the maximum possible profit in free-entry equilibrium.

Notice that awkward fact (2), that firms rarely produce identical products, is a theorem in this model with foresighted entry. Each firm in the sequence (except the first) has the option of producing a product identical to its predecessor's, but it chooses to maximally differentiate its product from its predecessor's, subject to entry being unprofitable in the interval between the two products.

When $1/2\underline{w}$ is not an integer, the problem is more complex. Prescott and Visscher argue that the following sequence is subgame perfect: firm 1 enters at $\underline{w}$, 2 at $1 - \underline{w}$, 3 at $3\underline{w}$, 4 at $1 - 3\underline{w}, \ldots$, and firm $\underline{n}$ anywhere in the segment between firms $\underline{n} - 1$ and $\underline{n} - 2$ that does not leave a market larger than $2\underline{w}$ on either side of it. In this sequence, the first $n - 3$ firms earn pure profit equal to K and the last 3 firms pure profit greater than zero and less than K.

Suppose now that we drop the restriction that each firm owns at most one address. Then the perfect equilibrium exhibits *monopoly preemption*: the first firm preempts the entire market by locating $\underline{n}$ plants in such a way that no interior segment exceeds $2\underline{w}$ and neither peripheral segment exceeds $\underline{w}$. Observe that from a resource cost, or total surplus, perspective the two solutions are essentially the same (they are necessarily identical when $1/2\underline{w}$ is an integer), despite the fact that the first industry structure is unconcentrated and the second exhibits maximal concentration.[2] The only significant difference between the two is distributional – all of the profit is captured by one firm in the monopoly preemption case. Thus, in this model, awkward fact (4), that a few firms produce many products, is a theorem where entry of firms is sequential. (As we will see in Section 7, when prices and/or quantities are endogenous, *monopoly* preemption is not an implication of sequential entry.)

4.7. History matters

The non-uniqueness of equilibrium implies that the present characteristics of markets may not depend solely on their current demand and cost conditions. Their past histories can matter in important ways. This can easily be illustrated using the Prescott–Visscher model when $1/2\underline{w}$ is an integer.

Consider two markets, A and B. Market A displays all of the features described in the previous section, so that in sequential equilibrium there are $\underline{n}$ firms in the market and each firm is earning a pure profit of K. Market B, however, has a past history that differs from A in one key respect. When the sequence of entry occurred in B, the density of customers was two instead of one. The result was $2\underline{n}$ plants, each earning pure profits of K. After entry, however, there was an unanticipated fall in consumer density to one. The two markets are now identical in terms of demand and cost conditions. Market B will, however, continue to have twice the number of firms as market A and the firms in B will earn zero pure profits while those in A will earn pure profits of K.

If we looked only at the present conditions in these markets, we would be unable to explain their differences. Only a knowledge of their past histories provides the correct explanation – as must generally be the case whenever equilibrium is not unique.

[2] They are not identical when $1/2\underline{w}$ is not an integer. The monopolist need not locate the first and last plants at $1/2\underline{w}$ and $1 - 1/2\underline{w}$, but when each firm chooses one location, the first and last plants will be so located.

5. Pure profit reconsidered

The possibility of pure profit in free-entry equilibrium is clearly a crucial aspect of this and other address models. It is crucial to the non-uniqueness of equilibrium and to the strategic behavior in the context of foresighted entry. Because it is so important and so often misunderstood, we consider the genesis of pure profit at length.

5.1. The importance of sunk costs

In spatial models a continuum of firms would allow all production to occur at the point of consumption so reducing transport costs to zero. In characteristics models a continuum of products would allow everyone to consume their own most preferred good.

The geographic concentration of production in firms located at discrete addresses is an obvious characteristic of the real world. To explain this along with awkward fact (3), that the number of products produced is much less than the number of possible products, some source of increasing returns to scale, or its cost equivalent, is necessary. In our simple model the cost I is the source of decreasing average total costs.

The address-specificity of I is at the heart of the pure-profit result since it is the sunk nature of this cost at any address which forces the entrant to regard addresses as fixed. It is, of course, entirely possible that this cost is not completely address specific. To the extent that it is not, the magnitude of pure profit possible in free-entry equilibrium is altered.

We can briefly consider this possibility by using a simplified version of the analysis in Eaton and Lipsey (1978) supposing that a portion s of I is address specific and therefore that $(1 - s)$ is not. Consider a symmetric configuration of many addresses in which the interior segments are of length w and peripheral segments are of length $w/2$, and each firm occupies only one address. Now ask under what circumstances one existing firm would change its address in response to entry. Suppose that the entrant locates just next to an existing firm and so takes one-half of that firm's market. Any rational relocation decision is obviously quite complex since the existing firm must itself ask whether or not it must take the addresses of other firms, including the entrant's, as given and, if not, it must solve their relocation problems. To avoid solving this problem and to maximize the existing firm's perceived incentive to relocate, suppose that the existing firm believes (or more properly that the entrant believes that the existing firm believes) that by relocating it will touch off a series of instantaneous relocations that would leave it and the entrant with a market identical to its pre-entry market – a market of length w. Given this belief, the existing firm will relocate if and only if

$w - sK > w/2$, since its anticipated market, if it relocates, is w and only $w/2$ if it does not. This inequality implies that the existing firm will not relocate if $w \leq 2sK$.

If the entrant foresees this result (notice that it could not be more optimistic about the possibility of relocation), it will not enter if $w \leq 2sK$ since its post-entry market would be $w/2$, which is not large enough to cover its costs. That is, $w/2 - K < 0$ when $w < 2sK$. This result in turn implies that the maximum pure profit in free-entry equilibrium is sK. Thus, if s is greater than 0, the magnitude, but not the existence of profit, is an issue.

5.2. *The durability of sunk capital, predatory entry and profit dissipation*

To focus on the influence of durability, we return to the case in which sunk capital is completely address specific. Infinitely durable, address-specific, sunk capital commits any plant to remain at its present address forever – it creates in essence a property right to the flow of pure profits available in free-entry equilibrium. As a result, the only sort of entry which is possible is *augmenting entry* – entry that increases the number of addresses.

When sunk capital is not infinitely durable, a second type of entry becomes a possibility. This is *predatory entry* – entry that causes an existing address to be abandoned when its associated capital expires. In these circumstances, a property right to some, but not all, of the pure profit that would be available if augmenting entry were the only possibility can be created by premature replacement of sunk capital in the market. As we show, this behavior tends to dissipate profit.

To see what is involved, we use a simplified version of the analysis in Eaton and Lipsey (1980). We suppose that sunk capital, which again costs I, has a durability of H periods. When unconcerned about predatory entry, a sitting firm replaces its plant every H periods. But a new entrant could establish its plant at the same location just as the sitting firm was about to replace its own capital. The market would then "belong" to the new entrant. But foreseeing that strategy, the sitting firm could renew its capital at some earlier date. We want to find the optimal premature replacement strategy.

To set the stage for a consideration of predatory entry, and its preemption by the early renewal of address-specific capital, we first derive a constraint on the size of intervals between plants which is implied by the possibility of augmenting entry. The constraint is analogous to the no-entry conditions derived in Subsection 4.2. Consider entry into an interior segment of length w. (For simplicity we ignore peripheral segments.) The present value of augmenting entry, discounted over an infinite time horizon and evaluated at the instant of entry, is

$$w/2r - I/[1 - E(rH)],$$

where $E(x) \equiv \mathrm{e}^{-x}$. The first term is the present value of the entrant's revenues and the second the present value of its costs. The condition for augmenting entry to be unprofitable is

$$w \leq 2rI/[1 - E(rH)] \equiv M.$$

For simplicity we suppose that $w = M$ so that firms are separated by the maximum distance consistent with no augmenting entry. We once again assume that each existing firm owns only one address. We wish to find a premature replacement strategy that deters predatory entry. The strategy we consider is for each firm to replace its address-specific capital δ periods prematurely. Thus, each firm incurs the cost I every $H - \delta$ periods and hence never has a commitment to the market less than δ periods.

At the instant in time when capital is replaced, the present value of any existing firm, if the strategy deters predatory entry, is

$$V(\delta) = M/r - I/[1 - E(r(H - \delta))],$$

which is obviously decreasing in δ. Thus, the problem is to find the minimum value of δ, which we denote by δ^*, that deters predatory entry. Think of δ^* as the optimal entry deterring strategy.

The point in time at which predatory entry is most attractive is the instant before an existing firm replaces its capital. At this point, the present value of a predatory entrant which adopts the address of an existing firm, and itself uses the optimal entry deterring strategy, δ^*, is

$$U(\delta, \delta^*) = -I + [M/2][1 - E(r\delta)] + M[E(r\delta) - E(r(H - \delta^*))] + V(\delta^*)E(r(H - \delta^*)).$$

The second term reflects the power of premature replacement as an entry deterrent – the successful predatory entrant must contend with the existing firm for δ periods, and its sales over this interval of time are only $M/2$ per period.

The existing firm's strategy, δ, will deter entry if $U(\delta, \delta^*) \leq 0$. Since $V(\delta)$ is decreasing in δ, δ^* is the value of δ such that $U(\delta, \delta^*) = 0$. That is, δ^* is defined by $U(\delta^*, \delta^*) = 0$. As the reader can verify, $\delta^* = H/2$.

If augmenting entry were the only issue, δ would be zero, and any existing firm's present value at time of replacement would be $V(0) = I/(1 - E(rH))$, which is identical to the present value of the firm's address-specific capital cost. But predatory entry is an issue, and the maximum present value of the firm is only $V(\delta^*) = I/[1 - E(rH/2)]$. As H gets large, $V(\delta^*)$ approaches $V(0)$ and as H gets small, $V(\delta^*)$ goes to zero. In the latter limit all of the potential profit is dissipated by premature replacement.

Profit dissipation also arises in the context of *foreseen* market growth. [See Eaton and Lipsey (1979).] In a growing market, entry itself is premature since only by prematurely entering can a firm create the necessary property right to appropriate the profit available in free-entry equilibrium. Indeed, with many potential entrants and parametric prices, all of the profit is dissipated by premature entry – that is, entry occurs when the present value of entry is zero.

These dissipation results are, of course, in no way inconsistent with the existence of flows of pure profit in static free-entry equilibrium. Indeed, it is this possibility that drives the premature entry and/or capital replacement which itself is partially or totally profit dissipating. And, of course, these results are not (necessarily) welfare improving. For example, the premature replacement of capital to deter predatory entry is associated with a pure deadweight welfare loss.

5.3. Endogenous prices

To articulate in the simplest possible way what we think are the essential issues peculiar to address models, we have assumed that prices are exogenous. In this subsection we discuss some thorny problems that arise when prices are endogenous, and we ask if the basic properties of free-entry equilibrium found in our simple model are robust with respect to endogenous prices. We continue to assume that any firm occupies at most one address. We add the assumption that the configuration of addresses is symmetric on a circular market (which avoids boundary problems) as in Salop (1979). Let $p(n)$ be the symmetric Nash equilibrium price in this circular model with n firms.

If we suppose that entrants are price takers – that entrants believe that the post-entry prices of existing firms will be $p(n)$ – then pure profit is impossible in free-entry equilibrium. For example, an entrant that adopted some existing firm's address and charged a price just lower than $p(n)$ would anticipate positive profit as long as the existing firm was earning positive profit. Hence, zero profit is not a property of free-entry equilibrium if entrants assume that they are price takers.

But in a model where existing firms have sunk costs which commit them to the market, foresighted entrants will not take price as given. The existing firm, which would be wiped out by the price-cutting entrant, will not maintain price. Rather, it will respond by reducing price, and the entrant can foresee that it will. For this reason, Eaton and Lipsey (1978) and Novshek (1980) impose a no-mill-price-undercutting restriction. They do not allow any entrant (or any existing firm) to believe that it can adopt a price which would reduce an existing firm's sales to zero without any reaction. The resulting free-entry equilibrium – where the prices of existing firms are Nash equilibrium prices subject to this restriction, and where entrants take established firms' prices as given but do not consider undercutting – exhibits pure profits.

The no-mill-price-undercutting restriction can be seen as an attempt to rule out an entrant's profit expectations which are wildly inconsistent with its post-entry profit realization. Consider again the price-cutting entrant which expects to appropriate virtually all of an existing firm's profits by replicating its address and undercutting its price. If, given addresses, the price equilibrium is a Nash price equilibrium, then the post-entry prices of the entrant and its intended victim will be equal to marginal cost, and the entrant's profit expectations will not be fulfilled. The entrant is a victim of its own myopic profit expectation. If the entrant must incur address-specific sunk costs, then it will not take the prices of existing firms as given. Instead, it will attempt to anticipate the post-entry price equilibrium. This, of course, suggests that the appropriate equilibrium concept for the subgame in which a potential entrant makes its entry decision is subgame perfection.

Eaton and Wooders (1985) use this approach to characterize the symmetric free-entry equilibria of a model similar to our basic model, in which the price equilibrium in both the pre- and post-entry games is a Nash equilibrium in pure price strategies, and in which each firm owns one address. Their free-entry equilibria exhibit the two fundamental features of our basic model: the maximum rate of return on sunk capital is $3.33r$, and there may be too much, too little, or the optimal degree of product diversity in equilibrium. (Relative to our basic model, the maximum profit in equilibrium is larger because existing firms respond to entry by reducing price, a result which entrants foresee.) This establishes that parametric prices are not a necessary condition for the basic characteristics of the model discussed in Section 3.

There is, in addition, a potential problem of non-existence of price equilibrium associated with price undercutting. The problem is articulated by d'Aspremont, Gabszewicz and Thisse (1979). If, for example, transportation costs in our basic model are linear in distance, then the Nash price equilibrium (for given addresses) does not always exist. If, on the other hand, transportation costs are quadratic in distance (as in the model proposed by d'Aspremont, Gabszewicz and Thisse and employed by Eaton and Wooders), then the price equilibrium does exist.

5.4. *The integer problem, balkanization and localized competition*

The pure-profit result in address models is sometimes *wrongly* attributed to what we call the integer problem. The integer problem, which we brushed aside in our discussion of the Chamberlin model, occurs in non-address models when n firms can make a profit and $n + 1$ firms cannot, and its source is increasing returns to scale or its cost equivalent. As a result, n firms can earn profit in a non-address model that is in equilibrium with respect to foresightful entry. The same problem arises in an address model in which we *arbitrarily assume* that the number of firms is maximized, subject to the constraint that no firm incurs losses.

In either case – and assuming for convenience that sunk costs per period are K, marginal cost is constant, and price parametric – the maximum pure profit consistent with entry equilibrium is K/n. So, for example, one firm can earn up to twice the normal rate of return on its sunk capital ($2r$), two up to $1.5r$, and 10 up to $1.1r$. The integer problem can thus account for substantial long-run profits in industries where demand is sufficient to sustain only a few firms – natural monopolies and natural oligopolies – but as more and more firms can be supported by the market, the excess profit attributable to the integer problem rapidly diminishes.

But in address models, the location of existing goods or products balkanizes the market into a number of overlapping submarkets. As a result, competition is localized – each good has only a few neighboring goods with which it competes directly, regardless of the number of goods serving the entire market. This localized competition imparts a natural oligopoly characteristic to address models – indeed, this is what Kaldor (1935) intended by the phrase "overlapping oligopolies". It also allows the maximum profit consistent with free-entry equilibrium to remain constant as the number of firms in the market is increased. For example, let the density of customers grow in the model considered above. The minimum number of goods consistent with free-entry equilibrium will now grow, and the distance between goods will diminish but each good will continue to have only two neighbors and the maximum profit consistent with free-entry equilibrium will remain at $2r$. [See Eaton and Wooders (1985) for an illustration of this result when price is endogenous.]

The driving feature of these results is that the expected size of the market for a new entrant, R^{e}, is significantly smaller than the market enjoyed before entry by the firms which will be the entrant's neighbors, R^{a} (given identical prices). In our simple model R^{e}/R^{a} is one-half, and, hence, existing firms can earn up to twice the normal return on capital without attracting entry.

5.5. *How robust is balkanization?*

What happens to balkanization and its implication of localized competition when our restrictive assumptions are relaxed? We discuss models of spatial competition first.

A key assumption concerns transportation costs. If these are identical for all consumers and convex in distance, then competition is clearly local in nature. Thus, in one-dimensional models, each firm is in direct competition with at most two others.

Suppose, however, that transport costs are a concave function of distance and identical for all consumers. It is now possible that a low-price firm may be in competition with several high-priced ones, diminishing the extent to which competition is localized.

Another way in which this result can occur is if "transportation costs" are subjective and differ among consumers. Again, a low-price firm could be in direct competition with a number of high-price firms selling identical products from different addresses. In both of the above cases a firm's market area is no longer a connected subset of the entire market, and competition is not localized to the same extent. Analogous possibilities arise in characteristics models – a phenomenon referred to as "cross-over" by Lancaster (1979).

Notice that in both of the above cases there is still some balkanization of the market in that the entry of either one low-priced or one high-priced firm at a specific address will not take sales in one set of *equal* increments from all existing low-priced firms and another set of equal increments from all existing high-priced firms.

Another issue relevant to balkanization concerns the dimensionality of the space itself. Models of spatial competition in a two-dimensional space exhibit the key properties of equilibrium in our simple one-dimensional model. In particular, competition is localized and plants can earn substantial pure profit in entry equilibrium. For example, in a two-dimensional model analogous to the one-dimensional model developed here, Eaton and Lipsey (1976) show that the pure profit of all plants in free-entry equilibrium can be as large as $0.96K$.

In the characteristics model, the number of characteristics embodied in goods, and hence the dimensionality of the space, may be an important determinant of the extent to which competition is localized. Archibald and Rosenbluth (1975) consider the number of neighbors a firm can have in the case where goods are combinable. With two characteristics, each good can be in direct competition with at most two neighboring goods and in the case of three characteristics, the average number of neighbors cannot exceed 6. However, with four characteristics, the average number of neighbors can be as large as $n/2$, where n is the number of goods. Apart from the Archibald–Rosenbluth results for the combining case, we know virtually nothing about how the number of characteristics affects the number of neighbors.

Archibald and Rosenbluth focus on the average number of neighbors a firm can have. This does not seem to us to be the important question. Instead, what matters is whether or not R^e is bounded away from R^a. We suspect this is a characteristic that will be robust to increasing the number of dimensions.

Schmalansee (1983) observes quite rightly that the extent to which competition is localized cannot be resolved by the theorists' paper and pencil. It is at root an empirical question. He develops some empirical tests for localization. The payoff to careful empirical work in this area is, we think, immense.

6. Vertical differentiation and natural oligopoly

In a model of vertical differentiation goods are differentiated in a one-dimensional space but, in contrast to the illustrative model discussed above, the characteristic θ that describes a vertically differentiated good is something of which more is better from every consumer's perspective. In contrast, the earlier model in which consumer's had different preferred θ's can be called one of horizontal differentiation.

There are at least two ways in which vertical differentiation can arise. First, the technology might be such that the product only contained one variable characteristic. Say it is quality, which we assume can vary over the range $0 < \theta < 1$. Second, consumers might live at points θ on a one-dimensional housing estate occupying the range $1 < \theta < 2$, while retail stores were constrained by zoning laws to locate at points $\theta < 1$. The first case is driven solely by technology and the second by institutional arrangements. [Notice in the second case, which is due to Gabszewicz and Thisse (1986), the estate must be one-dimensional and the stores must be constrained to locate on only one side of it.]

Vertical differentiation is an address model but the fact that everyone agrees on the most preferred address for the good or store gives it some special characteristics. One issue that has attracted attention in recent years concerns the circumstances under which the model produces a *natural oligopoly*. In this section we outline the results obtained in a simple model which is designed to illustrate some of the arguments developed by Gabszewicz and Thisse (1979, 1980, 1986) and Shaked and Sutton (1982, 1983).

In this model θ is interpreted as a measure of quality and we assume that the feasible quality range is $0 \leq \theta \leq 1$. Each consumer buys one unit of his or her most preferred good, given the prices and θ's of the available goods. The indirect indifference curves for each consumer have the following form:

$$\bar{u} = m\theta - p,$$

where θ and p are quality and price, and m is the willingness of consumers to trade off quality against price. We capture diversity by assuming that m is uniformly distributed on $[a, b]$ with density D, where $b > a > 0$.

Now consider a consumer's choice between two goods, (θ_1, p_1) and (θ_2, p_2), with $\theta_1 > \theta_2$. If $p_2 > p_1$, all consumers prefer good one, so suppose that $p_2 < p_1$. The "market boundary" in θ space, $\bar{m}$, satisfies:

$$\bar{m}\theta_1 - p_1 = \bar{m}\theta_2 - p_2$$

or

$$\overline{m} = (p_1 - p_2)/(\theta_1 - \theta_2).$$

For $m > \overline{m}$, (θ_1, p_1) is preferred to (θ_2, p_2), and for $m < \overline{m}$, the opposite is true. From this, we see that the demand functions are:

$$D_1(\theta, p) = D(b - \overline{m}),$$

$$D_2(\theta, p) = D(\overline{m} - a).$$

We assume that the only cost of production is K, a sunk cost which is independent of θ. Hence, the marginal cost of producing a higher quality good is zero. The property that is necessary and sufficient for the natural oligopoly result is that the marginal cost of producing higher quality be less than a.

Firms choose price non-cooperatively. If one firm enters it will choose $\theta = 1$, since it will be profitable to produce all the quality possible when people are more than willing to pay its marginal cost. Now let a second firm enter. If it also chooses $\theta = 1$, competition will drive price to zero so that sunk costs will not be covered.

This parallels the natural monopoly result in the undifferentiated Bertrand model. With foresighted entry and constant marginal costs, a second firm will never enter no matter how large the market. But unlike the spaceless model with a homogeneous good, the second firm has the option of building a worse mouse trap (or purposely polluting its mineral spring!). This allows it to differentiate itself from the first firm and so avoid ruinous price competition. It will pay a second firm to enter with a poorer product, if consumers' tastes are diverse enough. Specifically, for $\theta_2 < 1$ the second firm commands a positive market share in the non-cooperative price equilibrium if and only if $b > 2a$. Hence, if $b > 2a$, and K is not too large, there will be at least two firms in this market. Will there be three? It again depends on diversity of consumers' tastes. If they are not too diverse (if $b < 4a$), and if there are three firms, then the one with the smallest θ does not command a positive market share in the non-cooperative price equilibrium. Hence, if $4a > b > 2a$, and if K is not too large, we will have natural duopoly. Notice that the market share conditions for natural duopoly do not depend on the density of consumers, D. With more diversity of tastes we can have more firms, but there is always a maximum number of firms that can coexist in a non-cooperative price equilibrium, and the maximum is independent of D.

Let us now consider sequential entry in the natural duopoly case. What may not be immediately obvious is that, if the second firm does not fear entry from a third firm, it will go all the way to shoddy quality and choose $\theta_2 = 0$. If, however, firm 2 foresees the possibility of entry by a third firm, it will choose a θ_2^* sufficiently close to 1 so that a third firm does not anticipate positive profits at any θ_3.

Now let the density of customers, D, or the size of the fixed costs, K, vary. As D goes to infinity or K goes to zero, θ_2^* goes to 1, while prices go to zero. These results reflect asymptotic optimality. The common sense of the result is that, as the size of the sales to be obtained from a third firm locating between $\theta_1 = 1$ and θ_2 rises, or as the sunk cost K declines, firm 2 must choose its θ closer and closer to θ_1 in order to keep a third firm out. In the limit the two firms select $\theta = 1$ and price is equal to zero. But since fixed cost per unit of output approaches zero in either limit, the two firms will remain profitable.

These results reflect the more general result in this type of model. As long as every consumer would choose maximal quality if asked to pay the marginal cost of producing it, the number of firms is bounded above, and is independent of customer density.

Given the assumptions about tastes for and costs of added quality, the results are driven by the destructiveness of price competition. As in the spaceless model, a second firm would never enter if it had to produce an identical product but, if consumers' tastes are sufficiently diverse, entry with a worse mouse trap is profitable.

In the spaceless model, the natural monopoly result does not hold when competition is in quantities instead of prices. The number of firms in a subgame perfect sequential entry exercise then varies positively with the size of the market. An analogous result occurs in the model of quality competition outlined here. Because individual demand is price inelastic – total quantity demanded is $(b - a)D$ regardless of the prices and θ's of firms – there are many non-cooperative quantity equilibria in the model. To make the quantity setting equilibrium more interesting, assume that each consumer's demand function for the consumer's most preferred good is $1 - p$.

Some tedious manipulation then shows that in the sequential entry game with two firms, the first firm chooses $\theta_1 = 1$ and the second chooses $\theta_2 = 1$. That is to say, although firm 2 could choose to differentiate its product, it will not choose to do so when competition is in quantities instead of prices. Just as the natural monopoly feature of the Bertrand model vanishes when we suppose that competition is in quantities instead of prices, so does the natural oligopoly feature of this model vanish when we switch to quantity competition. No firm will ever choose θ less than one when competition is in quantities, and hence the model is, in essence, now an undifferentiated Cournot model. [Bonanno (1986) derives the same result in a slightly different model.]

We conclude that the natural oligopoly result in an address model with vertical differentiation is driven by the assumption that price is the strategic variable, as well as certain necessary taste and cost assumptions. It is also worth noting that cases in which commodities can be differentiated by only a single characteristic are few in number.

7. Price versus quantity competition

Most address models in which prices are endogenous are based on the presumption that competition is in prices. One reason for this modelling choice is analytical convenience – when one aggregates demand in an address model the natural way to proceed is to derive demand functions. The resulting functions are frequently quite difficult to invert – with n firms one must invert an n-equation system. Thus, it is more convenient to assume that competition is in prices.

Obviously, analytical convenience is not a sufficient reason for assuming that competition is in prices since, as the following examples illustrate, there are significant differences between address models in which competition is in prices and those in which competition is in quantities. (The circumstances in which oligopolists may play either a pricing or a quantity game are discussed by Carl Shapiro in Chapter 6 of this Handbook, so we make only a few points of special relevance to address models.)

We saw in the discussion of natural oligopoly that the very nature of the equilibrium is different: with price competition, products are differentiated; with quantity competition, they are not.

Deneckere and Davidson (1985) show that in a model of differentiated products and many firms, when entry is not a concern, any sort of merger is profitable if competition is in prices, whereas mergers are only rarely profitable if competition is in quantities.

d'Aspremont, Gabszewicz and Thisse (1979) have shown that the very existence of price equilibrium is in question in address models that do not invoke the no-mill-price-undercutting assumption. One source of non-existence in these models is the temptation to undercut, which requires that price be the strategic variable. Salant (1986) shows that the undercutting temptation is removed when competition is in quantities and goes on to prove the existence of quantity equilibrium in one-dimensional address models in a fairly general setting.

One argument for the appropriateness of price competition is that firms must announce price in an address model. Given cost and demand conditions, impersonal market forces cannot give rise to market-clearing prices in differentiated products as they can with undifferentiated products. But this fact does not necessarily mean that prices are the strategic variable.

Salant (1986), for example, argues that in many industries there are long lags in production – the next period's output is committed by this period's production decisions. In this case firms' quantity decisions precede their price decisions. Such firms are forced to ask at what price tomorrow can today's production be marketed? That is, they are forced to make conjectures concerning market-clearing prices for their quantities. In these circumstances, it seems appropriate to model competition in quantities.

In an analysis of undifferentiated products that is clearly also applicable to differentiated ones, Kreps and Scheinkman (1983) consider a two-stage model in which duopolists simultaneously choose quantities in stage 1 and prices in stage 2. The subgame perfect equilibrium in their model is identical to a Cournot equilibrium in spite of competition being in prices in stage 2. The intuition is that firms recognize the destructiveness of Bertrand competition and commit themselves in advance to curbing their non-cooperative pricing behavior by choosing a limited quantity. Where this model applies, we would expect Cournot results when demand is correctly predicted, but Bertrand-style price wars whenever unexpectedly low demand occurs.

Singh and Vives (1984) consider a two-stage game in which each firm first selects price or quantity as its strategic variable and then competes according to its selected variable in the second stage. In their model the dominant strategy is to select quantity, and this leads to the Cournot equilibrium.

These studies convince us that there is likely to be a range of real circumstances in which quantity, rather than price, will be chosen as the strategic variable. This in turn suggests a research agenda: rework address models using quantity competition. Here is one obvious example. In response to non-existence of price equilibrium in Hotelling's model, d'Aspremont, Gabszewicz and Thisse (1979) present a modified version of Hotelling's model in which the sequential equilibrium exhibits *maximum* differentiation. Is this result robust to competition in quantities? Will differentiation be less than optimal if competition is in quantities? We suspect the answers are "no" to the first question and "yes" to the second.

8. Multi-address firms

The great bulk of the literature on product differentiation in both large- and small-group situations has used the simplifying assumption of one-address firms. On a research agenda of tackling the easiest problems first, this is understandable. But awkward fact (4) states that the vast majority of real-world firms are multi-address, both in characteristic and geographic space. This is obvious in the consumers' goods sector – toiletries, tobacco products, refrigerators, automobiles, etc. The growth of retail chains has spread the same effect to the retail sector. For example, many single-outlet restaurants, which were the dominant form within living memory, have given way in the lower price range to such chains as

Macdonald's, Kentucky Fried Chicken, and the Dutch Oven. To come to grips with reality, the model needs to be extended to analyze multi-address firms.

One valuable step in this direction, using the non-address approach, has recently been taken by Brander and Eaton (1984). They deal with a model of four products. The pairs (1, 2) and (3,4) are close substitutes, while the pairs (1, 3), (1, 4), (2, 3), and (2, 4) are more distant substitutes as defined by cross-elasticities of demand. Because they use the goods-are-goods approach, they are able to use competition in quantities without difficulty.

The results are driven by the property that when two single-product firms compete, profits in equilibrium are lower the more substitutable are the two goods. As a result, a *segmented market structure*, with one firm producing products 1 and 2 and the other firm producing products 3 and 4, yields higher profits than an *interlaced market structure*, where each firm produces one good from the pair (1, 2) and one good from the pair (3, 4).

It follows immediately that when two firms enter the market sequentially choosing (by assumption) two products each, and knowing that no other firms will enter, the segmented structure will result. One firm will choose the pair (1, 2) while the other chooses (3, 4), thus minimizing the profit-reducing competition between them.

If firms are also allowed to choose the number of products they produce, the above result holds only for some intermediate levels of demand. At one extreme, demand may be so low that, if one firm chooses any one of the four products, the other would choose not to enter. This is a natural monopoly. At the other extreme, demand may be so great that, even if one firm chooses all four products, the other firm would still choose to enter. This will lead to an overlapping market structure. Only for intermediate levels of demand is segmentation the unique subgame perfect equilibrium.

Finally, if the two firms fear further entry, they may be led to select an interlaced, rather than a segmented, market structure. Since the interlaced market structure is more competitive and therefore discourages entry, it may result in higher duopoly profits than would result from the oligopoly that would evolve if the two firms encouraged entry by selecting a segmented structure. The general insight is that creating a more competitive n-firm situation may deter entry and result in higher profits for the n firms than the profit they would earn in any $n + 1$-firm free-entry equilibrium. Brander and Eaton's analysis leads to the conjecture that in a growing market, the natural evolution may be from monopoly, to a segmented duopoly, to an interlaced oligopoly.

Two earlier studies of multi-unit firms in address models are Schmalensee (1978) and Eaton and Lipsey (1979). Schmalensee studied a situation where a firm could deter entry by proliferating differentiated products so as to avoid presenting a new entrant with a market niche large enough to be profitable. The monopoly preemption result in our basic model of horizontal differentiation is a simple illustration of Schmalensee's argument.

Eaton and Lipsey analyzed a growing market initially large enough to support production at just one address. If the incumbent firm does not preempt the market, and if there are many potential entrants, then entry at a new address will occur at the point in time, T, when the present value of entry is zero. But the incumbent firm can deter entry by itself choosing a new address an instant before T. If the incumbent chooses the new address, the profit generated by the whole market will be larger (than if an entrant chooses the new address) for the simple reason that the incumbent will choose the new address, and the two prices, to maximize total profit. This implies that the incumbent will choose to deter entry. (The argument is similar to the one we used above in discussing the deterrence of predatory entry by premature replacement of sunk capital.)

Judd (1985) has pointed out that these preemption arguments neglect to consider the possibility that a new entrant, who would find entry unprofitable *given* the existing occupied addresses, might still be able to enter by inducing the incumbent to vacate one or more addresses – a strategy of predatory entry.

To see what is involved, consider a market of unit length with uniform customer density. Suppose that if firm 1 were to locate at two addresses, $1/4$ and $3/4$, an entrant which took these locations as given could not cover its costs, which are composed of a constant marginal cost and a cost of entry. Suppose also that two firms could cover their costs if one of them was located at $1/4$ and the other at $3/4$.

Now suppose that a first firm considering entering at locations $1/4$ and $3/4$ anticipates the possibility of predatory entry. We consider three cases in which the original firm is at $1/4$ and $3/4$ and the new firm at $1/4$.

Case 1: Competition in prices. Competition in prices would drive price down to marginal cost at $1/4$, and the incumbent would earn zero gross profit from his address at $1/4$. Price would, of course, exceed marginal cost at $3/4$ and the market boundary would be at some point to the right of $1/2$. If firm 1 were to abandon its address at $1/4$, the prices at $1/4$ and $3/4$ would both rise and the market boundary would be at $1/2$. Hence, firm 1's profit would increase. Foreseeing this result, one would not attempt the $(1/4, 3/4)$ preemption strategy.

Case 2: Cooperative pricing. Now suppose that post-entry pricing is cooperative. (Firm 1 could induce cooperative pricing if it could credibly announce a price-following strategy – "we will not knowingly be undersold" is, for example, a stated policy of some well-known retailers.) The prices at $1/4$ and $3/4$ are now identical regardless of whether there are two or one plants at address $1/4$. In this case, firm 1 would not abandon address $1/4$. Knowing this, there would be no predatory entry and the $(1/4, 3/4)$ preemption strategy works.

Case 3: Competition in quantities. In this case, the predatory entry strategy might or might not induce firm 1 to abandon address $1/4$. If the cross-elasticities between the goods at the two addresses are high enough, it would. If they are low enough, it would not. The $(1/4, 3/4)$ preemption strategy may or may not be profitable to firm 1.

Thus, the possibility raised by Judd of predatory entry in these preemptive models means that the obvious preemption strategy of locating plants at $1/4$ and $3/4$ may or may not be subgame perfect, depending on how fierce is post-entry competition. Of course, there may be other preemption strategies which are subgame perfect, even when competition is in prices. Consider, for example, the preemptive strategy of locating two plants at 0 and at 1. If competition is in prices, an entrant could induce firm 1 to abandon one of these addresses, but there may be no address for the entrant which both induces firm 1 to abandon an address and also offers the entrant positive profit in the ensuing duopoly equilibrium.[3]

9. Product diversity and economic policy

Rational economic policy requires an understanding of the welfare issues which arise in the context of differentiated products. The model using the representative consumer has the advantage of being tractable. Welfare results are easily derived from it. However, tractability in deriving incorrect results is no advantage, and we do know that in address models, whenever preferences are single peaked in (θ, p) space, aggregate consumer behavior cannot be caught by a representative consumer. Thus, it seems to us that while there may be cases for which the representative-consumer approach is appropriate, there are many problems for which an address model seems appropriate. Nothing can be learned about these problems from representative-consumer, or Chamberlinian models, since they do not capture all of the awkward facts.

Similar remarks apply to Chamberlin-style models that employ the symmetry assumption, even when they are rooted in individual taste differences, since these models appear to be inconsistent with awkward (6). So if the address model characterization of consumers' behavior captures important aspects of reality, there is no reason to believe welfare propositions derived from either of these approaches. For people interested in giving policy advice relevant to the bulk of manufacturing industries that sell differentiated goods this is a serious matter.

[3]In contrast to the rich set of possibilities considered in the text, Judd appears to claim that multi-address preemption is never subgame perfect. In his model, where curiously there are no addresses, just two goods, the claim is driven by his assumption that firm 1's revenue from plants at $1/4$ and $3/4$ when firm 2 also has a plant at $1/4$ is less than its revenue from one plant at $3/4$ while 2's single plant is at $1/4$. This amounts to solving the preemption issue by assumption: since predatory entry is assumed always to work, preemption is never subgame perfect.

Notice that the standard market failure associated with the divergence of price from marginal cost necessarily arises in markets for differentiated products whenever demand is not perfectly price inelastic. Market failure is a ubiquitous problem in address models with balkanization and localized competition since in free-entry equilibrium the position of each product is very much like the standard stylization of a natural monopoly. If the standard natural monopoly problem is difficult to solve in practice, the natural monopoly problem in address models is much more so since it is a pervasive problem.

In addition, the problem of optimal product diversity arises. This is, we believe, an even more difficult problem – one that, from a policy perspective, we know very little about. Our basic address model is a useful device for conveying the awkward nature of this problem since the only optimality issue is diversity. We showed in our discussion of that model that there is no general relationship between product diversity in free-entry equilibrium and optimal product diversity. However, it is clear that the diversity observed in free-entry equilibrium is unlikely to be the optimum amount. Even if the optimum diversity is a free-entry equilibrium, there are many other free-entry equilibria and no market force which pushes the equilibrium to the optimum. The awkward problem is that we do not even know the nature of the bias – whether there is likely to be too much or too little diversity in equilibrium.

If we interpret the basic model as a model of spatial competition, it is conceivable that one could discover at a modest cost all of the data necessary to determine optimal diversity. The principal difficulty concerns transport costs, the function $T(D)$ in the model. Where transport costs are out-of-pocket costs, it is not an overwhelming task to estimate the transport cost function. Where, however, $T(D)$ reflects the opportunity cost of shoppers' time, as in the retail sector, and especially where there is diversity over shoppers' opportunity costs, the task is far from trivial. There are also additional difficult issues concerning multi-purpose shopping. The point is that, even in this relatively simple environment, it is not obvious that we would recognize an optimum if we saw one. Even assuming that we could recognize one (that we could compute the optimum), the presence in any real situation of sunk capital implies that the optimal policy program which takes the market from an initial situation to the optimum is quite complex.

When we interpret the basic model as a model of differentiation in some characteristics space, we believe that we would be quite unable to recognize an optimum if we saw one. In this case $T(D)$ refers to a utility cost associated with the divergence of the characteristics of the best available good from the individual consumer's most preferred bundle of characteristics. Now we face the problem of recovering preferences from observed market behavior. In our basic model this is not an impossibly difficult task – we could, for example, imagine an experiment in which we systematically varied the price of one or more products. Provided that consumers did not play strategic games, the experiment would generate sufficient data for us to discover $T(D)$. There is, of course, no reason to believe that consumers' preferences are diverse in only one dimension, the most preferred good. In the basic model, the function $T(D)$ might very well be consumer specific. In this case, any experiment that would allow us to discover preferences would be far from trivial. It seems clear, however, that experimentation would be necessary in either case – we could not discover preferences from the data revealed in free-entry equilibrium.

The phenomenon of preemption via product proliferation has been raised as a real policy issue in an action by the U.S. Federal Trade Commission [see Schmalensee (1978]. From our perspective, the issue is whether or not this sort of preemptive activity is or is not anti-social. We can use the insights generated by Brander and Eaton (1983) to articulate, but not to answer, the question. Where prices are endogenous, the fewer the number of firms that engage in preemptive product proliferation, the higher will be the welfare losses associated with the divergence of price from marginal cost, but the greater will be actual product diversity. For example, the number of products necessary for a monopolist to preempt entry exceeds the number necessary for duopolists to preempt entry. Whether monopoly preemption is more socially desirable than some other free-entry equilibrium then depends on whether or not the added diversity associated with monopoly preemption outweighs the conventional welfare loss associated with non-competitive pricing. And so on. Thus, the question is well defined. The difficult problem concerns the discovery of consumers' preferences.

There is one exception to these disturbing results. In "large economies" equilibrium is (at least approximately) optimal when all goods are substitutes. See Hart (1979) and Jones (1987) for general results and Eaton and Wooders (1985) for results in an address model. We can illustrate what we mean by a "large economy" in the context of the simple address model we have used throughout the chapter. In that model we can create a large economy by letting the density of customers go to infinity, or by letting the product development cost, I, go to zero. With endogenous prices, prices of all firms approach marginal cost and the number of goods gets arbitrarily large in either asymptotic experiment. Hence, in the limit, every consumer is able to purchase his or her most preferred good at marginal cost. These results raise an important question: How large is large enough, and is this much "largeness" commonly – or ever – encountered empirically?

10. An historical postscript

A brief outline of some of the key points in the historical development of models of product differentiation may help to put the material discussed in this paper into perspective. Before discussing Figure 12.4, which systematizes the main points, we stress that we do not have space to give credit to all of the main

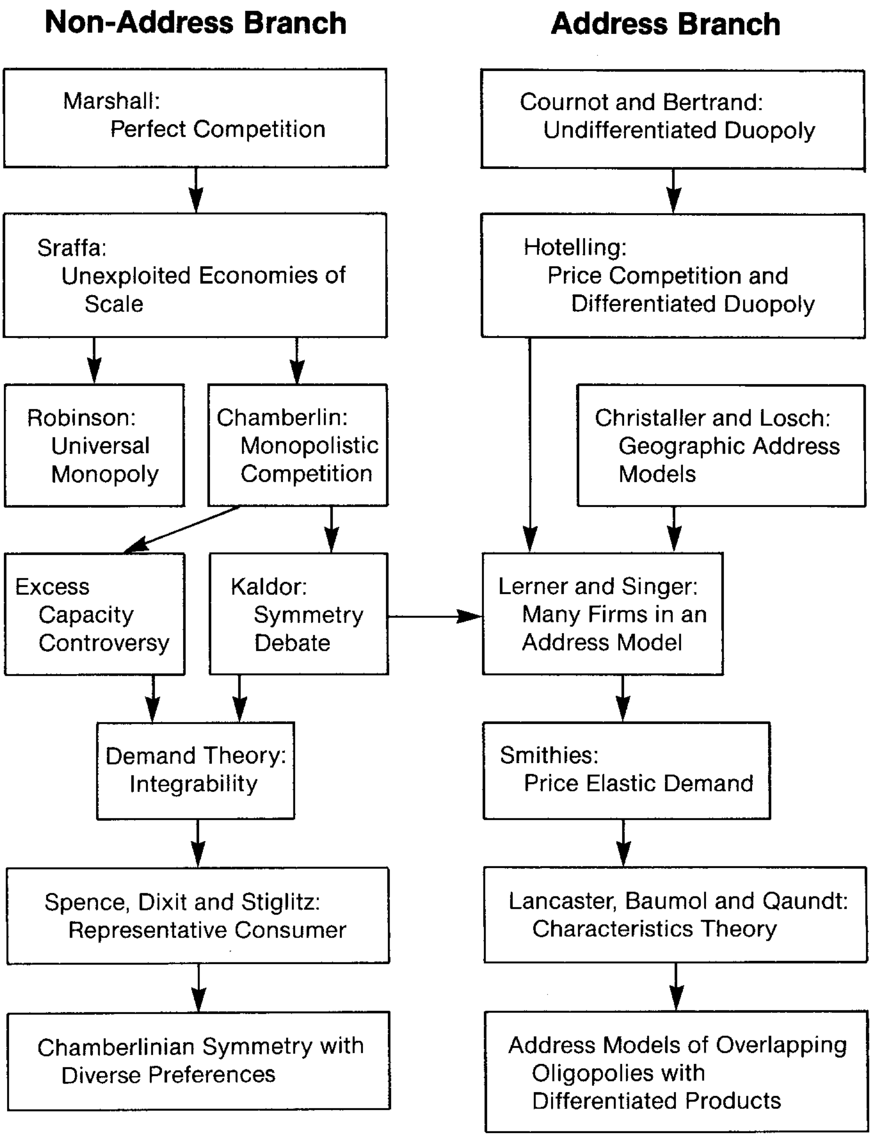

Figure 12.4. Historical perspective.

contributors. Our concern is with the flow of ideas, and we mention names only as illustrative benchmarks.

The left-hand side of the chart shows the development of the large group case of product differentiation. Shortly after the death of Alfred Marshall, Sraffa (1926) pointed out the inconsistency between the observed facts of unexploited scale economies in many manufacturing industries and the Marshallian theory of perfect competition, where all scale economies must be exhausted in long-run equilibrium.

Chamberlin (1933) and Robinson (1934) responded to Sraffa's challenge. Robinson assumed a single monopolist in each industry. Although her work greatly clarified the theory of monopoly, it proved to be a dead end as a response to Sraffa's point. Instead, the way was pointed by Chamberlin. In his theory, a large group of competitive firms, each producing one differentiated product and operating under conditions of free entry, produced an equilibrium where each firm's output was less than minimum efficient scale. The theory was a triumph in making a small amendment – differentiated products – to Marshall's theory of perfect competition, which then reconciled competitive theory with the empirical observation of unexploited economies of scale.

Ironically, the reason for its triumph soon became the greatest cause of concern about the theory. It resolved Sraffa's problem by showing that something very close to perfect competition could be consistent with the observation of unexploited scale economies. However, the presence of unexploited economies of scale, which became known as the excess capacity theorem, gave rise to innumerable controversies in response to its apparent implication of free-market inefficiency.

The controversy over the alleged adverse welfare implications of the excess capacity theorem finally faded away when it became understood that, in a society that values diversity, there is a trade-off between economizing on resources, by reducing the costs of producing existing products, and satisfying the desire for diversity, by increasing the number of products. The optimum diversity occurs when existing products are produced at points to the left of the minimum-efficient scale – therefore, "excess capacity" is *not* necessarily socially inefficient.[4]

A second, and quite different, strand of criticism of monopolistic competition was developed largely by Kaldor (1934, 1935). He attacked Chamberlin's symmetry assumption which made competition *generalized* in that, within one industry no firm had "near neighbors" who bore the main effects of any change in its behavior, and "distant" neighbors who bore smaller effects. Given symmetry, free entry would drive profits to zero (or at least to the small positive amount allowed by the integer problem). Kaldor argued that, although products at one end of the

[4]Bishop (1967) analyzed this issue in a Chamberlinian model. Lancaster (1979) does so in a characteristics model.

product spectrum in a given industry would be close substitutes for each other, they would be poor substitutes for those at the other end of the spectrum. He was intuitively working with an address model where goods are located in some appropriate space of characteristics. He saw competition as *localized*, even within an industry, and argued for a model of overlapping oligopolies rather than for a model of generalized intra-industry competition.[5]

Chamberlin, himself, did make a brief excursion into the area of small numbers competition and the short section on "mutual interdependence recognized" was at the time an original contribution to that theory. Furthermore, Chamberlin (1951) accepted Kaldor's arguments against the symmetry assumption. Nonetheless, his theory of large group competition had by then taken on a life of its own due to the extended and heated nature of the debates mentioned above.

By the 1960s, decreasing attention was being paid to the Chamberlinian model of monopolistic competition. Two reasons are worth mentioning. First, the realization slowly took hold that virtually all industries containing a multitude of differentiated products contained only a few firms [awkward fact (4)]. [See, for example, Markham (1964).] Thus, although the typical, real-world set of differentiated products was a large group, the typical set of competing firms was a small group. Second, growth of interest in location theory showed that localized, rather than generalized, competition was also common in many industries where firms are differentiated by their geographic location. Although, for example, there are many drugstores in a city, each has a few nearby, and many more-distant, neighbors. Here again a model of overlapping oligopolies, rather than one of symmetrically situated monopolistic competitors, seemed more appropriate.

The 1970s saw a revival of interest in all aspects of product differentiation. This was no doubt partly due to the experience of both the EEC countries after the signing of the treaty of Rome and of the GAAT participants during the Kennedy round of tariff reductions. Specialization following on these major tariff cuts among countries of roughly similar per capita incomes did not cause whole industries to close down in some countries and to expand greatly in others. Instead, in each existing industry in each country firms found product niches in which they could compete. So specialization took the form of a reduction of product lines in each country with great expansion of intra-industry, international trade. The increased production runs in each differentiated product afforded substantial reductions in unit costs. As a result the gains from specialization turned out to be substantially more than had been estimated from older constant-cost models of inter-industry specialization.

The outburst of theorizing in demand theory in the 1960s assisted the revival of interest in Chamberlin's model. Ever since the event of the "new welfare

[5]For a survey of all of the aspects of the debate up to 1960 – which is close to the time when the debate subsided – see Archibald (1961).

economics" in the 1940s, economists had worried about the construction of a community welfare function that could be derived from individual utility functions. In the 1960s the integrability literature showed, with standards of rigor not demanded in the 1940s, that under certain specific conditions, the community's demand behavior, and the welfare of its individuals, could be captured in a single community utility function.

In the 1970s Spence (1976a, b), and Dixit and Stiglitz (1977), developed models of monopolistic competition that used the concept of the representative consumer. A further development came with the models of Ferguson (1983), Sattinger (1984), Hart (1985) and Perloff and Salop (1985) who assumed different consumers with different tastes and then generated models of monopolistic competition that displayed the symmetry property.

To understand the development of small group competition with differentiated products we need to begin with Cournot's model of quantity competition between oligopolists producing identical products. Then came Bertrand's formulation of the alternative of price competition showing that non-cooperative behavior would drive price to marginal cost. Bertrand's critique of Cournot opened an important issue that still faces us: What conditions will favor the use of price or quantity as the strategic variable for oligopolistic competition?

The seminal article for the development of theories of competition among oligopolists producing differentiated products was Hotelling's (1929) address branch model. Address models of geographic location trace their lineage to the fundamental work of Christaller (1933) and Losch (1938). Hotelling's starting point was Bertrand's critique of Cournot. He made a crucial change of assumption by letting his duopolists compete to sell a differentiated rather than a homogeneous product.

Hotelling showed that when two competing firms were differentiated from each other, either by having different geographic locations or by producing products differentiated in some one-dimensional characteristics space, price competition could leave price high enough to cover capital costs, thus yielding a stable, long-run equilibrium. Lerner and Singer (1941) expanded Hotelling's model by increasing the number of firms beyond two. Smithies (1941) considered the consequence of altering Hotelling's restrictive demand assumption. Thirty years after its publication, however, only a modest amount of work could trace its lineage back to Hotelling's approach.

As with the other strand, a renewed interest in the address models of product differentiation was aided by developments in demand theory. The model developed by Lancaster (1966) and Quandt and Baumol (1966), in which consumers' preferences are defined over characteristics which themselves are embodied in goods, provided a structure in which the firm's decisions concerning product differentiation could be meaningfully analyzed. The following year Baumol (1967) studied a producer's optimal product design and observed that the new characteristic models provided "a promising approach to a problem that seems previously to have appeared to be intractable". Address models of competition among firms selling differentiated goods first concentrated on a proposition that had been developed from a variant of Hotelling's model, a proposition which Boulding (1966) christened the principle of minimum differentiation. [See Eaton and Lipsey (1975).] In response to that paper, Prescott and Visscher (1977) and Hay (1976) took up the issue of foresightful entry and interest quickly spread to many other issues as well.

References

Anderson, S.P., de Palma, A. and Thisse, J.F. (1988) 'A representative consumer theory of the logit model', *International Economic Review*, 29, forthcoming.

Archibald, G.C. (1961) 'Chamberlin *versus* Chicago', *Review of Economic Studies*, 29:1–28.

Archibald, G.C. and Eaton, B.C. (1988) 'Two applications of characteristics theory', in: G. Fiewel, ed., *Essays in memory of Joan Robinson*. London: MacMillan.

Archibald, G.C. and Rosenbluth, G. (1975) 'The 'new' theory of consumer demand and monopolistic competition', *Quarterly Journal of Economics*, 80:569–590.

Archibald, G.C., Eaton, B.C. and Lipsey, R.G. (1986) 'Address models of value theory', in: J.E. Stiglitz and G.F. Mathewson, eds., *New developments in the analysis of market structure*. Cambridge: MIT Press, 3–47.

Baumol, W.J. (1967) 'Calculation of optimal product and retailer characteristics: The abstract product approach', *Journal of Political Economy*, 75:674–685.

Bishop, R.L. (1967) 'Monopolistic competition and welfare economics', in: R. Kuenne, ed., *Monopolistic competitive theory: Studies in impact*. New York: Wiley, 251–263.

Bonnano, G. (1986) 'Vertical differentiation with Cournot competition', *Economic Notes*, 2:68–91.

Boulding, K. (1966) *Economic analysis*. New York: Harper.

Brander, J.A. and Eaton, J. (1984) 'Product line rivalry', *American Economic Review*, 74:323–326.

Chamberlin, E.H. (1933) *The theory of monopolistic competition*. Cambridge: Harvard University Press.

Chamberlin, E.H. (1951) 'Monopolistic competition revisited', *Economica*, 18:343–355.

Christaller, W. (1933) *Die zentralen Orte in Suddeutschland: Eine okonomisch-geographische Untersuchung uber die Gesetzmassigkeit der Verbrietung und Entwicklung der Siedlungen mit stadtischen Funktionen*. Jena: Fischer.

d'Aspremont, C., Gabszewicz, J.J. and Thisse, J.F. (1979) 'On 'Hotelling's stability in competition'', *Econometrica*, 47:1145–1150.

Deneckere, R. and Davidson, C. (1985) 'Incentives to form coalitions with Bertrand competition', *Rand Journal of Economics*, 16:473–486.

Deneckere, R. and Rothschild, M. (1986) 'Monopolistic competition and optimum preference diversity', discussion paper no. 684, Northwestern University.

Dixit, A.K. and Stiglitz, J.E. (1977) 'Monopolistic competition and optimum product diversity', *American Economic Review*, 67:297–308.

Eaton, B.C. and Lipsey, R.G. (1975) 'The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition', *Review of Economic Studies*, 42:27–49.

Eaton, B.C. and Lipsey, R.G. (1976) 'The non-uniqueness of equilibrium in the Loschian location model', *American Economic Review*, 66:77–93.

Eaton, B.C. and Lipsey, R.G. (1978) 'Freedom of entry and the existence of pure profit', *Economic Journal*, 88:455–469.

Eaton, B.C. and Lipsey, R.G. (1979) 'The theory of market preemption: The persistence of excess capacity and monopoly in growing spatial markets', *Economica*, 46:149–158.

Eaton, B.C. and Lipsey, R.G. (1980) 'Exit barriers are entry barriers: The durability of capital as a barrier to entry', *Bell Journal of Economics*, 11:721–729.

Eaton, B.C. and Wooders, M.H. (1985) 'Sophisticated entry in a model of spatial competition', *Rand Journal of Economics*, 16:282–297.

Ferguson (1983) 'On the theory of demand for differentiated goods', University of Victoria, mimeo.

Gabszewicz, J.J. and Thisse, J.F. (1979) 'Price competition, quality and income disparities', *Journal of Economic Theory*, 20:340–359.

Gabszewicz, J.J. and Thisse, J.F. (1980) 'Entry (and exit) in a differentiated industry', *Journal of Economic Theory*, 22:327–338.

Gabszewicz, J.J. and Thisse, J. (1986) 'On the nature of competition with differentiated products', *Economic Journal*, 96:160–172.

Hart, O.D. (1979) 'Monopolistic competition in a large economy with differentiated commodities', *Review of Economic Studies*, 46:1–30.

Hart, O.D. (1985) 'Monopolistic competition in the spirit of Chamberlin: Special results', *Economic Journal*, 95:889–908.

Hay, D.A. (1976) 'Sequential entry and entry-deterring strategies in spatial competition', *Oxford Economic Papers*, 28:240–257.

Hotelling, H. (1929) 'Stability in competition', *Economic Journal*, 39:41–57.

Jones, L.E. (1987) 'The efficiency of monopolistically competition equilibria in large economies: Commodity differentiation with gross substitutes', *Journal of Economic Theory*, 41:356–391.

Judd, K.L. (1985) 'Credible spatial preemption', *Rand Journal of Economics*, 16:153–166.

Kaldor, N. (1934) 'Mrs. Robinson's 'economics of imperfect competition'', *Economica*, 1:335–341.

Kaldor, N. (1935) 'Market imperfections and excess capacity', *Economica*, 2:33–50.

Koenker, R.W. and Perry, M.K. (1981) 'Product differentiation, monopolistic competition, and public policy', *Bell Journal of Economics*, 12:217–231.

Kreps, D.M. and Scheinkman, J.A. (1983) 'Quantity precommitment and Bertrand competition yield Cournot outcomes', *Bell Journal of Economics*, 14:326–337.

Lancaster, K.J. (1966) 'A new approach to consumer theory', *Journal of Political Economy*, 74:132–157.

Lancaster, K.J. (1979) *Variety, equity, and efficiency*'. New York: Columbia University Press.

Lane, W. (1980) 'Product differentiation in a market with endogenous sequential entry', *Bell Journal of Economics*, 11:237–260.

Lerner, A. and Singer, H. (1941) 'Some notes on duopoly and spatial competition', *Journal of Political Economy*, 45:423–439.

Losch, A. (1938) 'The nature of economic regions', *Southern Economic Journal*, 5:71–78.

Markham, J.W. (1964) 'The theory of monopolistic competition after thirty years', *American Economic Review*, 54:53–55.

Novshek, W. (1980) 'Equilibrium in simple spatial (or differentiated product) models', *Journal of Economic Theory*, 22:313–326.

Perloff, J.M. and Salop, S.C. (1985) 'Equilibrium with product differentiation', *Review of Economic Studies*, 52:107–120.

Prescott, E.C. and Visscher, M. (1977) 'Sequential location among firms with foresight', *Bell Journal of Economics*, 8:378–393.

Quandt, R.E. and Baumol, W.J. (1966) 'The demand for abstract transport modes: Theory and measurement', *Journal Regional Science*, 6:13–26.

Robinson, J. (1934) *The economics of imperfect competition*. London: Macmillan.

Salant, D.J. (1986) 'Equilibrium in a spatial model of imperfect competition with sequential choice of locations and quantities', *Canadian Journal of Economics*, 19:685–715.

Salop, S. (1979) 'Monopolistic competition with outside goods', *Bell Journal of Economics*, 10:141–156.

Sattinger, M. (1984) 'Value of an additional firm in monopolistic competition', *Review of Economic Studies*, 51:321–3332.

Schmalensee, R. (1978) 'Entry deterrence in the ready-to-eat breakfast cereal industry', *Bell Journal of Economics*, 9:305–327.

Schmalensee, R. (1983) 'Econometric diagnosis of competitive localization', working paper no. 1390-83, MIT.

Shaked, A. and Sutton, J. (1982) 'Relaxing price competition through product differentiation', *Review of Economic Studies*, 49:3–13.

Shaked, A. and Sutton, J. (1983) 'Natural oligopolies', *Econometrica*, 51(5):1469–1483.

Singh, N. and Vives, X. (1984) 'Price and quantity competition in a differentiated duopoly', *Rand Journal of Economics*, 15:546–554.

Smithies, A. (1941) 'Optimum location in spatial competition', *Journal of Political Economy*, 49:423–439.

Spence, A.M. (1976a) 'Product selection, fixed costs, and monopolistic competition', *Review of Economic Studies*, 43:217–235.

Spence, A.M. (1976b) 'Product differentiation and welfare', *American Economic Review*, 66:407–414.

Sraffa, P. (1926) 'The laws of returns under competitive conditions', *Economic Journal*, 34:535–550.