

ON EQUILIBRIUM IN MONOPOLISTIC COMPETITION

Richard Carson
Carleton University

INTRODUCTION

This paper is about equilibrium under monopolistic competition, incorporating the idea that each seller in such a market must have unique, product-specialized inputs whose uniqueness allows it to earn rent, even in long-run equilibrium. The existence of this rent affects our interpretation of equilibrium in a fundamental way. Monopolistic competition requires specialized inputs because some product differentiation is compatible with perfect competition [Rosen, 1974]. If we think of a good or service as a bundle of attributes in the manner of Lancaster [1966, 1971], each different product could be a different combination of the *same* attributes. Perfect competition in the supply of each attribute could then result in perfect competition in the supply of products. Firms would be price takers, even though no two supply exactly the same good or service. It is when each firm imparts a unique attribute to its output—one not exactly duplicated by any other supplier and therefore one which has no perfect substitute—that we leave the world of perfect competition, both in attributes and in products.

In order to supply an attribute that no competitor is able to provide, either a firm would have to be protected by a barrier that gives it a cost advantage in supplying this attribute, or else the advantage would have to come from possession of at least one indivisible input that is specialized to this attribute, and therefore to the firm's product. Since there are no entry barriers under monopolistic competition, each seller must be the sole possessor of one or more specialized inputs. Without these product-specialized inputs, it is hard to explain why monopolistic rather than perfect competition prevails.

In this paper, customers are assumed to demand goods, but we can still think of a differentiated product as possessing unique attributes, features, or properties whose cost-effective supply requires inputs specialized to these unique elements and thus to the product. For simplicity, to "differentiate" a product will mean to impart a unique attribute to it. For example, suppose that each bakery in a city supplies a bread with a distinctive flavor, which can be varied within limits without destroying its singularity. When this unique element is present, every bakery faces downward-sloping demand. Its product-specialized inputs are its bread formula or recipe plus the talent and tacit knowledge required to create the bread from the formula at lowest cost, exclusive of rent. These inputs potentially earn rent that cannot be competed away.

Richard Carson: Department of Economics, Carleton University, C870 Loeb Building 1125 Colonel By Drive, Ottawa, Canada, K1S 5B6. E-mail: rcarson@ccs.carleton.ca.

When product-specialized inputs are mobile between firms, their rents are part of the opportunity costs of the suppliers for which they work, as in the classic Chamberlinian equilibrium [Chamberlin, 1933]. There each supplier's demand curve is tangent to its average cost where both slope downward and output is below the level that minimizes average cost. Chamberlin therefore concluded that a monopolistic competitor would choose production facilities that are below efficient size—producing under increasing rather than constant returns to scale—and would operate these facilities with excess capacity. However, it would be misleading to say that a production facility is below optimal size or has excess capacity if average cost were to slope downward *solely* because of the inclusion of rent in cost. It is then only the specialized input, which in a sense has “excess” capacity. For example, suppose the above bakeries can also supply a non-differentiated or generic bread that does not require a product-specialized input. The generic bread would be supplied under perfect competition, but switching from the generic to the differentiated bread may not change a bakery's long-run equilibrium output at all.

Thus we must be careful to distinguish between cost measured *inclusive* of rent on product-specialized inputs and cost measured *exclusive* of this rent. With this understanding, our basic results are as follows. First, the price, output, and quality of a monopolistic competitor are found by maximizing the difference between its revenue and its cost measured exclusive of the rent on its product-specialized inputs. This maximized difference equals the rent in question. The inclusion of rent in cost then gives rise to the traditional Chamberlinian solution, in which (rent-inclusive) average cost is tangent to demand and therefore downward-sloping. But if rent is excluded, average cost will lie below demand in long-run equilibrium—whenever this rent is positive—and may be constant or even upward-sloping. Moreover, the point of tangency between demand and rent-inclusive average cost does not determine price, quality, or output; rather, tangency occurs at the price, quality, and output at which the rent reaches its maximum.

Suppose we measure economies of scale using rent-exclusive average cost. Then, under an additional assumption to be spelled out and motivated below, a monopolistic competitor's rent-exclusive average cost will be downward-sloping, constant, or upward-sloping in long-run equilibrium, depending entirely on whether differentiating its product increases, leaves constant, or decreases returns to scale in production and supply. This is because differentiating its product does not change its equilibrium output from the case in which it would supply a non-differentiated version of its good or service under perfect competition. Thus when economies of scale in supplying a differentiated product are no greater than those in supplying the non-differentiated version, a monopolistic competitor produces where its rent-exclusive average cost is constant or upward-sloping. Its Lerner index of market power can then be no greater than the equilibrium share of rent in its value added. The smaller is this share—and in this sense, the more intense is the competition facing it—the more closely will its equilibrium approximate that of a perfectly-competitive firm, although at the cost of suppressing the unique element in its product differentiation.

More generally, a monopolistic competitor will set socially optimal quality levels for types of quality change that play no role in product differentiation. However, when

quality change affects the slope of its demand curve, the firm will set a level of quality that leaves this slope flatter than social optimization would require. In this sense, firms do not differentiate their products enough. Thus the two sources of inefficiency in long-run equilibrium are the traditional tendency to price above marginal cost plus the tendency of each firm to under-differentiate its product. Under the additional assumption mentioned above, free entry and exit produces too many firms, the right number, or too few, depending on whether the tendency for greater competition to reduce product differentiation by already-existing firms has a greater, equal, or lesser impact on welfare than the increase in product diversity brought about by the appearance of a new product. Without the tendency to under-differentiate, entry would be socially efficient.

Finally, suppose product-specialized inputs are attached to specific firms, rather than mobile between them. Then rent is not included in opportunity cost, and a firm will earn positive profit in long-run equilibrium. Because of free entry and exit, this profit will equal the rent on its specialized inputs. It will maximize its profit, which is to say that it will find a price, output, and quality combination that maximizes this rent. Provided every firm produces efficiently, the industry will therefore reach the same long-run equilibrium, with respect to outputs, prices, and quality levels, as when product-specialized inputs are mobile between firms.¹ If a specialized input is mobile, firms compete for this input, and the winner must pay the maximized rent in order to bid successfully for the input. Rent is therefore included in opportunity cost. In each case, a monopolistic competitor must set price, quality, and output to maximize the difference between its revenue and its cost measured *exclusive* of rent, and this difference is the rent on its specialized inputs.

THE FIRM IN MONOPOLISTIC COMPETITION

In this paper, firms are allowed to vary quality as well as price and quantity, and monopolistic competition with quality variation is defined by three basic assumptions: [Chamberlin, 1933; Dorfman and Steiner, 1954]. (a). Owing to product differentiation between firms, each firm faces downward-sloping demand when quality is held constant. (b). In the long-run, the industry reaches a Nash equilibrium [Nash, 1951] that is self-enforcing, in the sense that no one firm acting on its own can increase its profit by moving away from it. As a result, we can treat each firm as setting price and quality independently of the others. (c). Because of free entry and exit, including costless mobility of inputs between firms, each monopolistic competitor maximizes profit at zero in long-run equilibrium. This leads to the Chamberlinian outcome, with the properties noted above, which has been controversial [e.g., Demsetz, 1959, 1972; Margolis, 1985, 1989]. I shall outline some of the historical controversy below and show its relation to the present paper, although in order to do this, I must first set up my basic model.

To this end, I shall use capital letters— X , Y , and Z —to designate specific products and small letters to denote output quantities of these products. Suppose that a single-product monopolistic competitor sells x units of a differentiated product, X , on the assumption that the income effects of its output and quality changes are small enough

to ignore. Let $P(x, q, r)$ be the total value of these x units to buyers—meaning the amount paid for them plus the consumer surplus—where P is a third-order continuous function of x , as well as of q , a quality index that also serves to differentiate the firm's product, and r , a quality index that plays no role in product differentiation. More precisely, changes in q affect the slope of the firm's demand curve, whereas changes in r do not. The partial derivative of P with respect to x is then the firm's second-order continuous demand price, $P_x = F(x, q, r)$, and we can write $F(x, q, r) = f(x, q) + g(r)$. If subscripts denote partial derivatives, the cross partial, $F_{xq} = F_{qx} = f_{xq} = f_{qx} \neq 0$.

Let x^* , q^* , and r^* be the long-run equilibrium values of these variables. Because the firm finds it profitable to differentiate its product, in the sense of imparting a unique attribute to it, $f_{xx}(x, q) < 0$ for values of q in some neighborhood of this equilibrium. However, for another value of q —say $q = q_0$ —the firm will supply the generic product, and without loss of generality, we can take $q_0 = 0$. For simplicity, suppose that this firm has just one product-specialized input. Then when $q = 0$, the specialized input adds no value to its good or service, and the firm operates under perfect competition. However, the specialized input does add value in equilibrium. For any given r , increases in q also add value to the bread (or shift its demand curve upward), and therefore $q^* > 0$. Product differentiation is assumed to emerge continuously as q increases from 0. That is, as q tends to zero, so does $f_{xx}(x, q)$. Once a firm imparts a unique attribute to its good or service, its quality index, q , also becomes unique to this attribute. In our bakery example, if two firms set $q = 0$, both supply the generic bread, but when each sets $q = 1$, their breads taste differently, since they are now differentiated.

In general, P_x will depend on price-quality combinations chosen by competitors, as well as on x , q , and r . To reach a Nash equilibrium, however, firms must maximize profit independently of one another. This obliges them to form expectations of the prices, outputs, and quality levels of competitors, which prove to be correct in equilibrium. *Given* these expectations, each firm's demand price depends only on own output and quality.

Besides its product-specialized resource, the firm will use non-specialized inputs, which are assumed to be in horizontal supply to it (although possibly in upward-sloping supply to the industry). Thus in addition to the rent on its specialized input, suppose the firm incurs a cost of $K(x, q, r)$ to differentiate its output. If $C(x, q, r)$ is the rent-exclusive cost of x units of output with quality levels q and r , $K(x, q, r) = C(x, q, r) - C(x, 0, r)$, where $C(x, 0, r)$ is the cost of x units of the generic substitute for X —call it Y —with quality level r , supplied under perfect competition. The average cost of Y , $AC_y = C(x, 0, r)/x$, is assumed to have the traditional U-shape, with a unique minimum. Without the specialized input, the value of x units of output would be $P_y x$, where P_y is the price of the generic substitute, and the contribution of the product-specialized input to value added is therefore $(P_x - P_y)x - K(x, q, r)$.

EQUILIBRIUM QUALITY

The rent on the firm's product-specialized input is the maximum value of $\pi = P_x x - C(x, q, r)$ over q , r , and x . The first-order conditions for maximizing π with respect to q and r are

$$(1) \quad P_{xr}x = C_r = MC_r,$$

$$(2) \quad P_{xq}x = C_q = MC_q,$$

where $P_{xr} = g'(r)$ and $P_{xq} = f_q$ are the partial derivatives of P_x with respect to r and q , and $C_r = MC_r$ and $C_q = MC_q$ are the partial derivatives of C with respect to these quality indices, or the marginal costs of r and q . In order to evaluate equations (1) and (2), let P_r and P_q be the increases in P generated by unit increases in r and q , or the marginal values to buyers of r and q . Then the socially optimal levels of quality occur where $P_r = MC_r$ and $P_q = MC_q$. When $x = 0$, $P_r = P_q = 0$. Using the mean-value theorem, we can therefore write $P_r(x) = P^l_{xr}x$, where P^l_{xr} is an intermediate value of P_{xr} lying between 0 and x . However, $P_{xr} = g'(r)$ is independent of x , so that $P^l_{xr} = P_{xr}$, and equation (1) therefore becomes

$$(3) \quad P_r = MC_r.$$

In long-run equilibrium, the firm sets the socially optimal level of r .

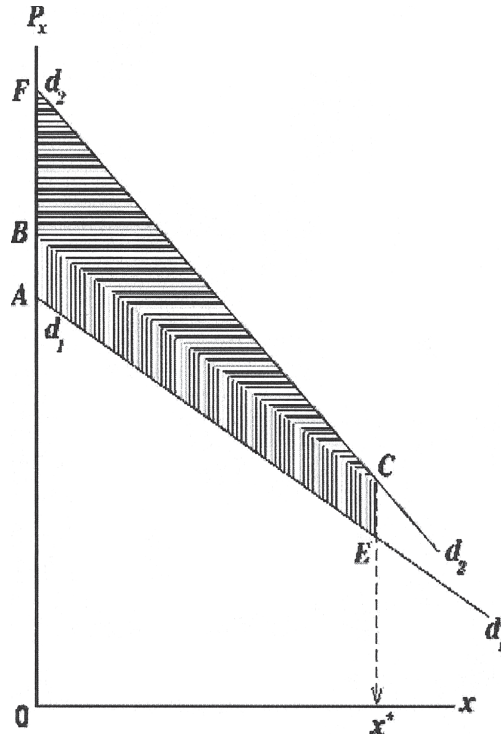
The same is not true of q , however, a result originally derived by Spence [1975]. We again have $P_q(x) = P^l_{xq}x$, but $P^l_{xq} = P_{xq}$ no longer holds as a general rule. To see this, consider the special case in which $P_{xxq} = P_{xqx}$ has the same sign for all $q > 0$. Around $q = 0$, small increases in q must reduce P_{xx} , which is initially zero at $q = 0$, and then becomes negative. Therefore $P_{xxq} = P_{xqx} < 0$; increasing q makes the demand curve steeper. As shown in Figure 1 below, the upward shift in demand from d_1 to d_2 , caused by a unit increase in q , is then relatively great at small values of x . The rectangular, vertically-shaded area, ABCE, is $P_{xq}x$ when $x = x^*$, and the larger trapezoidal area, AFCE, is $P_q = P^l_{xq}x$. The firm ignores the horizontally-shaded area of triangle BFC in setting q because it does not contribute to profit, although it is part of the increase in consumer surplus.

Because $P_{xqx} < 0$, we have $P^l_{xq} > P_{xq}$ or

$$(4) \quad P_q > P_{xq}x = MC_q,$$

and the profit-maximizing level of q is below the socially-optimal level. The problem is that purely infra-marginal shifts of demand do not affect the firm's revenue. From the standpoint of revenue change, the firm sees every quality-induced shift in demand as a parallel shift by an amount equal to the resulting change in P_x at $x = x^*$. However, an upward shift in demand caused by an increase of q can always be thought of as a parallel shift followed by a tilt. In Figure 1, the tilt gives rise to the increase of BFC in consumer surplus, which has no effect on the firm's revenue and is therefore ignored in setting q . Similarly, an increase of q that makes demand flatter—and which is, in general, possible at values of q above 0—amounts to a parallel shift followed by a tilt downward that lowers consumer surplus. Again the tilt is ignored in setting q , and such an increase may therefore be profitable when not socially worthwhile.

FIGURE 1



The net result is a tendency to set q where product differentiation—as measured by the slope of the demand curve—is below the socially-optimal level. When increases in q reduce product differentiation (or make the demand curve flatter), they are over-valued by the firm, in comparison to its customers. When increases in q raise product differentiation, they are under-valued by the firm, in comparison to its customers. In particular, around $q = 0$, we always have $P_{xxq} < 0$, and therefore $P_q > P_{xq}x$. Some firms supplying Y might well differentiate their outputs by imparting unique attributes to them if they were able to capture all of P_q , and the ability to internalize all of P_q could even cause Y to vanish from the market.

Finally, the increase of q from 0 to q^* creates both consumer and producer surplus. It creates the former by differentiating the product, which makes the firm's demand curve steeper. By comparison, a single perfectly competitive firm faces horizontal demand and therefore generates negligible consumer surplus. Similarly, the increase of q creates producer surplus in the form of rent to the product-specialized input, which rent can not be realized if these inputs are used to produce the generic alternative. This increase is therefore welfare-improving, even though q^* is not the socially optimal level of q , and product differentiation causes output to be set where P_x exceeds the marginal cost of x .

EQUILIBRIUM QUANTITY

In fact, the familiar first-order condition for maximizing $\pi = P_x x - C(x, q, r)$ with respect to x is

$$(5) \quad MR_x = MC_x,$$

where $MR_x = P_x + P_{xx}x$ and $MC_x = C_x$ are the marginal revenue and marginal cost of x . Because $P_{xx} < 0$, too little of X is supplied, in the sense that welfare would increase if inputs were transferred from the generic substitute, Y , to X . Let $C^R(x, q, r) = C(x, q, r) + \pi^*$ be the firm's total cost including rent, where π^* is the equilibrium value of π , and note that MC_q , MC_r , and MC_x are also the marginal costs associated with $C^R(x, q, r)$. Thus as long as the firm is viable, equations (1), (2), and (5) are first-order conditions for maximizing either $P_x x - C(x, q, r)$, which is its profit when product-specialized inputs are immobile between firms, or $P_x x - C^R(x, q, r)$, which is its profit when product-specialized inputs are mobile between firms. In long-run equilibrium, $\pi = \pi^*$, and equations (1), (2), and (5) must hold, along with

$$(5a) \quad P_x = AC_x^R,$$

where $AC_x^R = C^R/x$ is the firm's rent-inclusive average cost. Moreover, $P_{xx} = AC_{xx}^R < 0$, where AC_{xx}^R is the slope of AC_x^R . AC_x^R is downward-sloping in long-run equilibrium—as long as product differentiation is profitable—and returns to scale as measured by AC_x^R are increasing.

However, let $AC_x = C(x, q, r)/x$ be the firm's average cost, exclusive of rent. Then AC_x could still be constant or upward-sloping in equilibrium—in which case returns to scale as measured by AC_x would be constant or decreasing—and AC_x^R would then be falling solely because π^* is included in cost. In turn, this would have implications for the firm's own-price elasticity of demand. Let AC_{xx} be the slope of AC_x and $\varepsilon_p = -P_x/P_{xx}x$ denote this elasticity. Since $AC_{xx} = AC_{xx}^R + (\pi^*/x^2) = P_{xx} + (\pi^*/x^2)$, $AC_{xx} \geq 0$ implies

$$(6) \quad \varepsilon_p \geq P_x x / \pi^*,$$

in long-run equilibrium. The firm's elasticity of demand must be greater than or equal to the reciprocal of the share of rent in its value-added if constant or decreasing returns to scale, as measured by AC_x , prevail. Thus if this share is 20%, ε_p must be at least 5; if this share is only 10%, ε_p must be at least 10. Some readers may consider these elasticities to be on the high side, although, in principle, ε_p can take any value between one and infinity.

In order to investigate this issue further, we shall make an additional assumption that will allow us to locate the equilibrium level of x precisely. For this, we turn to the earlier literature on the question of excess capacity under monopolistic competition. As noted above, Chamberlin argued that long-run equilibrium was characterized by excess capacity and production facilities below optimal size. Such conclusions reflect

the envelope theorem plus the fact that AC_x^R is downward-sloping, in the sense that $AC_{xx}^R < 0$. Writing a quarter of a century later, however, Demsetz [1959] challenged this excess capacity result. To understand his argument, let $q(x)$ and $r(x)$ be the profit-maximizing levels of q and r at each given x , and $P_x(x)$ and $AC_x^R(x)$ be defined by $P_x(x) = P_x(x, q(x), r(x))$ and $AC_x^R(x) = AC_x^R(x, q(x), r(x))$. Then $P_x(x)$ and $AC_x^R(x)$ could be tangent where each has a slope of zero, since increases in x could cause increases in q and/or r that would raise P_x and offset the negative direct effect on price ($P_{xx} < 0$) of increases in x .² In this case, Demsetz claimed, no excess capacity would be present. And while capacity should be measured for a specific and well-defined product—which would appear to mean for a fixed q and r —Demsetz argued that increases in q and r could represent higher selling costs, rather than increases in quality *per se*. Moreover, economies of scale in selling costs could cause $q(x)$ and/or $r(x)$ to be increasing functions of x over some ranges of output [Demsetz, 1959, 24–25].

Demsetz' article aroused interest in the subject, but was later criticized by Barzel [1970] and Schmalensee [1972], among others. The thrust of Barzel's criticism was that if buyers are not misled, increases in q and/or r must represent increases in broadly-defined quality for at least some customers. (On this point, see as well Margolis' [1985, 266–270] summary.) Demsetz [1972] accepted this criticism, but then made an entirely new argument to the effect that, in comparing monopolistic to perfect competition, C_x^R is an inappropriate measure of cost. This is because consuming a "branded" rather than a "non-branded" version of a product—analogueous to consuming X rather than Y above—saves the buyer certain costs.³ Demsetz subtracted these costs from C_x^R and claimed that the resulting measure of average cost would be minimized where demand and AC_x^R are tangent. However, if X were to disappear from the market, it is unclear whether replacing units of X with units of Y and incurring these costs would be a utility-maximizing strategy for buyers. Instead, they might be better off simply adjusting their consumption bundles. In place of these costs, I shall assume the existence of a product, Z , introduced below, which will anchor production of X to a specific rate of output.

Subsequently, Barzel and Schmalensee were criticized by Ohta [1977]. Both Ohta and Murphy [1978] tried to re-establish Demsetz' original [1959] conclusions, but in his 1985 survey article, Margolis maintained that efforts by Ohta, Murphy, and Greenhut [1974] to overturn Chamberlin's excess capacity theorem were flawed. Finally, Margolis himself argued [1989] that a multi-product monopolistic competitor using a single brand name does not necessarily operate under excess capacity "if marketing efforts spill over from one product [sold under this brand name] to another."⁴ Of the various challenges to the excess capacity theorem, this one appears to be the most successful, but it relies on the existence of benefits that are external to individual products, although internal to the firm supplying these products. Absent such externalities—which can exist for multi-product firms only—Chamberlin's original claims (made for a single-product firm) have fared better than one might have expected.

In general, this literature did not deal with product-specialized inputs or the rents on these and therefore did not distinguish between rent-inclusive and rent-exclusive cost. Its claims and counter-claims relate to rent-*inclusive* cost. The present paper takes AC_x^R as downward-sloping at the equilibrium value of x —in the sense that $AC_{xx}^R < 0$ —

while maintaining that this may be due entirely to the existence of rent on product-specialized inputs and therefore misleading. The goal here is to find conditions under which rent-exclusive average cost, or AC_x , will be minimized or upward-sloping in equilibrium, in the sense that $AC_{xx} \geq 0$.

With this in mind, we note that $(P_x - P_y) \geq (AC_x^R - AC_y)$ must hold in long-run equilibrium for the firm supplying X , where $AC_y = C(x, 0, r)/x$. Since $P_x = AC_x^R$, $(P_x - P_y) < (AC_x^R - AC_y)$ would imply that $P_y > AC_y$ is possible and therefore that positive economic profit could be earned in supplying Y . The additional assumption mentioned above is then that at least one product, Z , is available, with price, P_z , which meets three conditions. First, Z is supplied under free entry and exit, with the result that $P_z = AC_z^R$ in equilibrium, where AC_z^R is the rent-inclusive average cost of Z . In particular, Z may be supplied under perfect or monopolistic competition. Second, Z can be used together with Y , and this combination competes with X in the market. Z is therefore assumed to be an imperfect substitute for X , and the same is true of other products that compete in the same industry with Z . An expansion of the supply of these substitutes which puts downward pressure on P_z is also assumed to put downward pressure on P_x .

Third, a supplier of Z is at least as cost effective as the supplier of X in adding value to Y . If the firm supplying X would supply Y instead, its average cost would fall to AC_y . Increasing q from 0 to q^* raises the firm's price from P_y to P_x and its rent-inclusive average cost from AC_y to AC_x^R . Thus suppose the supplier of X is producing any output at which AC_x^R is downward-sloping and that demand is such that $P_x = AC_x^R$. Then to meet the third condition, a supplier of Z must be able to produce an output of its own at which

$$(7) \quad AC_z^R \leq (AC_x^R - AC_y),$$

when output units of Z are scaled in such a way that $P_z = (P_x - P_y)$. Returning to our bread example, suppose an entrepreneur markets a spread, Z , that improves the taste of the generic bread, Y , but clashes with the taste of the differentiated bread, X . As a result, the combination of generic bread with spread competes with the differentiated bread.

When the third condition is met and the Y -industry is in long-run equilibrium, a supplier of X must produce on a scale that is efficient for Y in order to cost compete with the composite good whose units consist of a unit of Z plus a unit of Y and whose average cost equals $(AC_z^R + P_y)$. Any other output puts X at a cost disadvantage. Thus suppose x is set where $P_x = AC_x^R$ and $(P_x - P_y) > (AC_x^R - AC_y)$ or, equivalently, $AC_y > P_y$. But then long-run equilibrium can not prevail, since equation (7) implies that $P_z > AC_z^R$ is possible for any product, Z , meeting the above three conditions. When $P_z > AC_z^R$, the prospect of quasi-rents will attract entry of new competitors into the market in which Z competes, and this will put downward pressure on P_z and on P_x , without affecting AC_x^R or AC_z^R , since this entry would be foreseen. By the time $P_z = AC_z^R$ is reached,

competitive pressures on X will have increased, and $P_x < AC_x^R$ must hold. The supplier of X will have to change its output or exit the market.

Thus if this firm survives over the long run, it must produce where $(P_x - P_y) = (AC_x^R - AC_y)$. If MC_y is the marginal cost of Y , x^* is the output at which

$$(8) \quad P_y = MC_y = AC_y,$$

or at which $AC_y = C(x,0,r)/x$ reaches its minimum over x . When the above three conditions hold, differentiating a monopolistic competitor's product does not change its equilibrium output, since a firm supplying Y also produces where AC_y reaches its minimum. In addition, this is the only output at which the value added by the product-specialized input equals this input's rent—that is at which $[(P_x - P_y)x - K(x,q,r)] = P_x x - C(x,q,r)$.

If there is no product with the properties of Z , long-run equilibrium can occur where $(P_x - P_y) > (AC_x^R - AC_y)$. But then there is an incentive to discover and market such a product, which will be viable over the long run and earn quasi-rent in the short run, as long as $(P_x - P_y) > (AC_x^R - AC_y)$, which implies $P_z > AC_z^R$. This motivates the assumptions behind equation (7). If Z is also supplied under monopolistic competition, with its own generic alternative, W , the above analysis is potentially symmetrical in Z and X . That is, if we reverse the roles of X and Z and substitute W for Y —and if the third condition then holds— z^* will be where AC_w reaches its minimum.

Finally, while the existence of Z allows us to anchor production of X to the specific equilibrium output where AC_y is minimized, this existence is not necessary to the result that equilibrium under monopolistic competition may occur where a seller's rent-exclusive average cost is minimized or upward-sloping. Even without Z , the rent on the firm's specialized input may be maximized in this range.

FURTHER RESULTS

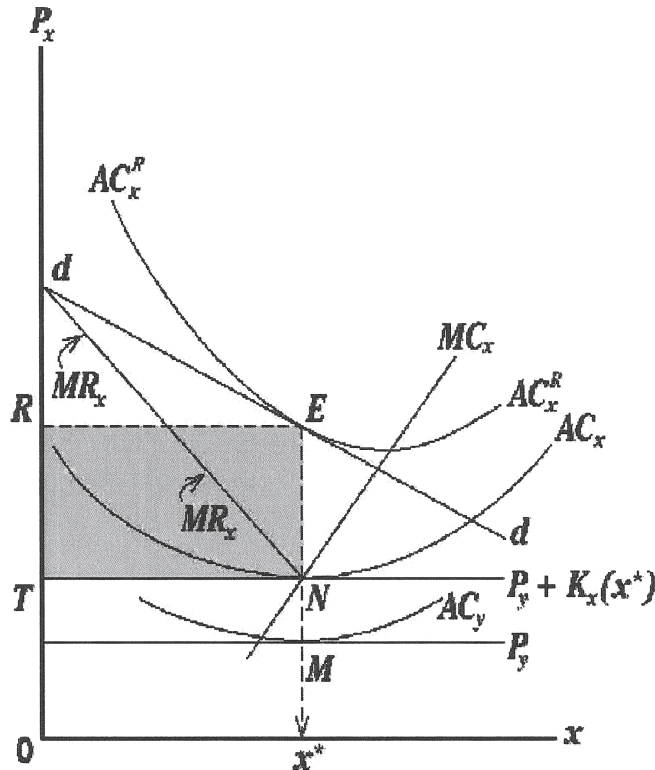
For the remainder of this paper, we measure returns to scale using AC_x , the rent-exclusive average cost of the monopolistic competitor supplying X . As just shown, when equation (7) holds, differentiating this firm's product does not change its equilibrium output. Thus it produces where returns to scale are increasing, constant, or decreasing, depending entirely on whether differentiating its product increases, leaves constant, or decreases returns to scale in production and supply at the value of x that minimizes AC_y . In particular, if product differentiation does not alter returns to scale, these will be the same for X as they are for Y at any given output. AC_x and AC_y will both reach their minima at $x = x^*$, where we consequently have

$$(9) \quad AC_x = MC_x.$$

Constant returns to scale in the supply of X prevail. This is shown at N in Figure 2 below, which is drawn with q and r fixed at their long-run equilibrium values. There AC_x is tangent to the horizontal line, $P_y + K_x(x^*)$, where $K_x = (MC_x - MC_y)$ is the marginal cost of product differentiation. The firm's price-quantity combination is at E, where its demand is tangent to AC_x^R . Area RENT gives the rent earned by its product-

specialized input. However, it should be stressed that equation (9) holds only when product differentiation does not alter returns to scale, whereas equation (8) holds whenever inequality (7) does.

FIGURE 2



More generally, suppose that economies of scale in supplying the differentiated product, X , are less than or equal to those in supplying the non-differentiated version, Y . In long-run equilibrium, the firm supplying X will then produce where AC_x is constant or upward-sloping ($AC_{xx} \geq 0$), and it must be earning positive rent, inasmuch as AC_x^R and AC_x are clearly different. Because $MC_x \geq AC_x$ also holds, inequality (6) puts a floor on the firm's own-price elasticity of demand equal to the reciprocal of the share of rent in its value added. Its Lerner index of market power ($= (P_x - MC_x)/P_x$), which is the same as $1/\varepsilon_p$ in equilibrium, can be no greater than this share (or no greater than $\pi^*/P_x x$). Thus the more intense is the competition facing a monopolistic competitor—in the sense that the greater is the pressure on its rent—the smaller will be its index of market power and the more closely will its equilibrium approximate that of a perfectly-competitive firm.

The latter outcome is assured by the fact that, regardless of rent, the monopolistic competitor produces where $AC_y = C(x, 0, r)/x$ reaches its minimum, which is also where demand is tangent to AC_x^R . As rent disappears, AC_x^R tends to AC_x , but production

never occurs where AC_x is downward-sloping, and demand can not be upward-sloping if income effects are small enough to ignore. Thus the limit must be where demand is horizontal, but this implies that the unique attribute in product differentiation has disappeared and that q^* has tended to zero. As a result, the limiting outcome is at M in Figure 2, where AC_y is minimized and equal to both P_y and MC_y . The Lerner index tends to zero, but the unique attribute in product differentiation is eliminated, and this is, in general, not consistent with efficiency, owing to the inability of suppliers to capture all of P_q .

Only if product differentiation increases returns to scale will the firm supplying X produce where AC_x is downward-sloping, when equation (8) holds. This could happen, e.g., if differentiating the product increases the ratio of fixed or set-up to variable costs at x^* .⁵ Then it is possible that the smaller is $\pi^*/P_x x$, the more closely will the long-run equilibrium approach a limiting outcome where q^* is positive, and demand and AC_x are tangent and downward-sloping. However, such a solution can arise only when product-specialized inputs earn no rent and scale economies in supplying the differentiated product are greater than those in supplying its generic substitute. This limiting outcome has the advantage of preserving the unique element in each firm's product differentiation, although it has the disadvantage that price exceeds marginal cost.

In order to consider the welfare effects of entry of new competitors into this industry, we return to the more natural assumption that differentiating a product by giving it a unique attribute reduces returns to scale or leaves them the same. Suppose then that a new entrant marginally reduces the demand for X (or the demand facing a previously existing supplier). If this would cause an output decrease of Δx , the loss to consumers would be roughly $P_x \Delta x$, whereas the cost saving is $MC_x \Delta x$, leaving a net welfare loss of $(P_x - MC_x) \Delta x$ that the entrant does not take into account in making its entry decision. This is sometimes called the "business-stealing" effect. Entry also adds a new product, which in of itself increases product variety, but may cause quality changes in products already supplied. The net impact of these is a second effect on welfare—the "product-diversity" effect—that the entrant fails to capture. When customers value variety, the number of competitors will be greater or less than the optimal number, depending on whether these business-stealing and product-diversity effects produce a net decrease or increase in welfare [Mankiw and Whinston, 1986]. However, there is no business-stealing effect for Y or for any good that is priced at marginal cost (including r), nor is there any for products whose output remains constant in the face of entry.

Therefore, would entry cause the equilibrium value of x to change? Suppose that there are no long-run fixed or set-up costs in supplying Y , or else that entry does not change relative input prices. In either case, the output that minimizes $C(x,0,r)/x$ for an existing supplier will be invariant to entry of new competitors. Thus, unless the firm is forced out of this market, entry would not change the equilibrium value of x , as long as inequality (7) holds before and after entry. And while entry may cause firms to exit the industry, these are most likely to be marginal suppliers already pricing at or near marginal cost, especially since firms operating under monopolistic competition are small relative to industry size. Entry could steal business from firms outside the industry—and/or cause the supply of complementary goods to expand—but if the net

welfare gain or loss resulting from this is small, the same will be true of the business-stealing effect.

As a result, entry would then be socially efficient if suppliers were able to capture all of P_q —or if q were fixed, as in the classic paper by Dixit and Stiglitz [1977]. Inefficient entry arises only because in long-run equilibrium, $P_q \neq MC_q$. A given entrant may cause q to rise or fall, but as rents are squeezed by greater competition, there is a tendency for q to tend to zero, and therefore for the unique element in product differentiation to decline and eventually vanish. Thus whether and how entry into this industry departs from the socially optimal level depends on whether and how the tendency for greater competition to reduce product differentiation by already-existing firms has a greater or lesser impact on welfare than the increase in product diversity brought about by the appearance of a new product with a new unique attribute. However, if entry changes relative input prices *and* there are positive fixed or set-up costs in supplying Y , entry could alter the ratio of fixed to variable cost at any output and thereby change the output at which $C(x,0,r)/x$ reaches its minimum. This output could either rise or fall, but the resulting effect on welfare would not be taken into account by an entrant, and the tendency to produce where $P_x > MC_x$ would also be a source of inefficient entry.

CONCLUSION

To sum up, the price, output, and quality of a monopolistic competitor are determined by maximizing the difference between its revenue and its cost, where cost is measured exclusive of the rent on its product-specialized inputs. Such a firm must have unique inputs that are specialized to its unique product—since product differentiation is otherwise compatible with perfect competition—and the uniqueness of these inputs allows them to earn positive rent, even in long-run equilibrium. The inclusion of rent in cost gives rise to the traditional Chamberlinian solution, in which (rent-inclusive) average cost is tangent to demand and therefore downward-sloping. But if rent is excluded, average cost may be constant or even upward-sloping in equilibrium, and in this sense, monopolistic competition need not give rise to excess capacity or to production facilities that are too small.

Perhaps the point to emphasize in closing is that differentiating a monopolistic competitor's product by increasing q from 0 to q^* creates both consumer and producer surplus. The increase of q therefore improves welfare, even though q^* is not the socially optimal level of q , and product differentiation prevents marginal-cost pricing of output. To restore marginal-cost pricing by suppressing this product differentiation would reduce welfare and, in particular, would reduce the productivity of the product-specialized inputs by destroying the consumer and producer surplus that they create. This surplus can only be realized when firms are allowed to differentiate their products, not when they supply non-differentiated or generic alternatives.

In addition, when equation (7) holds, differentiating a firm's product does not alter its equilibrium output. If economies of scale, as measured by rent-exclusive cost, are no greater in supplying the differentiated product than those in supplying a generic version, a monopolistic competitor produces where AC_x is minimized or upward-sloping,

in the sense that $AC_{xx} \geq 0$. In this case, there is also no excess capacity, except for the capacity of the product-specialized input itself, and returns to scale as measured by AC_x —although not as measured by AC_x^R —are constant or decreasing. In short, conventional criticisms of monopolistic competition are often formally correct, but they ignore the most important thing about it, namely that when all economic actors are rational and knowledgeable, it improves welfare by creating variety—and the consumer and producer surplus that goes with this—without necessarily reducing output at all. The story these criticisms tell, as Schumpeter [1950, p. 86] once noted in a related context, “is like *Hamlet* without the Danish prince.”

NOTES

I wish to thank Erwei Yin, Han Li, and two anonymous referees for assistance or helpful comments. Sole responsibility for errors rests with me. An earlier version of this paper was presented at the Southwestern Economic Association meetings in Corpus Christi on March 17, 2004.

1. An industry operating under monopolistic competition may have multiple equilibria. In this case, the equilibria will be the same regardless of whether product-specialized inputs are mobile between firms, provided all firms produce efficiently.
2. In terms of earlier notation, the slope of $P_x(x)$ is $P_{xx} + P_{xq}q'(x) + P_{xr}r'(x)$, where $q'(x)$ and $r'(x)$ are the derivatives of $q(x)$ and $r(x)$ with respect to x . We have seen that both P_{xq} and P_{xr} are positive if both q and r are positive (equations (1) and (2) above), and $q'(x)$ and/or $r'(x)$ could be positive owing to scale economies in supplying q and/or r . Thus $P_{xx} + P_{xq}q'(x) + P_{xr}r'(x)$ could be non-negative. Since $P_x(x)$, $q(x)$, and $r(x)$ are the profit-maximizing values of P_x , q , and r at each given x , $P_x(x)$ is the same as Demsetz' [1959] *mutatis mutandis* average revenue curve, *MAR*.
3. Demsetz [1972] only sketches his argument, but his basic idea is that, in comparing monopolistic with perfect competition, C_x^R is an inappropriate measure of cost. As noted below in the text, this is because consuming a “branded” rather than a “non-branded” version of a product—analogueous to consuming X rather than Y above—saves the buyer certain costs. These are “costs that would need to be incurred to ascertain the quality of the product, to establish prestigious consumption in some way other than by consuming branded commodities, and to be confident of clear lines of responsibility should the product be defective in some respect. The consumer can reduce these costs by purchasing differentiated products, since product homogeneity makes it more difficult both to discern clear lines of responsibility for product quality and to consume conspicuously.” [595–596] However, if X were to disappear from the market, it is unclear whether replacing units of X with an equal number of units of Y and incurring these costs would be a utility-maximizing strategy for buyers. Instead, they might be better off simply adjusting their consumption bundles. Moreover, it is not obvious that product homogeneity must make it more difficult to discern clear lines of responsibility for product quality. Instead of these costs, I assume the existence of the product, Z , introduced below.
4. Margolis [1989, 199].
5. Write $C(x,q,r)$ as $C = B(q,r) + V(x,q,r)$, where B is fixed cost—or cost that is independent of output—and $V(x,q,r)$ is cost that varies with x . We then have $MC_x = V_x$, the partial derivative of V with respect to x . The returns to scale embedded in C —or the firm’s elasticity of production—are given by $S = AC_x/MC_x = C/xC_x = (B + V)/xV_x = (R + 1)S_v$, where $R = B/V$ is the ratio of fixed to variable cost, and $S_v = V/xV_x$ are the returns to scale embedded in V . Thus if increases in q raise $(1 + R)$ or S_v , they will also raise S , provided they do not reduce the other of these two components by a larger percentage amount. Intuitively, however, it seems more natural to assume that differentiating a product by giving it a unique attribute would reduce economies of scale, or at least not increase them, although all three outcomes are possible.

REFERENCES

- Barzel, Y.** Excess Capacity in Monopolistic Competition. *Journal of Political Economy*, September/October 1970, 1142–1149.
- Chamberlin, E.** *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press, 1933.
- Demsetz, H.** The Nature of Equilibrium in Monopolistic Competition. *Journal of Political Economy*, February 1959, 21–30.
- _____. The Inconsistencies in Monopolistic Competition: A Reply. *Journal of Political Economy*, May/June 1972, 592–597.
- Dixit, A. K. and Stiglitz, J. E.** Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, June 1977, 297–308.
- Dorfman, R. and Steiner, P. O.** Optimal Advertising and Optimal Quality. *American Economic Review*, December 1954, 826–836.
- Greenhut, M. C.** *A Theory of the Firm in Economic Space*. Austin: University of Texas, 1974.
- Hotelling, H.** Stability in Competition. *Economic Journal*, March 1929, 41–57.
- Lancaster, K.** A New Approach to Consumer Theory. *Journal of Political Economy*, April, 1966, 132–159.
- _____. *Consumer Demand: A New Approach*. New York: Columbia University Press, 1971.
- Mankiw, N. G. and Whinston, M. D.** Free entry and Social Inefficiency. *Rand Journal of Economics*, Spring 1986, 48–58.
- Margolis, S.** The excess Capacity Controversy: A Critique of Recent Criticism. *Economic Inquiry*, April 1985, 265–275.
- _____. Monopolistic Competition and Multiproduct Brand Names. *Journal of Business*, April 1989, 199–209.
- Murphy, M.** The Consistency of Perfect and Monopolistic Competition. *Economic Inquiry*, January 1978, 108–112.
- Nash, J.** Non-Cooperative Games. *Annals of Mathematics*, July 1951, 286–295.
- Ohta, H.** On the Excess Capacity Controversy. *Economic Inquiry*, April 1977, 153–165.
- Rosen, S.** Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, January/February 1974, pp. 34–55.
- Schmalensee, R.** A Note on Monopolistic Competition and Excess Capacity. *Journal of Political Economy*, May/June 1972, 586–591.
- Schumpeter, J. A.** *Capitalism, Socialism, and Democracy*. New York: Harper and Row, 1950.
- Smithies, A.** Optimum Location in Spatial Competition. *Journal of Political Economy*, June 1941, 423–439.
- Spence, A. M.** Monopoly, quality, and regulation. *Bell Journal of Economics*, Autumn 1975, 417–429.