Kelvin Lancaster
Columbia University

Competition and Product Variety*

This paper is an analysis of the economic consequences of infinitely variable product specification in the presence of diverse consumer preferences and some minimal degree of economies of scale near the origin and thus might be considered a study of the contemporary high-technology economy. The study covers both the welfare economics and the market structures associated with such an economy. The emphasis throughout is on the degree of product variety which is optimal (in the welfare economics sections) or will be generated by the market and on how the degree of variety differs between different market structures and between the market and the optimum.

Three results of the study which are of general interest are the following:

1. It is never optimal to produce any good at minimum average cost, but always better to increase variety at the expense of average cost when any good reaches this level of output.

2. A structure very similar to that of Chamberlin's monopolistic competition is the "most perfect" market structure that can be generated, being the Nash equilibrium of firms under conditions of perfect information, noncollusion, perfect flexibility, and free and willing entry. Thus, this structure cannot be regarded as "imperfect competition"[1] and is here referred to as "perfect monopolistic competition." The traditional "perfect competition" structure cannot exist under the conditions posited for the

*The background analysis on which the results of this paper are based is set out in detail in Lancaster (1979), which was unpublished at the time of the original presentation of the paper. The author wishes to acknowledge the assistance of the National Science Foundation, grant SOC 75-14252. A preliminary version of the analysis, given in Lancaster (1975), reaches some incorrect conclusions concerning the properties of market equilibria because it did not include provision for "outside goods" (goods outside the product class being considered) without which there are no stable market equilibria.

1. In this sense Chamberlin was correct in asserting that monopolistic competition was not a form of imperfect competition. Since he was not able to handle the analysis of variable product differentiation, he was not able to show that the monopolistic competition structure was inherent in certain types of situation.

(*Journal of Business*, 1980, vol. 53, no. 3, pt. 2)

0021-9398/80/5332-0012$01.50

economy but could be attained by imposing constraints (in particular by requiring all goods to be produced to the same specification within some range), in which case the apparently perfectly competitive structure would result in a welfare loss as compared with monopolistic competition.

3. Monopolistic competition does not lead inevitably to a greater-than-optimal degree of variety—the traditional view[2]—but may give less than optimal variety under some circumstances.

The Analytical Framework

The basic analytical structure is that of a two-sector economy, the sector of special attention consisting of the group of goods within which product differentiation takes place, and the other sector being the rest of the economy. The product-differentiated sector consists of a single group—that is, all goods within it possess the same characteristics—but goods can possess these in variable proportions and do not share any of these characteristics with goods in the rest of the economy. Preferences are assumed to have appropriate separability properties so that choice within the group is independent of choice over the remaining goods. The goods in the rest of the economy are lumped together as "outside goods" (as contrasted with group goods), and prices and specifications of these goods are assumed constant, so that they can be treated as a single aggregate good.

Within the group, the possession of the same characteristics is presumed to imply similarities at the production end, so that a smooth "product-differentiation curve" can be drawn relating the various characteristics quantities that can be produced when embodied in the output of a good of any specification, given a fixed level of resources to be devoted entirely to the production of that single good. The product-differentiation possibilities are assumed to be homothetic, so that if V resource units can give characteristics vector z^1 when used to produce a good of specification (characteristics proportions) S_1, or vector z^2 when used to produce a good of specification S_2, the resources required to produce a characteristics vector kz^1 from good S_1 will be the same as the resources required to produce a characteristics vector kz^2 from good S_2, although the resources in each case will not necessarily be equal to kV since returns to scale are not assumed to be constant. This property is used to define a unit of a good of arbitrary specification as that quantity that can be produced with a unit resource level. The unit level is arbitrary, but relative quantities of goods of different specifications are then clearly defined. These ideas are identi-

2. But Spence has shown that monopolistic competition (in a rather different version from that here) leads to less-than-optimal differentiation when goods are complements; see Spence (1976).

cal with their equivalents given in Lancaster (1975), where the treatment is more expanded.

Once the specification of a good has been chosen and the unit quantity defined as above, the resources required to produce a level of output Q are given by an input function $F(Q)$. It is assumed that resources can be treated as a single aggregate and thus that the input function $V = F(Q)$ is a single valued function, the unique inverse of the conventional production function. In a market context, $F(Q)$ becomes the cost function. The properties of the input or cost function will appear in the analysis mainly as the "degree of economies of scale" parameter θ, defined as

$$\theta = F/QF'. \tag{1}$$

That is, θ is the ratio of average to marginal cost, or average to marginal resource requirement. If $\theta > 1$, there are economies of scale; if $\theta < 1$, diseconomies of scale; and $\theta = 1$ if there are constant returns to scale or output is at minimum average cost with a U-shaped cost curve.[3]

It is assumed throughout that production within the group shows some initial economies of scale, so that $\theta(Q) > 1$ for $0 \leqq Q \leqq Q_o$, where $Q_o > 0$. If there are true increasing returns to scale, θ will be a constant greater than unity, equal to the degree of homogeneity of the production function. In general, θ will be a function of Q, and it is assumed that $\theta' \leqq 0$ everywhere so that the degree of returns to scale does not increase with output.[4] (This turns out also to be a requirement for satisfaction of second-order conditions.) The restrictions on production are consistent with a wide variety of production conditions, including homogeneity of degree greater than one and fixed costs combined with constant or even rising marginal cost, the last giving a U-shaped cost curve. It will be assumed that the production of the aggregate outside good, representing the rest of the economy, is subject to constant returns to scale.

An individual consumer is assumed to have preferences over characteristics of goods within the group. Because of the assumed separability, the choice within the group is essentially independent of the quantities or prices of outside goods, the latter having effects only of an "income" type as far as group characteristics are concerned. If this economy consisted only of a single individual, there would be a particular specification for a good within the group such that the consumer could attain a given utility or welfare level with a minimum use of resources. This specification would be given by the tangency be-

3. See Hanoch (1975) for an extended discussion of relationships of this kind. He uses the term "elasticity of scale" in much the same way as the degree of economies of scale is used here.

4. It is not surprising that economies of scale which themselves increase with scale will cause destabilizing problems.

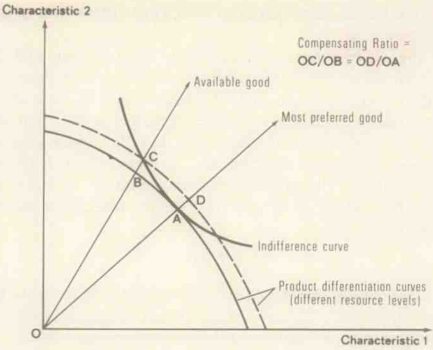

FIG. 1

tween a product-differentiation curve and an indifference curve, and a good of that specification will be referred to henceforth as the "most preferred good" of that consumer. Figure 1 illustrates the idea, and a more detailed treatment is given in Lancaster (1975), where the term "optimal good" is used, a nomenclature that has been changed to avoid overuse of the word "optimal."

A consumer supplied with an arbitrary "available" good, not to the specification of his most preferred good, will require more resources to attain a given welfare level. Since equal quantities of different goods require equal resources (by the quantity definition), a consumer will require a larger quantity of an arbitrary available good than of his most preferred good to attain a given welfare level. The ratio of the quantity of available good to the quantity of most preferred good giving the same welfare level will be called the "compensating ratio" for that consumer with respect to that available good. It will be assumed that preferences can be regarded as homothetic over the relevant range, so that the compensating ratio can be treated as independent of the level of welfare chosen within the range.

Due to the assumed strict quasi concavity of preferences and the assumed shape of the product differentiation curves (nonconvex toward the origin), the compensating ratio will increase as the difference between specifications of the available and most preferred goods increases.[5] Let u be some measure of distance between the two

5. Note that the properties of the compensating function depend on the combined properties of the indifference curves and the product-differentiation curve, as well as the choice of distance measure.

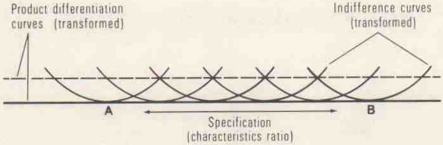

FIG. 2

specifications, so that the compensating ratio can be written as a compensating function $h(u)$. The $h(u)$ will be taken to be strictly convex, and will possess the two boundary properties $h(O) = 1$ (from the definition of the compensating ratio) and $h'(O) = 0$ (because of the tangency at the most-preferred-good specification). Thus $h'' > 0$ everywhere and $h(u) > 1$, $h'(u) > 0$ for all $u > 0$.

If there are many consumers with diverse preferences, as is assumed to be the case, then each consumer will have his own compensating function and his own compensating ratio with respect to any given available good. To bring some order to the potential chaos, the fundamental simplifying assumption of the analysis is made, that of uniformity of the preference spectrum. It is assumed that preferences vary over individuals in such a way that, for a suitable choice of the distance measure between specifications of different good, the compensating ratio for any individual with respect to any available good is the same as that for any other individual with respect to any available good (the same as for the first individual, or different), if the distance in specification between the most preferred goods and the available goods is the same for both individuals. That is, a uniform compensating function $h(u)$ gives the compensating ratio for any individual anywhere on the preference spectrum with respect to an available good at distance u from his most preferred good.[6]

The uniformity assumption is a heroic simplification, but no apologies are made for it. It has some resemblances to, and plays much the same role as, the assumption of a featureless plain in location theory, providing a background of regularity against which variations in other features of the system can be studied.

Implicit in the idea of uniformity is that preferences are, in some sense, similar between individuals except as to specifications of most preferred goods. The preference spectrum can be visualized as made up of geometrically similar indifference curves, shifted in one direction or the other along the spectrum as shown in figure 2. Thus the

6. Uniformity requires a special relationship between the preferences of consumers as one moves across the spectrum and the shape of the product-differentiation curve, as well as a suitable choice of distance measure.

specification of the most preferred good is sufficient identification of an individual or of a group of individuals with identical preferences.

It will be assumed that there is a continuum of preferences—that is, every small region of the spectrum contains individuals having most preferred goods with specifications falling within that region, at least over a portion of the spectrum of possible characteristics combinations which will be called the range of preference diversity.

Uniformity of the spectrum in the above sense refers to the properties of preferences and stipulates nothing concerning the relative numbers of individuals having most preferred goods in different parts of the spectrum or concerning the distribution of incomes or market weights over the spectrum. The central analysis is conducted for the particular case of uniform density of consumers over the spectrum, but this assumption can be relaxed to study the effect of density variations over the spectrum. In any case, "uniformity" and "uniform density" are totally independent and unrelated properties.

Optimum choice for an individual among whatever goods in the group are available depends on a basic property of the consumption technology, namely, whether goods are combinable or noncombinable. If goods are combinable, the individual can obtain characteristics in any proportions by combining goods: He can always make his own most preferred combination in this way. If goods are noncombinable and his most preferred good is not available, he cannot attain his most preferred combination and must settle for the best available good, which will be determined by closeness in specification to his most preferred combination in conjunction with relative prices. The emphasis here is on the analysis of the noncombinable case, which is the most interesting, but the combinable case can also be treated. In the noncombinable case, the individual consumes a single good within the group.

Attention has so far been concentrated on events within the group. As pointed out earlier, however, it is the embedding of the group within an economy that is the most important single contribution of the present analysis. This embedding involves taking account of the role of outside goods, which appear as a single aggregate good. It is greatly simplified by the assumption of uniformity, which makes it entirely reasonable to assume that consumers have a uniform view of outside goods; this is interpreted to mean that the substitutability between outside goods and their most preferred good is the same for all consumers. This does not mean that the substitutability between an arbitrary available good and outside goods is the same for all individuals. On the contrary, the substitutability between the available good and an outside good is compounded of the substitutability between the available and the individual's most preferred good (determined by the properties of the compensating function) and the substitutability between the most preferred good and the outside good, so that outside goods become better substitutes for available group goods as the difference between the specifications of the available and most preferred goods increases. The outside substitution concept is modeled by assuming a constant elasticity of substitution between outside goods and the most preferred group good, the elasticity being the same for all individuals. The relationship of the group to the rest of the economy is then fully parameterized by two parameters, the elasticity of substitution (σ) and a parameter expressing the relative importance of the group in the total, the latter being the ratio of expenditure on group goods to total expenditure when in a market context.

The utility function (w) for an individual given quantity q of an available good which is distant by a measure u from his most preferred good, and quantity y of the aggregate outside good, is given by

$$w(q,y,u) = T\{aq^s[h(u)]^{-s} + (1-a)y^s\}, \tag{2}$$

where $\sigma = -1/(s - 1)$ is the elasticity of substitution and a the parameter expressing the importance of group goods, T being any monotone increasing function.

For a consumer choosing freely in a market context defined by an outside good and a single good having specification distance u from his most preferred good, at income I and group good price P (both in units of outside good), his utility is given by

$$w(I,P,u) = T(I^sP^{-s}h^{-h}m^{1-s}), \tag{3}$$

where m is the proportion of total expenditure devoted to group goods and is related to a but also a function of P, u, and σ.

The Optimum Configuration

Before introducing market analysis, it is necessary to establish the optimum configuration as a reference. It is obvious that if there were constant returns to scale everywhere, an arbitrarily determined distribution of welfare over individuals would always be attained optimally (that is, with the minimum use of resources) by producing every individual's most preferred good. This proposition is inherent in the definition of the most-preferred-good concept.

If there are economies of scale, however, resources can always be saved by producing a smaller number of goods, each at a larger output level. In so doing, there will be individuals whose most preferred goods are no longer available and these individuals will be worse off if given a quantity of the available good only equal to the quantity of most preferred good they received before the elimination of that good from the production schedule. In order to maintain them at the original welfare level, they must be compensated by being given an increased quantity of the available good or of outside goods or both. The compensation will require additional resources in whatever form it is received, and the cost of compensation must be weighed against the resource saving from producing fewer goods. The optimum is determined by the appropriate balance between the saving from economies of scale and the compensation costs.

Before proceeding further, it should be noted that there is an optimal compensation mix for any individual—that is, a mix of available group good and outside goods that requires the least resources for fully compensating that individual. This optimal mix can be determined from his preferences, the distance between the available group good and his most preferred good, and the marginal resource costs of producing the two goods. It will be the mix that he himself would choose if compensated appropriately with additional income and permitted to buy the goods at prices proportional to marginal resource costs. (A special case arises if the group goods are indivisible, like automobiles, so that compensation must be entirely in outside goods and marginal optimal conditions cannot be satisfied.)

The reader will find no difficulty in accepting that a formal analysis of the relationship between the quantity and composition of optimal compensation and the distance u between the available good and an individual's most preferred good will show (1) that the value of the total compensation, at marginal resource cost prices, will be an increasing function of u; and (2) that the ratio of the quantity of outside good to that of the available group good in the mix can be an increasing or a decreasing function of u, depending on the value of the elasticity of substitution.[7]

If there is a continuum of preferences over the spectrum and there is to be a finite number of goods produced, there will be a set of individuals to be supplied with each good. Under the assumption of uniformity over the spectrum, each set will be a segment of the spectrum since if it is optimal (that is, compensation is minimal) to supply any individual whose most preferred good is distant u from a specific available good, it will be optimal to supply any other individual whose most preferred good is at a distance less than u from this good. Given the compensation properties for each individual as a function of the distance u from the available good and the welfare level to be associated with each individual (strictly the welfare density for a small subsegment of the spectrum), the aggregate quantities of both the available good and the outside good required to supply both the "base" and the compensation for all individuals having most preferred goods at a distance up to and

7. If $q(u)$, $y(u)$ are the quantities of group and outside goods required for a consumer at distance u from the available good, the optimal compensation mix is defined by the differential equations $d[\log q(u)]/du = [1 - (1 - m)\sigma]\, d[\log h(u)]/du$, $d[\log y(u)]du = m\sigma\, d[\log h(u)]/du$.

including u from the available good can be determined as a function of u.

For the basic model it is assumed that the welfare density is uniform—that is, if there were no economies of scale so that it would be optimal to supply everyone with his most preferred good, then the target welfare distribution would call for all goods to be produced in equal quantities. With economies of scale this same welfare distribution is to be maintained by compensation. It is easy to appreciate that under a uniform density assumption it will be optimal to divide the spectrum into equal segments, whatever the number of goods, and supply each segment with a good having its specification centered in the segment. If the distance of the edge of the segment from the specification of the available good is Δ, the segment will be of width 2Δ. All segments will be of the same width, so that the optimum can be determined by considering the typical segment.

From the properties of individual compensation and the uniform-density assumption, the aggregate quantity of goods (available group plus outside) required to maintain all individuals in a segment at the base welfare level will increase more than in proportion to the segment half-width Δ and, furthermore, will increase at an increasing rate. While it is certain that the quantity of outside goods required will rise more than proportionally to Δ, the quantity of the group good need not do so.[8] Denote by $Q(\Delta)$ the aggregate quantity of the group good to be provided to all individuals within the segment and by $Y(\Delta)$ the aggregate quantity of outside good, the respective elasticities of Q, Y with respect to Δ being denoted by e_Q, e_Y. Both e_Q and e_Y will vary with Δ, and will depend on the properties of the compensating function $h(u)$, the elasticity of substitution between group goods and outside goods σ, and the relative importance of the group in total expenditure m. While e_Y will never be less than unity, e_Q will be less than unity for high values of σ.

As the width of the segment is varied, the quantities of group and outside goods will change and so will the average resource cost of providing them due to the economies of scale in the production of the group good, even though there are assumed to be no scale economies for the outside good. The optimum configuration will be such that the scale economies and the compensation costs balance each other at the margin.

The formal solution of the optimal problem can be shown to be such that the segment width (the same for all segments) satisfies the optimum equation

$$me_Q(\Delta) + (1 - m)e_Y(\Delta) = m\theta[Q(\Delta)] + (1 - m). \tag{4}$$

8. This is obvious from the differential equations given in n. 7 above.

The meaning of this equation can be appreciated intuitively by considering the effect of a 1% change in the segment width 2Δ. The elasticities e_Q, e_Y then give the percentage change in the group and outside goods, respectively, necessary to maintain welfare for every individual at the constant target level. Since the parameter m represents the proportion of the value of group goods to total goods, the left-hand side represents the percentage change in the value of total goods which is necessary to maintain constant welfare when the segment width is changed by 1%. On the right-hand side, the function θ is the elasticity of the production function for the group good, giving the percentage change in output from a 1% change in resources. Since there are assumed to be constant returns to scale in the production of the outside good, the equivalent elasticity for the outside good is unity. Thus, the right-hand side represents the percentage change in the value of total output at marginal-resource-cost shadow prices and when group and outside goods are produced in a mix of proportions m, $1 - m$, when there is a 1% change in resources.

Thus, when equation (4) is satisfied, a 1% change in segment size will require a change of exactly 1% in resources. That is, the equation represents the segment width at which the total resource use per unit of segment width is stationary. Since the sum of segment widths is equal to the total spectrum (fixed) and all segments are equal in size with equal use of resources, a stationary value of resources per unit of segment width is equivalent to a stationary value of total resource use. Thus, equation (4) gives the segment width at which resources are minimized for the given welfare distribution, subject to second-order conditions being satisfied (which can be shown to be the case).

It can be shown that the left-hand side of (4) is a strictly increasing function of Δ, while the right-hand side is constant or decreasing,[9] so that an optimum solution always exists. It can also be shown that the weighted average elasticity, $me_Q + (1 - m)e_Y$, is always greater than unity[10] (that is, the value of total goods to be provided to a segment always increases more than proportionally to the segment width), so that the right-hand side must be greater than unity. But this necessarily implies that $\theta > 1$—that is, that production of the group good is always at a level at which average resource cost exceeds marginal resource cost, so that the group goods are never produced at the minimum average cost output level even when the cost curves are U-shaped.

Given that the range of the preference spectrum is constant, the optimum segment width is inverse to the number of group goods being

9. Θ is constant if the production function is homogeneous, in which case it is equal to the degree of homogeneity.

10. $e_Y = 1$ at $\sigma = 0$ and increases with σ, while e_Q varies inversely with σ and is equal to unity at $\sigma = 1/(1 - m)$, being less than unity for higher values. Nevertheless, the weighted average is always greater than unity.

produced, and it is convenient to discuss the properties of the optimum configuration in terms of the "degree of product differentiation," the number of goods produced. Obviously, the optimum degree of product differentiation will be a function of the system parameters, including the degree of convexity of the compensating function, the degree of economies of scale, the elasticity of substitution with respect to outside goods, and the relative importance of the group in the economy. The effect of some of these parameters on the optimal degree of product differentiation is both easy to determine and in conformity with intuitive expectations, such as, that the optimum degree of product differentiation will be lower if the degree of scale economies is higher and will be lower if the degree of convexity of the compensating function is lower—that is, if individuals are less sensitive to differences in specification between their most preferred good and the available good.

The effect of other parameters on the optimal degree of product differentiation is much more difficult to determine and the results less intuitive. In particular, the effect of the elasticity of substitution on the optimal degree of product differentiation—which is important for comparing market and optimal solutions—falls into this category. It can be shown that the optimal degree of product differentiation first decreases as the elasticity of substitution increases, to reach a minimum at unit elasticity, and then increases with the elasticity. Although no simple explanation can be given for the full shape of the relationship, the reasons for a high degree of product differentiation at high elasticity are clear enough. When the elasticity of substitution is high, outside goods are good substitutes for group goods so that optimal compensation will result in a rapidly increasing ratio of outside goods to group goods as the segment width is increased. But there are no economies of scale in outside goods, so that the average scale economies over the total goods mix fall rapidly as the segment width increases. (In the extreme case of infinite elasticity of substitution, only consumers for whom the available good meets most preferred specifications will be supplied with that good, and there are no potential gains from economies of scale at all.) Thus the optimal segment widths will be relatively small and the number of products relatively large.

Compensation Problems and the Second Best

The preceding welfare analysis is strictly Paretian in the sense that every individual is maintained on his same indifference contour throughout as the number of goods is varied until the configuration which uses resources to the minimum extent is determined. There is no trading off one individual's welfare against that of another. Although the analysis given for a specific welfare distribution, the principles apply to an arbitrary distribution.

In order to maintain every individual's welfare at a constant level, it is necessary to compensate for specification as the number and thus the specification of available goods changes. Apart from the information problem implicit in knowing the preferences of every single individual and not merely the distribution of preferences in an anonymous fashion, there is a special problem of what can best be termed "manifest equity." Consider two individuals, labeled as I_1 and I_2, whose preferences are such that both will be supplied with the same group good at the optimum. This good happens to be the most preferred good for I_1 but not for I_2, so that I_2 will be compensated by being given more of the good and/or more outside good, even when the desired welfare distribution is one of equality.[11] Now I_1 may well find it difficult to accept that he is not being treated inequitably when he sees I_2 being compensated for receiving the very good that I_1 considers to have an ideal specification. Compensation for specification inevitably results in truly equal treatment appearing to be inequitable.

Because of practical and manifest equity problems associated with the full optimum, it seems desirable to investigate also a second-best optimum in which the constraint is imposed that all individuals are to receive the same income, so that there is no compensation for specification. Since the distribution of welfare then changes as the number of goods is increased or decreased, it is necessary to assume the existence of a social welfare criterion that weighs welfare gains by some individuals against welfare losses by others. Since the uniformity assumption implies that preferences vary over individuals with respect to the specification of the most preferred good but are "similar" in other respects, a naive utilitarian criterion does not seem inappropriate for such a simple model.

It is assumed that individuals supplied with the same quantities of their respective most preferred goods and with the same amounts of outside good will derive the same welfare and that a person supplied with a good which is not his most preferred derives welfare that can be assessed by converting the available good to its "most-preferred-good equivalent" by dividing by the compensating ratio. The criterion used in the analysis is the average welfare per capita derived in this way. The second-best solution is the number of goods which minimizes the resource use per capita for a target level of per capita welfare.

The second-best solution can be shown to give a smaller degree of product differentiation than does the full optimum solution. The intuitive explanation for this is that while the average value of output per head required to maintain a constant level of average welfare per capita

11. By "equal distribution of welfare" is meant the distribution obtained by giving every individual the same quantity of both outside good and most preferred good or most-preferred-good equivalent. Nothing is implied concerning interpersonal comparisons of actual welfare indexes.

rises as the segment width increases, just as does the average output per head required to maintain every individual on a constant welfare level in the full optimum case, it rises more slowly in the second-best case. This is because the effect of the individuals at the margin is partly balanced by averaging with individuals at the center in the second best, but not in the full optimum. Thus, in the second best, the segments will be larger and the number of goods smaller.

Market Demand

In order to lay the foundation for a market analysis under the same general conditions for which the optimal configuration of the economy was determined, it is necessary to derive the properties of market demand from the properties of individual preferences, the distribution of those preferences, and the distribution of income.

The analysis will be confined to the case in which the consumption technology is of the noncombinable kind (within the group), so that the individual will consume only one of the group goods. The individual's decisions then consist of the following: (1) which of the group goods to purchase, (2) how much of that good to purchase, and (3) how much of the outside good to purchase. The individual is assumed to have full information concerning all goods and their specifications, to know his own preferences fully, and to choose freely within the market subject to his budget constraint while taking prices and his income as given.

An individual's choice concerning which of the group goods to purchase will depend only on the specifications of the available good, the specification of his most preferred good, and the relative prices of the available goods. It will be independent of the quantity of that good he will subsequently choose to purchase and, thus, of his income and the price of outside goods. If the group is defined by only two characteristics, so that specification is determined by a single parameter (the ratio of the characteristics or some transform of this), then the spectrum of specifications is a segment of a line. In general, the choice for any particular consumer will be between two goods, those having specifications closest in each direction to the specification of his most preferred good.[12] The quantity of either of the goods which is equivalent to any specified quantity of his most preferred good is given by application of the compensating function. Obviously, he will buy whichever of the goods provides a given most-preferred-good equivalent at the lowest cost. If the two goods have specifications which differ from the specification of his most preferred good by distance measures u_1, u_2, and sell at prices P_1, P_2, the relative costs of attaining the

12. That is, unless the price of a more distant good on the spectrum is so low relative to the closer good that the latter is passed over. In such a case the passed-over good will not be purchased by anyone and will drop out of the market.

equivalent of $\bar{q}$ units of his most preferred good will be $\bar{q}P_1h(u_1)$ and $\bar{q}P_2h(u_2)$, respectively. He will choose that good for which the product $Ph(u)$ is the least and will be indifferent between the goods if P_1, P_2 and u_1, u_2 are related in such a way that $P_1h(u_1) = P_2h(u_2)$.

Once the best available group good has been chosen, the quantity of that good and of the outside good will be determined in the usual way, by maximization of this utility subject to the budget constraint. The utility function is assumed to have the constant-elasticity-of-substitution form given in (2).

Because of the uniformity assumption, the demand functions for all individuals are fully determined by the distance of the chosen group good from their most preferred good (u), and particular individuals do not need to be identified. If q represents the quantity of the chosen group good and y the quantity of outside good, the demand functions for individuals are given by:

$$q(u,P,I) = IP^{-1}\,[1 + AP^{\sigma-1}\,h(u)^{\sigma-1}]^{-1} \tag{5a}$$

$$y(u,P,I) = IAP^{\sigma-1}h(u)^{\sigma-1}\,[1 + AP^{\sigma-1}\,h(u)^{\sigma-1}]^{-1}, \tag{5b}$$

where A is a constant reflecting the importance of the group in total utility.[13]

The market demand for any available good depends on the market width, or the highest value of u for which individuals will buy that good at the going relative prices of group goods, and the market depth or average quantity purchased per unit of market width. The market width depends only on the individual choice concerning which good to buy and thus only on relative prices of goods within the group and the difference in specification between goods in the market. The market depth depends on the individual demand functions (5a) and thus on the price of the good relative to the price of outside goods and on the distribution of incomes over consumers buying the good. The market depth also depends on the market width through the compensating function $h(u)$ which appears in the demand equation.

The market width is determined by the following dividing condition. Denote the price of the jth good by P_j and assume that the goods are numbered in sequence along the spectrum, so that the $(j + 1)$th good is the adjacent good in one direction, sold at price P_{j+1}. If the spacing between the goods is 2Δ, the dividing consumer (the consumer who is indifferent between the goods) will be the consumer whose most preferred good is at distance u from the jth good, where u satisfies the dividing condition:

$$P_jh(u) = P_{j+1}h(2\Delta - u). \tag{6}$$

13. $A = (a/1 - a)^{-\sigma}$. Note that P and I are assumed given in terms of the outside good.

Note that the division between the markets for the two goods depends only on the properties of the compensating function, the difference in specification between the goods (2Δ), and their relative prices. There will also be a dividing condition for the jth good relative to the (j − 1)th, the total market width for the jth good being the sum of the widths of the two half-markets. The analysis here will be confined to symmetrical situations in which the half-markets are identical, but they may differ in width and other respects in the general case, and there will be more than two dividing conditions if there are more than two characteristics. Note also that $P_j = P_{j+1}$ implies $u = \Delta$—that is, the dividing consumer's most preferred good has a specification midway between the specifications of the two available goods when the prices of the adjacent good are the same.

A change in the price of P_j has the following effects: (1) It changes the width of the market by changing the price of the jth good relative to the prices of adjacent goods. (2) It changes the quantity of the good purchased by consumers within the market area by changing the price of the jth good relative to outside goods. (3) It changes the quantity of the good purchased by consumers in the market because of the traditional kind of income effect, unless there is compensation for the price change. (4) It changes quantity through a special kind of effect, the specification effect, unless there is compensation for specification.

The first three effects can be identified as the inside substitution effect, the outside substitution effect, and the income effect, respectively, and they call for no special comment. The fourth, the specification effect, requires some discussion. As the market width increases as a result of a fall in P_j, the additional consumers are buying goods which are even more distant from their most preferred specification than the existing consumers. If money incomes were uniform over the market, the real incomes of these consumers would be less than the average real incomes of the existing buyers because of this specification distance. Thus, a change in the market width will change the average real incomes of consumers in the market even if all receive the same money income and even if that money income is adjusted to compensate for price in the usual way. In addition, the specification effect involves a special kind of price effect because the marginal customers are paying a price $P_j h(u)$ for a unit of most-preferred-good equivalent and thus a higher price for this equivalent than intramarginal customers. The "average price" per most-preferred-good equivalent rises as the width of the market increases, partly offsetting the fall in P_j which causes the increase in width. Note that the specification effect works in the opposite direction to the two substitution effects and the income effect (assuming all goods are normal), tending to reduce demand when price falls. It cannot, however, be greater than the inside substitution effect, so that the sum of the inside substitution and specification effects has the regular sign. If consumers are compensated for specification in the same way as under the full optimum analysis, the specification effect vanishes.

Aggregate demand over the market depends, of course, on the distribution of individuals over the spectrum and on the distribution of incomes over the individuals within the market. It will be assumed for the analysis here that individuals are distributed uniformly over the preference spectrum, and demand will be considered for two different distributions of income over a market area—(*a*) a uniform distribution of money income and (*b*) a distribution of money incomes that represents compensation for specification. In distribution *b*, individuals near the market fringes receive higher incomes than those near the center and all consumers reach the same welfare level after optimal expenditure of their money incomes.

By taking the derivative with respect to u in the individual demand equation (5a), the variation of q with u can be shown to be given by

$$\frac{\partial \log q}{\partial u} = -(1-m)(\sigma-1)\frac{d\log h(u)}{du}, \tag{7}$$

when income is not compensated for specification, and

$$\frac{\partial \log q}{\partial u} = [1-(1-m)\sigma]\frac{d\log h(u)}{du} \tag{8}$$

when there is compensation for specification.[14]

In all the succeeding discussion, $Q(u)$ will denote the integral of q from 0 to u and $e_Q(u)$ the elasticity of Q with respect to u. The properties of $q(u)$ and thus $Q(u)$ and $e_Q(u)$ will be determined by (8) or (7) above, according to whether there is compensation for specification or not. Note that when the elasticity of substitution, σ, is less than unity for demand uncompensated for specification, or less than the value $1/(1-m)$ for demand which is compensated for specification, q is an increasing function of u, Q is a convex function of u, and e_Q is greater than unity. When the elasticity of substitution is greater than unity or $1/(1-m)$, whichever is the relevant value, q is a decreasing function of u, Q is concave, and e_Q is less than unity. At the crossover value—unity or $1/(1-m)$—q is a constant, Q is linear in u, and e_Q is constant and equal to unity.

The market demand function for the jth good depends on the properties of the compensating function, the elasticity of substitution with respect to outside goods, the relative importance of group goods

14. Note that the equation (8) for $q(u)$ is the same as in the optimum mix for the nonmarket analysis; see n. 7 above.

in total expenditure, and the prices and specifications of the $(j - 1)$th and $(j + 1)$th goods,[15] as well as its own price, the existence or otherwise of compensation for specification and/or compensation for price, and the level of income. For the particular case in which the difference in specification between the jth good and each of the adjacent goods is the same, and when the prices of the jth good and adjacent goods are equal, the own-price elasticity of demand is given by

$$E(P,\Delta;\sigma,m) = s^* e_Q(\Delta) + (1 - m)\sigma + (1 - \gamma)m, \tag{9}$$

where the spacing between pairs of adjacent goods is 2Δ and the degree of price compensation by γ, $\gamma = 0$ meaning no price compensation and $\gamma = 1$ meaning full compensation. In the above formulation, the compensating function is assumed given and the elasticity depends on two market variables, P and Δ (since firms can vary both price and specification), and on the system parameters which are varied for comparative static analysis, σ and m.

The first term in (9) represents the combined inside substitution and specification effects, e_Q being chosen for the quantity function $Q(u)$ appropriate to the case with compensation for specification or without, whichever is relevant in the context. The factor s^* needs explanation, since this is its first appearance. For a half-market in which the marginal customer is at distance u from the market center and the next good in that direction is at a distance 2Δ, $s(\Delta,u)$ is the elasticity of u with respect to P multiplied by the ratio u/Δ, the elasticity being taken with the positive sign. In the case in which adjacent prices are equal, $u = \Delta$ and $s^*(\Delta)$ is the elasticity of the market width with respect to price. From the dividing condition (6) it can be shown that

$$s^*(\Delta) = \frac{1}{2} e_h(\Delta), \tag{10}$$

where e_h (Δ) is the elasticity of the compensating function $h(u)$ with respect to u, at the value $u = \Delta$.[16]

The second term in (9) corresponds to the outside substitution effect and the third term to the income effect. If income is fully compensated for price, $\gamma = 1$ and the third term vanishes. All three terms are nonnegative.

It can be shown that, in general: (1) The quantity E varies inversely with the spacing between adjacent goods, becoming infinite when that spacing approaches zero. It is, however, bounded below by the sum of the outside substitution and income effects as the spacing becomes

15. With more than two characteristics there will be more than two adjacent goods.

16. The factor 2 appears because both $h(u)$ and $h(2\Delta - u)$ change, equally but in opposite directions, in response to a change in u.

very large. (2) The quantity E varies with price through second-order effects even though price does not appear explicitly in (9). For values of the elasticity of substitution at least equal to unity when there is no compensation for specification, or at least equal to $1/(1 - m)$ when there is such compensation, the elasticity of demand increases with price.[17] (3) The quantity E increases when the elasticity of substitution increases, provided $m \neq 1$, that is, provided outside goods are actually purchased.

For the analysis which follows, it is convenient to make use of the market parameter R, defined as the ratio of price to marginal revenue in the market and referred to as the marginal revenue ratio. Since $R = [1 - (1/E)^{-1}]^{-1} = E/(E - 1)$, it is a function of E only. Changes in R and E are inversely related, so that (1) the quantity R increases with the spacing between goods, is equal to unity when the spacing becomes zero, and is bounded above for large spacings; (2) the quantity R decreases as price increases, if $\sigma \geqq 1$ or $1/(1 - m)$, whichever is relevant to the compensation pattern; and (3) the quantity R decreases as the elasticity of substitution increases.

Market Structures

Within the general context of the model set out above, the following market structures are investigated:

I. *Single-Product Firms*

Perfect monopolistic competition. This is the structure defined by perfect information on the part of both consumers and firms, full flexibility in choosing and varying specifications of goods by firms, and free and willing entry into the group when profits are positive.

Imperfect monopolistic competition. These structures are determined by noncollusive single-product firms when one or more of the conditions for perfect monopolistic competition are not met. The resulting structures will differ according to which of the "perfect" conditions are not attained—information less than perfect, costly specification changes, or the existence of barriers to entry.

II. *Multiproduct Firms*

Full group monopoly. A single firm possesses monopoly power over the whole group.

Island monopoly. A single firm possesses a monopoly over some part of the spectrum covered by the group but not the whole group. It is

17. For lower values of the elasticity of substitution, the effect of price on the demand elasticity depends on the degree of price compensation.

assumed that the remainder of the spectrum is covered by single-product firms.

Multifirm, multiproduct structures. These structures arise when there are interproduct economies, so that there is a two-part fixed cost, one arising from operating in the group at all, the other associated with each product.

The term "perfect" monopolistic competition is used for the first structure, since this arises from noncollusive behavior among firms under conditions of perfect information and perfect flexibility in a context in which there are economies of scale, variable product specifications, and diversity in consumer preferences. It cannot be regarded as imperfect competition because no more perfect form of competition is possible within the same context.

The equilibrium under perfect monopolistic competition satisfies conditions of the same form as those for the traditional Chamberlin model, except that here the demand properties are fully derived from preferences[18] and the consumption technology and, in particular, the demand for any product is given as a function of its specification relative to the specifications of other goods as well as of prices. The equilibrium conditions can be written in the form

$$P = RF' \tag{11a}$$

$$R = \Theta, \tag{11b}$$

where R is the marginal revenue ratio (defined earlier) and Θ the degree of economies of scale (ratio of average to marginal cost). The first condition is the equality of marginal revenue to marginal cost; the second is the combination of this and the equality of price to average cost. All firms will charge the same prices and will produce goods which are evenly distributed along the spectrum,[19] the spacing being determined from (11b) since both R and Θ (through the quantity produced) are functions of this, given the price relationship of (11a) (see Lancaster 1979, chap. 4).

If consumers are ignorant about the specifications of the goods available, it is assumed they choose at random among the goods. The specifications and distribution of the goods along the spectrum do not appear in the demand functions, but the elasticity of demand (with respect to price) can be considered to be an increasing function of the

18. There are no "perceived" demand curves in this analysis—the demand conditions are true conditions.

19. Note that although this model has many features in common with neo-Hotelling spatial models, it differs in that the compensating function is strictly convex as compared with linear transport costs and that there is smooth substitutability with respect to outside goods. These differences are crucial. See Salop (1977) for a survey of neo-Hotelling models.

number of goods. An equilibrium can be shown to exist, giving the same formal conditions as in (11a) and (11b) but with a different value of R because of the different demand conditions. In general, it can be argued that the value of R will be lower under conditions of ignorance (E higher) and that this implies fewer goods than under perfect monopolistic competition. The specifications of the goods are not determined, and may be chosen in any way by the firms.

If some consumers are informed, firms will distribute themselves evenly over the spectrum (assuming informed consumers to be evenly distributed over the population) and reach an equilibrium which again has the same form as (11a) and (11b), except that the value of R is now between that for perfect monopolistic competition and that for the complete ignorance case.

It can be shown that if consumers are ignorant about the specifications of the goods actually produced but know their own preferences concerning specifications, it will be in the interest of firms to provide the information concerning specification, so ignorance of this kind will be temporary.

The other forms of imperfection—entry barriers and costs of specification change—will lead to fairly obvious effects. Entry barriers will result in fewer firms, and thus goods, than in their absence, while specification-change costs will result in some degree of irregularity in the spacing of firms on the spectrum and, generally, some reduction in the number of goods produced (see Lancaster 1979, chap. 8).

Full group monopoly exists when a single firm has no actual or potential competition with respect to goods produced within the group. It is obvious that it will be optimal for the firm to distribute whatever number of goods it chooses to produce within the group evenly over the spectrum and to charge the same prices for all. The elasticity of demand for the group as a whole, which is the monopolist's elasticity of demand, depends only on the substitution properties for the group with respect to outside goods and on income effects and has no intragroup component. The profit-maximizing condition with respect to price has the form of (11a), but the value of R is greater (because E is smaller) than under monopolistic competition. Solution for the most profitable number of goods in the monopoly case is somewhat complex, but the appropriate condition is given in equation (12) (see also Lancaster [1979], chap. 9):

$$e_u = (R - \Theta)/(R - 1). \tag{12}$$

The right-hand side of (12) is constant except insofar as the degree of economies of scale varies with the output because R does not depend on the spacing between goods in this case. The quantity elasticity e_u does, however, depend on this spacing and thus the relationship determines the most profitable spacing and, thus, the number of goods.

An island monopoly exists when there is a portion of the spectrum into which other firms cannot intrude (because of patent or other protection, presumably), but firms outside this area can certainly compete for the potential customers near the edge of the monopoly segment. The island monopolist can isolate his segment from outside competition by setting up "boundary stakes"—goods at or near the edge of his monopoly area which are sold at the monopolistic competition equilibrium price and which are preferred to all goods beyond the monopoly area by all consumers whose most preferred goods lie within the area, since they are closer in specification and no more expensive. If the monopoly segment is large enough, the monopolist can treat most of the interior as though it were a group monopoly. There will be some edge effects involving goods near the fringes of the area and, if the segment is small, there may no interior portion free of these, but the island monopolist will sell fewer goods at higher prices over his segment than would be the case if the segment were occupied by monopolistic competitors.

Finally, multiproduct firms may exist without explicit entry barriers if there exist interproduct economies so that part of the fixed cost of producing within the group can be spread over several products within the group. Suppose that the economies are such as to be effectively used up over M products, then the group can be expected to consist of firms each producing at least M different products when an equilibrium is reached. There exists, in effect, an entry cost for any new firm which cannot find space for M of its products, but there also exists a limit price for the products in the group above which it will pay a new firm to enter with even a single product. The structure developed under these conditions depends on how the products of the firms come to be distributed over the spectrum. At one extreme, each firm might produce M products which are adjacent on the spectrum, putting it somewhat in the same position as an island monopoly except for the upper-limit price. At the other extreme, every pair of products produced by the same firm might be separated by the products of several other firms, in which case the equilibrium will resemble that of monopolistic competition.

Each of these possible structures will lead to a different degree of product variety, and the next section will consider the relationships between the degree of product differentiation under the various competitive configurations with each other and with the optimum.

Competition and Product Variety

Before discussing the effect of different types of competition on the degree of product variety, there are some features common to all configurations which should be noted.

1. There can be no equilibrium market solution of any kind unless (a) there are outside goods, and (b) the elasticity of substitution between

the group goods and outside goods, σ, is greater than some lower limit which is always at least unity (otherwise second-order conditions are not satisfied) and less than some upper limit which is related to the degree of economies of scale (otherwise even a monopolist cannot break even). The existence of variable product specifications narrows the viability possibilities for the market, as compared with a finite number of goods of given specifications.

2. The equilibrium degree of product variety (and the optimum degree) is a function of the parameters of the system and varies with the parameters, which include: (*a*) the degree of economies of scale; (*b*) the preference relations over the group (compensation function properties), the elasticity of substitution between group goods and outside goods, the relative importance of group goods in total budgets, and whether goods are divisible or indivisible; and (*c*) the width of the spectrum and the size of the total market.

Only the degree of economies of scale and width of the spectrum (difference between the most preferred specifications of the extreme consumers) have the same effect on the optimum and all market configurations, variety being increased by a decrease in the degree of economies of scale or an increase in the width of the spectrum. An increase in the elasticity of substitution sometimes leads to an increase in product differentiation, sometimes to a decrease, depending on the configuration and the parameter values, for example (see Lancaster 1979, chaps. 5, 6, and 9).

The greatest degree of product variety in the market will be reached under perfect monopolistic competition if there are no interproduct economies. Contrary to the received wisdom and to my earlier results (Lancaster 1975), perfect monopolistic competition may, however, lead to a less than optimal degree of product variety if the elasticity of substitution with respect to outside goods is high. All forms of imperfection in the monopolistic competition setting—imperfect information, entry costs, and costs of specification change—will decrease the degree of product variety as compared with perfect monopolistic competition. In the long run, imperfect monopolistic competition will converge to perfect monopolistic equilibrium since entry and specification-change costs wash out eventually, and it will be in the interests of firms ultimately to supply missing information.

Under a full group monopoly, the monopolist will normally provide a variety of products within the group. This variety will always be less than under perfect monopolistic competition and will be less than the full optimum (with compensation for specification) except at $\sigma = 1$, which is the lower limit for equilibrium.[20] The monopolist will, how-

20. This is contrary to my earlier assertion (Lancaster 1975) that the full group monopolist will produce an optimal degree of product differentiation. This is true if there are no outside goods, but then there is no stable monopoly equilibrium because the monopolist's elasticity of demand is less than unity.

ever, produce more variety than appropriate for a second-best optimum (with no compensation for specification) if the elasticity of substitution is relatively small and the group a relatively large sector of the economy.

The island monopolist, whose monopoly power covers only some segment of the spectrum, will produce fewer goods over that segment than would be produced if the segment were under perfect monopolistic competition. As the segment size increases relative to the spectrum, the number of goods per unit segment length will approach that of the full group monopolist, so that the typical island monopolist will provide a degree of variety per unit segment length in between that of monopolistic competition and full group monopoly—it could, in fact, be the optimal variety, but only by coincidence.

Perhaps the most interesting market structure is that which will result from the existence of interproduct economies, in which the economies of scale for a product are reduced if the firm already produces a product in the group. It is eminently reasonable to suppose that some of the components of the fixed cost associated with production within the group (such as technical and market information) can be used over several products, so that the degree of economies of scale (per product) is a decreasing function of the number of products produced by the same firm. Denote by Θ_k the degree-of-economies-of-scale function for each product when the firm produces K products within the group, so that it can be assumed that $\Theta_J > \Theta_K$ if $J < K$, at least up to some limiting value M, where all Θ's refer to the same levels of output.

Such interproduct economies will obviously lead to the emergence of multiproduct firms, even though no monopoly elements are present and entry remains free. If M, the number of products which exhausts the interproduct economies, is small relative to the total number of goods that would be produced under perfect monopolistic competition, then the industry will be characterized by the existence of several firms, each producing several products.

The possible equilibrium states for such an industry depend on how the products of an individual firm are spread over the spectrum. If the industry has grown randomly, as a mixture of new entries and acquisitions within the group, it can be assumed that the products of each firm are scattered over the spectrum and each such product typically has products of other firms as neighbors on the spectrum. However, if each firm produces M products which form a segment of the spectrum, with no other firms product in that segment, the outcome may be different.

Consider first the scattered case, and suppose that if all firms produced a single product there would be N such firms and products, the economies of multiproduct operation remaining unused. Compare this with a situation in which the N products are produced by N/M firms, each of which produces M products and exhausts all interproduct economies. If the products of each firm are scattered, and if the economies affect only fixed costs and not variable costs, the elasticity of demand (and hence R, the marginal revenue ratio) will be the same in both cases, as will marginal costs. Thus the profit-maximizing condition $P = RF'$ will be the same in substance as well as form for both cases, and so will the price. But the firms will make positive profits in the multiproduct case, since they break even on a single product without taking any economies, and new multiproduct firms will be attracted into the group. Ignoring the problem of integer solutions, the equilibrium will have the form

$$R(\Delta_m) = \Theta_M, \tag{13}$$

where Δ_M is the goods spacing with multiproduct firms, as compared with

$$R(\Delta_1) = \Theta_1. \tag{14}$$

Since $\Theta_M < \Theta_1$ and R is an increasing function of Δ, the product spacing will be closer and thus the number of products greater with multiproduct firms than with single-product firms. (Price will certainly be higher with multiproduct firms if the marginal cost curve is falling since $P = RF'$, R is higher, output is smaller, and thus F' is higher. If marginal cost is rising, the price relationship may go either way.)

In the scattered-product case, the multiproduct firms end up behaving rather like monopolistic competition firms, providing the number is not too small. But if the firms can assemble blocs of products which are contiguous on the spectrum, they possess some additional market power. Products in the interior of the spectrum have potential competition from products of other firms only if these enter the bloc, and the retaliatory power of the firm is considerable with respect to such entry, which it can attack not only by varying the price of the good which is closest in specification to the intruder but by varying specifications and prices of other goods close in specification. A new entrant would, in order to gain the economies from multiproduct operation, have either to try to enter a bloc with M products or enter several blocs at the same time. Thus, each multiproduct firm would have some implicit monopoly power within its segment. Contacts with the other firms are essentially limited to the products at the boundary of each segment, minimizing the oligopolistic elements. The overall result of bloc operation in terms of product variety would be to give less product differentiation than with scattered products, certainly more than with true island monopolies, probably more than with single-product firms since the single-product case provides a kind of limit behavior. If the firm's spacing and pricing policy were such as to attract M single-product firms into its segment, it would find this relatively hard to combat.

References

Hanoch, G. 1975. The elasticity of scale and the shape of average costs. *American Economic Review* 65:492–97.

Lancaster, K. J. 1975. Socially optimal product differentiation. *American Economic Review* 65:567–85.

Lancaster, K. J. 1979. *Variety, Equity, and Efficiency*. New York: Columbia University Press.

Salop, S. 1977. Monopolistic competition reconstituted, or circular fashions in economic thought. Federal Reserve Board Research Paper. Washington, D.C.: Federal Reserve Board.

Spence, A. M. 1976. Product differentiation and welfare. *American Economic Review* 66:407–14.