MARKETS WITH CONSUMER SWITCHING COSTS*

PAUL KLEMPERER

Ex ante homogeneous products may, after the purchase of one of them, be ex post differentiated by switching costs including learning costs, transaction costs, or "artificial" costs imposed by firms, such as repeat-purchase discounts. The noncooperative equilibrium in an oligopoly with switching costs may be the same as the collusive outcome in an otherwise identical market without switching costs. However, the prospect of future collusive profits leads to vigorous competition for market share in the early stages of a market's development. The model thus explains the emphasis placed on market share as a goal of corporate strategy.

I. INTRODUCTION

In many markets consumers face substantial costs of switching between brands of products that are ex ante undifferentiated. There are at least three types of switching costs: transaction costs, learning costs, and artificial or contractual costs.

There may be transaction costs in switching between completely identical brands. Two banks may offer identical checking accounts, but there are high transaction costs in closing an account with one bank and opening another with a competitor. Similarly, it is costly to change one's long distance telephone service, or to return rented equipment to one firm and rent identical equipment from an alternative supplier.

The learning required to use one brand may not be transferable to other brands of the same product, even though all brands are functionally identical. A number of computer manufacturers, for example, may make machines that are functionally identical, but if a consumer has learned to use one firm's product line and has invested in the appropriate software, he has a strong incentive to continue both to buy machines from the same firm and to buy software compatible with them. Similarly, when choosing a cake mix, it is easiest for a consumer to buy the brand that he already

*This paper is a shortened version of a chapter of my Ph.D. thesis. Research for this paper was begun at Stanford University in the summer of 1983, and was supported there by the Stanford Graduate School of Business and the Center for Economic Policy Research. I am grateful to Doug Bernheim, Adam Brandenburger, Tim Bresnahan, Jeremy Bulow, David Kreps, Meg Meyer, James Mirrlees, Barry Nalebuff, John Roberts, Larry Summers, and the referees for help and advice.

knows exactly how to prepare, even if he knows that all brands are of identical quality.[1]

These two types of switching costs reflect real social costs of switching between brands, even though their size can be influenced by firms (by their product-design choices, for example). A third type of switching cost, artificial or contractual switching costs, arises entirely at firms' discretion, and is distinguished by the absence of social costs of brand switching. Airlines enroll passengers in "frequent-flyer" programs that reward them for repeated travel on the same carrier, some retailers offer trading stamps (Greenshield stamps, S & H Green Stamps, etc.) that customers can exchange for prizes after sufficiently many have been accumulated from repeated purchases, and many grocery products are sold with a discount coupon valid for the next purchase of the same item. Similar switching costs can be created by contracts. (If a customer signs a contract committing himself to either buy from a firm or pay damages s, this is exactly equivalent to paying s for a discount coupon of value s.) In these examples, consumers who switch between different companies are penalized relative to those who remain with a single firm.

In all these markets rational consumers display brand loyalty when faced with a choice between functionally identical products.[2] Products that are ex ante homogeneous become, after the purchase of one of them, ex post heterogeneous.[3]

In this paper we take switching costs as exogenous, and

1. If a consumer must purchase a brand to discover whether or not it is suitable for him (that is, each brand is an experience good in the terminology of Nelson [1970]), then, in expectation, the consumer faces a loss of utility from switching from a brand that he has tried and liked to an untested rival. In this case, however, there are additional complications due to the possibility of prices being used as signals of quality, and due to the existence of groups of consumers who tried a brand and did not like it.

Our model can apply to markets in which consumers' preferences are shaped by brand-specific knowledge acquired from exposure to other consumers' purchases, as well as to markets in which consumers repeat-purchase. There is, however, an important difference between a model of switching costs and models of learning on the demand side in which advertising or other nonconsumer-specific learning raises *all* consumers' willingnesses to pay for a particular brand (see, for example, Spence [1981] and Clarke et al. [1982]): with switching costs the distinction between those consumers who have and who have not previously bought or been otherwise exposed to a particular brand is critical, and so the concept of a customer base is important.

2. There may also be psychological costs of switching between brands, as when, for no obvious economically rational reason, consumers repeat-purchase from simple habit or loyalty.

3. Corporate planning models often rest heavily on the assumption that the future demand for a product will be positively correlated with its present sales. This correlation of demands across time is an immediate implication of a model with switching costs.

examine their implications for the competitiveness of markets. We make two main points.

First, switching costs make each individual firm's demand more inelastic and so reduce rivalry. Switching costs segment the market into submarkets. Each submarket contains consumers who have previously bought from a particular firm and may in effect be monopolized by that firm. *The resulting (noncooperative) equilibrium may be the same as the collusive solution in an otherwise identical market with no switching costs.* In a standard differentiated products model the social cost of firms' increased monopoly power is mitigated by the benefit of increased consumer choice. Differentiating functionally identical products through switching costs, however, yields no benefits to set against the cost of restricted output.

Second, the monopoly power that firms gain over their respective market segments leads to vigorous competition for market share before consumers have attached themselves to suppliers. We thus provide an explanation for the importance that many companies attach to building market share and for the emphasis that is commonly placed on market share as a measure of corporate success. However, switching costs do not necessarily make firms better off. The ferocious competition to attract new customers so as to be able to fleece them (and other consumers whom they teach or influence) after they have "bought in" to a particular brand may more than dissipate firms' extra monopolistic returns and leave them worse off than in a standard oligopoly.

Von Weizsäcker [1984] has built a simple model that incorporates switching costs, but he assumes that firms commit to charging the same price in every period (including the introductory period), and so abstracts from the issues we shall consider. In his model, in contrast to ours, the central result is that switching costs make markets more competitive.[4] We, however, argue that both the assumption that firms can credibly commit to charging the same

4. Von Weizsäcker considers products that are differentiated functionally as well as differentiated by switching costs, and also assumes that consumers' tastes for the underlying product characteristics may change. Since with switching costs a consumer's choice today is also influenced by the future when his tastes may be different, today's tastes become less important relative to any price difference that is expected to last. Hence the assumption that firms charge the same price in every period leads, in this model, to markets that are more competitive with switching costs than without switching costs. Klemperer [1987b] considers a model based on von Weizsäcker's but without the assumption that firms must choose the same price in every period, and obtains results closer to those of our paper. Prices in the second period are always higher both than in the first period and than if there were no switching costs.

price in every period, and the assumption that they will necessarily want to, are unjustified in many markets.[5]

Section II uses an example to illustrate the point that switching costs can lead to a noncooperative equilibrium that is the same as the collusive solution. Section III analyzes the collusive nature of equilibrium in a mature market in which switching costs have already been built up. This is the second period of a two-period model.[6] Then Section IV considers the first period, before consumers have attached themselves to suppliers, and shows how the second-period monopoly profits lead to competition for market share in the first period. We conclude in Section V.

II. Numerical Example

A simple example introduces the main idea. Consider two airlines, A and B, flying a route in each of two periods. Industry demand is $q = 100 - p$, and each firm's constant marginal cost is 10. Note that a monopolist (or collusive oligopoly) would choose quantity, $q = \frac{1}{2}(100 - 10) = 45$, and price, $p = 55$, yielding an industry profit of 2025, but in Cournot competition (imagine each airline choosing a number of flights to fill in each period) each firm chooses a quantity of 30, the market price $p = 40$, and profits are $(30)(40 - 10) = 900$ for each firm.

Now imagine that after Cournot competition in period 1, each firm announces that it will offer a "frequent-flyer" discount of 10 to anyone who flew on its flight during period 1. Are these discounts a sign of ferocious second-period competition?

5. Farrell [1985], Summers [1985], and Wernerfelt [1985] have independently developed models of switching costs that contain some of the results of our paper. For discussion of other models, including related models in which a consumer is reluctant to buy an untested brand because of uncertainty about product quality (see note 1), see Klemperer [1986].

In labor markets there may be switching costs of all three kinds. Job-specific training leads to "learning" costs, physical relocation costs, and other hiring-and-firing costs are "transaction" costs, and rules that employees' pensions are vested only after a certain number of years are "artificial" costs of switching. Our model can be interpreted as representing an oligopsonistic labor market. (For consumers purchasing products at high (low) prices, read workers selling labor for low (high) wages.) However, we are assuming away the possibility of mechanisms (long-term contracts, reputation, etc.) that would allow agents to make credible commitments about the future. We assume that consumers (workers) bear the switching costs and that firms cannot price discriminate, which assumptions imply that a firm cannot pay a new worker's training or moving costs without giving current employees a bonus of equal financial value. We also ignore uncertainty and risk aversion. Thus, the model is more appropriate to product markets.

6. This period is formally quite similar to a single-period model of functional product differentiation in which heterogeneity in consumers' tastes (as in standard models à la Hotelling [1929]) or firm-specific elements in consumers' utility functions (see, for example, de Palma et al. [1985]) gives firms some local monopoly power over consumers.

On the contrary: in the second-period noncooperative quantity-setting equilibrium, each firm chooses a quantity of 22½, so the total quantity $q = 45$, the market price $p = 65$, and profits are $(22\frac{1}{2})(65 - 10 - 10) = 1012\frac{1}{2}$ for each firm. Not only do the "discounts" allow the airlines to achieve the collusive level of output, but the customers do not benefit in any real way from them—the quoted prices simply end up higher than the collusive price by the amount of the discounts. Given the discounts, each firm finds that it is too expensive to increase output far enough to take any consumers that it did not serve in the previous period.[7]

If firms engage in price competition rather than quantity competition, the same result can be obtained, but a discount of at least 45 is then required. With a discount of $s \geq 45$ each firm sets a price of $55 + s$ in the noncooperative equilibrium, hence the effective net-of-discount price to "old" (previous-period) customers is 55, and so each firm sells a quantity of 22½. This equilibrium is straightforward to check. Neither firm can benefit by deviating from the monopoly price $(55 + s)$ to its old customers, unless it sells to any of its rival's customers. However, attracting the rival's customers requires pricing at least s below the rival. For $s \geq 45$ this cuts its effective price to old customers to its marginal cost or lower, and so gives up at least as much profit on these customers as it can gain by stealing its rival's customers. So each firm serves only its old customers, and acts as a monopolist against them.

Of course this is only a partial analysis. The discounts have implications for the first-period equilibrium also.[8] However, the

7. For further intuition, see the Interpretation subsection of Section III. To check the period-2 equilibrium, let $q_2^B = 22\frac{1}{2}$. Then if $q_2^A \in [17\frac{1}{2}, 27\frac{1}{2}]$, we have $p_2^B = 65$ and $p_2^A = (100 - 2q_2^A) + 10$ (since the discount has value 10). However, the maximum price difference between the firms is 10 (otherwise the higher price firm will sell nothing), so if $q_2^A \geq 27\frac{1}{2}$, market clearing requires $p_2^B = p_2^A + 10$, and if $q_2^A \leq 17\frac{1}{2}$, market clearing requires that $p_2^B = p_2^A - 10$. It is then easy to solve for the (unique) price p_2^A such that the total purchased is $q_2^A + q_2^B$ (noting that consumers with reservation prices below 40 can get discounts from neither firm), and to check that $q_2^A = 22\frac{1}{2}$ maximizes A's profits. The optimality of q_2^B given q_2^A follows by symmetry. See Klemperer [1984] for detailed computations. Any discount in excess of 9.3 is sufficient for the period-2 equilibrium to be $q_2^A = q_2^B = 22\frac{1}{2}$.

We do not allow a firm the strategy of honoring its rival's discount coupons (that is, offering discounts to consumers who flew with the rival in the first period). A tacit agreement not to do this may be easy to monitor and to enforce (a best response to a rival honoring your discounts is to honor its), see note 19. Furthermore, if each firm must first announce whether or not it will honor its rival's discount coupons before each firm chooses its quantity or price, precommitment to not honoring the rival's coupons is each firm's optimal (in fact dominant) strategy and so arises naturally.

8. When switching costs are "artificial," discount coupons, it may be possible to avoid any effects on the first-period equilibrium by distributing them in other ways than through first-period sales. For example, TWA has offered discounts to AT&T long distance phone service users, and many firms mail discount coupons to potential purchasers of their products.

basic point—that repeat-purchase discounts or, more generally, consumers' costs of switching between competing brands, lead to the collusive solution even with noncooperative behavior—is the focus of our paper.[9]

III. The Second Period: Noncooperative Behavior Looks Collusive

The next two sections look at the development over two periods of a market with switching costs. In the first period consumers have no ties to any particular firm. (Any "start-up" transaction or learning costs are assumed to be the same whichever firm a consumer buys from.) However, second-period switching costs are created by first-period sales. This section analyzes the second period, that is, the "mature market" after consumers' switching costs have been built up. It finds the symmetric equilibrium in a general model of consumers with different switching costs, and emphasizes that *noncooperative* behavior in the presence of switching costs leads to outcomes that *look* collusive. In Section IV we shall return to analyze the initial period, taking account of how first-period incentives are affected by the dependence of second-period profits on first-period sales.

In period 2, switching costs and firms' previous-period market shares are given. We consider two firms, A and B, producing functionally identical products. We assume that q consumers have reservation prices greater than or equal to $f(q)$ for the product that they previously bought or were otherwise exposed to (e.g., by observation of others' purchases or through specific training, as for computers). Thus, $f(q)$ would be inverse demand if there were no switching costs. However, consumers face a "switching cost" of purchasing the product that they were less exposed to. A fraction of the market σ^A, "A's consumers," must pay a switching cost to buy B's product, while the complementary fraction $\sigma^B(=1 - \sigma^A)$ must pay a switching cost to purchase A's. These fractions are most naturally the firms' respective shares of the previous period's sales, and this will be assumed when we examine the full two-period

9. Our numerical example is intended as an illustration of our more general analysis, and *not* as a complete explanation for "frequent-flyer" programs. A large part of their explanation is probably that they exploit a principal-agent problem within the purchasing firms. Companies pay higher ticket prices, but employees get the free vacations; and it is typically not worth companies' paying the monitoring costs to avoid this. There is also the advantage that the employees are receiving a tax-free benefit.

model in Section IV, but they could also depend on other factors. (We assume that each firm has the same proportion of consumers with each reservation price, which would arise naturally, at least in the cases we consider.)

The switching costs need not be the same for each consumer. We let $\Gamma(w)$ be the proportion of a firm's consumers whose cost of switching to the other firm's product is less than or equal to w. Thus, $\gamma(w) = \partial\Gamma(w)/\partial w \geq 0$ is the density function of the switching costs. Most naturally, $\Gamma(0) = 0$. (We discuss below the case in which $\Gamma(0) > 0$, that is, there is an atom of consumers without switching costs. Klemperer [1984] generalizes the model to allow for different distributions of switching costs for different firms and at different reservation prices.) We let $h(\cdot) = f^{-1}(\cdot)$.

Without loss of generality, let $p^A \leq p^B$. Market equilibrium requires that

$$(1a) \quad q^A = \sigma^A h(p^A) + \sigma^B \Gamma(p^B - p^A) h(p^B) + \sigma^B \int_{r=p^A}^{p^B} \Gamma(r - p^A)[-dh(r)]$$

and

$$(1b) \qquad q^B = \sigma^B(1 - \Gamma(p^B - p^A))h(p^B).$$

Firm B sells only to its own consumers with reservation prices greater than or equal to p^B and switching costs greater than or equal to $p^B - p^A$. Firm A, on the other hand, sells to all its own consumers with reservation prices greater than or equal to p^A (the first term of (1a)), to those of B's consumers with reservation prices greater than or equal to p^B and switching costs less than or equal to $p^B - p^A$ (the second term), and also to B's consumers with reservation prices in the range (p^A, p^B) and reservation price less switching cost greater than or equal to p^A (the third term).[10]

10. This model differs slightly from that of Section II. Here we are assuming that all a firm's customers pay it the same price, as is the case when switching costs are real learning or transaction costs. In the example of Section II, customers stolen from a competitor pay a higher price than the net-of-discount price paid by the firm's old customers, as is the general case when switching costs are artificial. This distinction is of no importance to the qualitative results, but for a given size of switching costs a market with artificial switching costs will be somewhat the more competitive since new customers are relatively more valuable and so the incentive to cut price or increase quantity is greater; see notes 14 and 18. Another distinction is that with artificial, but not real, switching costs, the quoted price is higher than the effective (net-of-discount) price by an amount equal to the switching cost. With artificial switching costs of s the fully collusive outcome involves prices s above the monopoly price and all purchasing consumers receiving discount s. With real switching costs of s the fully collusive outcome involves prices equal to the monopoly price, and no consumers paying the switching cost to switch between firms.

Noncooperative Equilibrium

We can use (1a) and (1b) to solve for either a price-setting or a quantity-setting equilibrium.[11] We begin with the price-competition equilibrium, as it is easier to derive.

Firm A's first-order condition is

$$\frac{\partial \pi^A}{\partial p^A} = q^A + \left[p^A - \frac{\partial c^A}{\partial q^A}\right]\frac{\partial q^A}{\partial p^A} = 0,$$

where π^A are A's profits and c^A is A's total cost. Substituting from (1a) yields

$$(2)\quad 0 = \sigma^A h(p^A) + \sigma^B \Gamma(p^B - p^A)h(p^B) + \sigma^B \int_{r-p^A}^{p^B} \Gamma(r - p^A)[-dh(r)] + \left[p^A - \frac{\partial c^A}{\partial q^A}\right]\left[\sigma^A h'(p^A) - \sigma^B \gamma(p^B - p^A)h(p^B) + \sigma^B \int_{r-p^A}^{p^B} - \gamma(r - p^A)[-dh(r)]\right].$$

So at a symmetric equilibrium (in pure strategies), $p^A = p^B = p$, with $\sigma^A = \sigma^B = ½$,

$$(3)\qquad \frac{1}{2}\left[h(p) + \left(p - \frac{\partial c^A}{\partial q^A}\right)(h'(p) - \gamma(0)h(p))\right] = 0.$$

If $\gamma(0) = 0$, then we can rewrite (3) as

$$(4)\qquad h(p) + \left(p - \frac{1}{2}\left(c^{A'}\left(\frac{q}{2}\right) + c^{B'}\left(\frac{q}{2}\right)\right)\right)h'(p) = 0,$$

where $q = 2q^A = h(p)$, and where we have assumed that $c^A(\cdot) = c^B(\cdot)$. *This is just the first-order condition for a monopolist (or collusive oligopoly) in a market without switching costs.* (We assume that the monopoly must operate both firms' plants equally. This is cost-minimizing if $c^{A''} = c^{B''} \geq 0$, and an oligopoly might anyway be constrained to do this by an inability to make side-payments.)

It is easy to use (2) to check that this result—that the first-order condition for a symmetric pure-strategy equilibrium is the same as that for a monopolist in a market without switching

11. The choice between the two equilibria should, of course, depend on the economics (see Singh and Vives [1984] and Klemperer and Meyer [1986]), but in the absence of any fully satisfactory theory to choose between them we consider both kinds of competition.

costs if $\gamma(0) = 0$—also holds when firms have unequal market shares ($\sigma^A \neq \sigma^B$), if they have constant, equal, marginal costs.

As $\gamma(0) \to \infty$ in the symmetric equilibrium, on the other hand, (3) implies that $(p - (\partial c^A/\partial q^A)) \to 0$. That is, the market price approaches the competitive price (firms' marginal cost) as we approach the case of no switching costs.

With $\gamma(0)$ between these extreme cases the equilibrium is between the competitive and collusive equilibria.

Thus, in a symmetric pure-strategy equilibrium the only information about the distribution of switching costs that matters is the density of consumers with zero switching costs, $\gamma(0)$—these are the marginal consumers who are sensitive to a small deviation in one firm's price from its competitor's price. However, the rest of the distribution is crucial in determining whether the prices satisfying the first-order conditions are global best responses for the firms.[12]

Consider the important special case in which all consumers have a switching cost of at least $s > 0$ and firms have constant, equal, marginal costs, but not necessarily equal market shares. Here $\gamma(0) = 0$, so the only possible symmetric pure-strategy equilibrium is for each firm to choose the monopoly price p_m for an otherwise identical market without switching costs.[13] At any lower common price, each firm has an incentive to slightly increase its price, which more fully exploits its own customers without losing any to its competitor.) We can check that for linear demand and costs, $f(q) = \alpha - \beta q$, $c^A(q) = c^B(q) = Cq$, the first-order conditions do in fact define an equilibrium—neither firm has an incentive to make a large deviation—if $s \geq ((\alpha - C)/4)(\sqrt{R^2 + 4R} - R)$, where $R = \max_{F=A,B}((1 - \sigma^F)/\sigma^F)$ is a measure of the relative market shares of the firms. Thus, for all σ^A and σ^B, $s \geq ((\alpha - C)/2)$ is sufficient, and for $\sigma^A = \sigma^B = ½$, $s \geq ((\sqrt{5} - 1)/4)(\alpha - C)$ is sufficient for the joint-profit-maximizing outcome to be a noncooperative equilibrium.[14] (We shall see that with quantity competition it is much easier for the joint-profit-maximizing outcome to be a noncooperative equilibrium.) For s large enough that $p^A = p^B = p_m$ in equilibrium, firm A's sales are $\sigma^A q_m$, that is, its market share of the monopoly output, and its profits are $\pi^A = \sigma^A \pi_m$, where q_m is the monopoly

12. Other consumers' switching costs are also more important in asymmetric equilibria.

13. This assumes a monopolist's profit function would be quasi concave if there were no switching costs. More generally, a price that is locally optimal for a monopolist could be an equilibrium.

14. When switching costs are artificial, $s \geq (\alpha - C)/2$ is sufficient for the joint-profit-maximizing outcome to be a noncooperative equilibrium if $\sigma^A = \sigma^B = ½$; see note 10.

output and π_m the monopoly profits in the otherwise identical market without switching costs. With smaller switching costs mixed-strategy equilibria arise but these are difficult to calculate.[15]

At the other extreme, if there is an atom of consumers without switching costs (for example, a proportion of consumers is new in the second period and uncommitted to either firm), the only possible symmetric pure-strategy equilibrium has price equal to marginal cost. (At any greater common price, each firm has an incentive to slightly lower its price and capture the entire atom of consumers without switching costs.) Clearly this is not an equilibrium in general—it is not, for example, an equilibrium with constant marginal costs if any consumers have positive switching costs—so mixed-strategy equilibria arise in this case.[16]

To obtain pure-strategy equilibria with prices between the competitive and monopoly prices, therefore, we consider the intermediate case in which there is a positive density of consumers with zero switching costs but no atom at zero switching costs. Consider linear demand, equal linear costs, equal market shares, and switching costs uniformly distributed on the interval $[0, k]$. (That is, $f(q) = \alpha - \beta q$, $c^A(q) = c^B(q) = Cq$, $\sigma^A = \sigma^B = 1/2$, $\gamma(w) = 1/k$ for $w \leq k$ and $\gamma(w) = 0$ for $w > k$.) Then the second-order conditions are globally satisfied and a unique equilibrium in pure strategies exists and is described by (3) for all $k \in [0, \infty)$.[17] The equilibrium is

15. Shilony [1977] in another context solves for the mixed-strategy equilibria of a price-competition model that is mathematically equivalent to the special case of ours in which all consumers have the same reservation price as well as the same switching cost and all firms have the same market share and constant identical marginal costs. His solution confirms the intuition that, for switching costs less than those supporting the joint-profit-maximizing outcome, the expected market price and firms' expected profits increase continuously and monotonically as the switching costs increase, from the competitive equilibrium when switching costs are zero up to the collusive outcome.

16. The technical problem here is that there is a discontinuity in firms' profit functions. Klemperer [1987b] avoids this problem by considering a model in which products are functionally differentiated as well as differentiated by switching costs. In that model pure-strategy equilibria in prices are obtained even when a fraction of the second-period consumers are "new" (uncommitted) consumers, and we confirm the intuition that the second-period price is decreasing in the proportion of new consumers.

To confirm that mixed-strategy equilibria do in general exist, use Theorem 5 of Dasgupta and Maskin [1986].

17. No asymmetric pure-strategy equilibria exist with $\sigma^A = \sigma^B = 1/2$ and equal linear costs. To see that this is true for any demand (and also that if $\sigma^A \neq \sigma^B$ the higher-market-share firm chooses the higher price in any asymmetric equilibrium), assume that $p^A < p^B$ in equilibrium and write $\Gamma(p^B - p^A) = \Delta$. So

$$q^A = \sigma^A h(p^A) + \sigma^B h(p^B)\Delta + \sigma^B (h(p^A) - h(p^B))\, x\Delta$$

and

$$q^B = \sigma^B h(p^B) - \sigma^B h(p^B)\Delta,$$

$$(5) \qquad p^A = p^B = \left\{k + \left(\frac{\alpha + C}{2}\right) - \sqrt{k^2 + \left(\frac{\alpha - C}{2}\right)^2}\right\}.$$

As k increases, the market price and industry profits increase monotonically from the competitive equilibrium, $p = C$ at $k = 0$, to the fully collusive outcome, $p = (\alpha + C)/2$, as $k \to \infty$.

The analysis of the noncooperative quantity-setting equilibrium is closely analogous, and is presented in the Appendix. As the density of consumers with zero switching costs increases from 0 to ∞, the equilibrium moves from the joint-profit-maximizing equilibrium to the Cournot equilibrium (that is, to the quantity-setting equilibrium in a market without switching costs). With quantity competition, linear demands and costs, and equal market shares, the joint-profit-maximizing outcome is a noncooperative equilibrium if all consumers have a switching cost of at least $(3/2 - \sqrt{2})(\alpha - C)$ (that is, 17.2 percent of firms' monopoly profit markup, $p_m - C$, per unit), which condition seems quite plausible.[18]

Interpretation

The intuition for these results is that firms' demands are less elastic than if there were no switching costs—if either firm increases

---

where

$$x = \left[\int_{r=p^A}^{p^B} \Gamma(r - p^A)[-dh(r)]/(h(p^A) - h(p^B))\Delta\right] \in (0, 1).$$

For A not to prefer charging p^B we require that

$$(p^A - C)q^A \geq (p^B - C)\sigma^A h(p^B) \Rightarrow (p^A - C)h(p^A)$$
$$+ (\sigma^B/\sigma^A)((p^A - C)\Delta(xh(p^A) + (1 - x)h(p^B))) \geq (p^B - C)h(p^B).$$

For B not to prefer charging p^A we require that

$$(p^B - C)q^B \geq (p^A - C)\sigma^B h(p^A) \Rightarrow (p^B - C)h(p^B)$$
$$\geq (p^A - C)(h(p^A)) + (p^B - C)\Delta h(p^B).$$

Together these inequalities require that

$$(\sigma^B/\sigma^A)((p^A - C)(xh(p^A) + (1 - x)h(p^B)))$$
$$\geq (p^B - C)h(p^B) \Rightarrow (\sigma^B/\sigma^A)(p^A - C)h(p^A) > (p^B - C)h(p^B).$$

Since we found earlier that $(p^A - C)h(p^A) < (p^B - C)h(p^B)$, it must be that $\sigma^B < \sigma^A$. The analysis is robust to the addition of an atom of consumers without switching costs, but there may exist asymmetric pure-strategy equilibria, even when firms have equal previous-period sales, if lower-reservation-price consumers have lower switching costs; see Klemperer [1986]. Scotchmer [1986] shows that there are in general no symmetric pure-strategy equilibria in this model if there are more than two firms.

18. When switching costs are artificial, it is sufficient that all consumers have a switching cost of at least $((\sqrt{2} - 1)/4)(\alpha - C)$, that is, 20.7 percent of the monopoly profit markup per unit, for the joint-profit-maximizing outcome to be a noncooperative equilibrium; see note 10.

output, its price falls faster than its opponent's, and therefore faster than without switching costs (in which case both prices would fall together). The higher are switching costs, the fewer consumers are attracted by a price cut, hence the smaller the incentive to cut price (or increase output), and the closer the equilibrium price is to the collusive price.

When the density of consumers without switching costs is zero, that is $\gamma(0) = 0$, then locally each firm's demand is as if it were a monopolist in its share (fraction σ) of the total market demand. Each firm therefore acts as a monopolist in its share of the market (provided that the first-order conditions define an equilibrium). This is precisely what happened in the example of Section II. The discounts divided the customers into two identical groups: one that had bought from A in period 1, and another that had bought from B in period 1. Provided that the firms' strategies were not too different, each firm locally faced a demand $q = \frac{1}{2}(100 - p')$, where p' was the price net of the discount, so $p = p' + 10$ was the quoted price. Each firm therefore chose an output exactly one-half of the original monopoly output.

Collusive Behavior

So far, we have shown that switching costs mean that competitive behavior may *look* collusive. However, they may also facilitate collusive *behavior*, even when they are not large enough for the collusive outcome to be a noncooperative equilibrium.

Absent switching costs, and with imperfect monitoring, the difficulty with colluding on output levels is that each firm has incentive to increase its quantity slightly (see, for example, Stigler [1964] and Green and Porter [1984]). Switching costs reduce or remove this incentive, Consider for simplicity the special case in which all consumers have a switching cost of at least $s > 0$. Recall that the collusive output is *always* a local optimum for each firm, even if the switching cost is not large enough for it to be a global optimum. So with switching costs a firm has no incentive to increase its output only slightly. A firm must drive down the price at which it sells to its own consumers by s before it takes any consumers away from its competitors, so that only large changes in output can help a firm. Such large changes are far easier to monitor and so (tacitly or otherwise) agree to eschew.

In the general model if there are some consumers with switching costs arbitrarily close to zero, then firms still have some incentive to chisel on the collusive agreement by slightly increasing their outputs. However, this incentive is reduced relative to the case of no switching costs, and the forces for collusion are correspondingly strengthened.[19]

IV. The First Period: Competition for Market Share

The previous section described the second period of a two-period market in which second-period switching costs are created by first-period sales. We now consider the first period, in which consumers are not attached to any particular firm. With switching costs in the second period, firms will compete more aggressively in the first period, because increased sales increase market share and so increase second-period profits.

In period 1, firm A chooses its first-period strategic variable v_1^A to maximize its total discounted future profits

$$\pi^A \equiv \pi_1^A(v_1^A, v_1^B) + \lambda\pi_2^A(\sigma^A(v_1^A, v_1^B)) \tag{6}$$

taking B's first-period strategic variable v_1^B as given. Here π_1^A are the firm's first-period profits, and π_2^A are the firm's second-period profits which can be written as a function of the firm's first-period market share σ^A, and are discounted by a factor λ in first-period terms. It is convenient to assume that a higher value of v_1^A represents more aggressive play (so for quantity competition we write $v_1^A = q_1^A$ and for price competition we write $v_1^A = 1/p_1^A$) so that $\partial\sigma^A/\partial v_1^A > 0$. In noncooperative equilibrium,

$$0 = \frac{\partial\pi^A}{\partial v_1^A} = \frac{\partial\pi_1^A}{\partial v_1^A} + \lambda\frac{\partial\pi_2^A}{\partial\sigma^A}\cdot\frac{\partial\sigma^A}{\partial v_1^A}.$$

Therefore, $\partial\pi_1^A/\partial v_1^A < 0$, provided that $\partial\pi_2^A/\partial\sigma^A > 0$; that is, that a higher market share makes the firm better off in the second period. Thus, firm A, and firm B also, chooses its first-period strategic variable at a level higher than that which maximizes first-period profits given the opponent's behavior: with switching costs firms compete more aggressively in the first period than they otherwise

19. Switching costs may also facilitate collusion by breaking up a market into well-defined submarkets of groups of customers who bought from different firms, and so providing natural "focal-points" for tacit collusive division of the market.

In our analysis of the noncooperative equilibrium, we assumed that firms are unable to price discriminate between their old customers and consumers that prefer their competitor's product. Each firm typically has an incentive to do this, but by collectively acting in this way, firms would reduce their profits. Similarly, in Section II, we did not allow firms the formally similar strategy of honoring their competitors' discount coupons; see note 7. Thus, even in our main model and numerical example, the role of switching costs can be thought of as making collusion easier. With switching costs it is sufficient to agree (tacitly or otherwise) not to price discriminate in favor of competitors' consumers (or honor competitors' discount coupons) which agreement may be easy to monitor and enforce, whereas without switching costs it is necessary to collude on output levels that may be very hard to observe.

would, in order to gain market share that will be valuable to them in the second. Thus, for example, U. K. banks give college students money, book-tokens, and free banking services to induce them to open accounts (the "first period"), and subsequently impose high bank charges to "milk" their customers after they graduate (the "second period").[20] Firms compete most aggressively for consumers who influence many others; for example, educational institutions. This motive explains the fierce competition for the university and high school markets that we observe in the computer industry.

On average, of course, firms end up with no more market share as a result of this fiercer competition. As the oligopoly would anyway produce more than the joint-profit-maximizing output in the first period, the result is that the firms dissipate some of their second-period gains. Specifically, each firm balances the first-period profit it gives up by more aggressive play with its second-period gains *at the margin* ($\partial\pi_1^A/\partial v_1^A = -\lambda\ (\partial\pi_2^A/\partial v_1^A)$). The *total* first-period profit given up can be in any relationship to the *total* second-period gains from switching costs, so switching costs can either help or hurt firms overall.

Two caveats should be noted to the result that firms will compete more aggressively in the first period than if there were no future switching costs. First, it is possible for a higher market share to hurt a firm ($\partial\pi_2^A/\partial\sigma^A < 0$) by making its competitor more aggressive. Thus, firms might compete less aggressively in the first period than without switching costs, to avoid gaining market share. (See Farrell [1985], Summers [1985], and Klemperer [1987a].) Second, if higher-market-share firms charge higher prices in the second period, rational consumers recognize that a lower price in the first period increases a firm's market share and so foretells a higher price in the second period. In this case consumers are less tempted by a price cut, so first-period demand is less elastic than in an otherwise identical market without switching costs in the future period. Firms still compete more aggressively in the first period than they would if they were maximizing short-run (first-period) profits *given* first-period demand, but they may compete less aggressively in the first period than in the otherwise identical market without second-period switching costs. (See Klemperer [1987b].) If either caveat applies, switching costs are unambiguously beneficial to firms.

20. An example from the U. S. banking industry is provided by the introduction of "NOW" money market checking accounts in December 1982, after industry deregulation, with a promotional frenzy of high interest rates (more than 10 percent above the rates of money market funds) and cash bonuses for opening accounts. Two years later the average rate paid was one-half percent below that of money market funds (see the *Wall Street Journal*, November 21, 1984).

Consider the case analyzed in Section III in which firms have constant, equal, marginal costs C per unit, and consumers have switching costs that are large enough that $\pi_2^A = \sigma^A \pi_m$, where π_m is the monopoly profit in an otherwise identical market without switching costs.[21] Then neither of the caveats above applies, since a firm's second-period profits are always increasing in its market share and its second-period price is unaffected by its market share. With quantity competition, $\pi_2^A(q_1^A, q_1^B) = (q_1^A/(q_1^A + q_1^B))\pi_m$, where q_1^A and q_1^B are firms' first-period quantities, so if $f_1(\cdot)$ is first-period inverse demand (taking account of any first-period learning or transaction start-up costs),

$$\pi^A = q_1^A(f_1(q_1^A + q_1^B) - C) + \lambda\left(\frac{q_1^A}{q_1^A + q_1^B}\right)\pi_m. \tag{7}$$

Each firm therefore acts in period 1 as if an additional segment of demand of constant elasticity -1 had been added to the market. It follows that if demand has this form in both periods, firms make the same total profits over the two periods with infinite switching costs as with no switching costs. It is easy to alter the demand curve slightly either so that firms prefer large switching costs or so that firms prefer no switching costs. With linear demand, firms always make less profits with large switching costs than with no switching costs. If demand is sufficiently concave below the duopoly equilibrium point when there are no switching costs, on the other hand, firms prefer switching costs.[22]

With price competition, firms with constant, equal marginal costs are exactly as well off with large switching costs in the second

21. A sufficient condition for the first-order conditions to define equilibrium for the two-period problem, for either price or quantity competition, is that all consumers have switching costs of at least $f(q_m) - C$, where q_m is the output either firm would produce as a monopolist in the second period if there were no switching costs. We assume that consumers' reservation prices are identically ordered in the two periods, so that the consumers who have the q_m highest reservation prices in the second period all purchased in the first period.

22. In these examples we are thinking of adding second-period switching costs that do not affect customers' first-period consumption utilities. This is natural if switching costs are artificial costs, learning costs (the compatibility of firms' products may affect second-period switching costs but not first-period start-up costs), or purely psychological costs. If switching costs are transaction costs, then higher second-period switching costs will generally imply higher first-period start-up costs and so lower first-period demand. It then remains true that firms may either prefer large switching costs or prefer no switching costs, but the details of the comparison are slightly different. Note that we have ignored the fact that, even in our example in which second-period prices are independent of first-period market shares, the second-period switching costs affect firms' common first-period demand $f_1(q)$ by affecting consumers' decisions about whether or not to buy in the first period (by buying in the first period, consumers are also buying the right to buy more cheaply in the second period). This has no effect on the equilibrium, since marginal first-period consumers get no utility from buying in the second period so first-period demand is unaltered around its equilibrium price.

period as they are without switching costs in the second period. However, with switching costs, the competition for market share is sufficiently fierce that first-period prices are below firms' costs, while second-period prices are at the monopoly level.

In all these examples, firms sell to more consumers in the first period than in the second period. Their first-period prices are low, not because they want to attract the lower-reservation-price consumers but because they want to win the largest possible share of the higher-reservation-price consumers. Thus, banks give gifts and cash to undergraduates opening their first bank-accounts, and book clubs give away books as introductory offers, in order to attract high-reservation-price consumers who will become valuable repeat customers, even though they also attract low-reservation-price consumers who will never purchase again.

The example with price competition most clearly illustrates a welfare cost of switching costs. In it, industry output is socially optimal in both periods if there are no switching costs, but firms produce excessive output in the first period and too little in the second if there are switching costs. Thus, switching costs cause an allocative inefficiency even though no consumers actually pay the switching cost, and even though no consumers would want to switch between the products if switching costs were zero. The result is similar with quantity competition. A more competitive first period typically increases welfare, since output is less than socially optimal in Cournot competition without switching costs, but this welfare gain is generally outweighed by the second-period welfare loss. The extra units sold in the first period are less socially valuable (their social value is closer to their marginal cost) than the units of output that are lost in the second period when output is contracted to the monopoly level.[23] Although it is possible to find demand curves for which switching costs increase total welfare,[24] simple examples, including linear demand and costs and demand of constant elasticity -1, suggest that switching costs generally reduce welfare.[25]

23. This intuition is exact for the case of demand of constant elasticity -1, for which the number of extra units sold in the first period equals the number of units fewer than the no-switching-costs output sold in the second, but in general, of course, these numbers will not be exactly equal.

24. Consider a demand curve that is sufficiently concave above the no-switching-costs duopoly equilibrium point, but that is linear or convex below it.

25. In a model in which products are functionally differentiated as well as differentiated by switching costs (see Klemperer [1987b]), switching costs may also cause a welfare loss akin to that in a standard model in which there is a cost increase. If any consumers' preferences for the underlying product characteristics change over time, then some of these consumers pay the switching cost to switch between firms in the second period, while others buy the good that is not the one they would choose if there were no switching costs. Also, in a model in which switching costs are

V. Conclusion

This paper shows that switching costs in a mature market lead to monopoly rents, but that these rents induce greater competition in the early stages of the market's development. Thus, switching costs do not necessarily make firms better off. The presumption is that welfare is reduced, so the model provides some support for regulatory policies that lower switching costs: for example, policies that encourage standardization.

This work has relied on a two-period model, and it would be useful to determine the extent to which the results carry over into a multiperiod setting. Future research should also look at markets in which in each period a proportion of consumers leaves the market and is replaced by new consumers.[26] In such markets the distinction between the first and subsequent periods would be of less consequence than in our model, and it would be interesting to establish the properties of steady-state equilibria. Finally, this paper has taken switching costs as an exogenous feature of markets. Future research should directly examine firms' incentives to increase or reduce switching costs through product-design choices, standards, contracts, etc., and should also examine the role that switching costs may play in determining industry structure.[27]

Appendix: Noncooperative Quantity-Setting Equilibrium

This Appendix analyzes the noncooperative quantity-setting equilibrium in the second-period of a market with switching costs. Using (1b) to find $\partial p^B/\partial p^A|_{q^B\text{-const}}$ and then (1a) for $\partial q^A/\partial p^A|_{q^B\text{-const}}$, the first-order condition can be simplified to show that, at any symmetric equilibrium with $\sigma^A = \sigma^B = \frac{1}{2}$,

$$\text{(A1)}\quad f(2q^A) - \frac{\partial c^A}{\partial q^A} + q^A f'(2q^A)\left[\frac{2 - 4q^A f'(2q^A)\gamma(0)}{1 - 4q^A f'(2q^A)\gamma(0)}\right] = 0.$$

If $\gamma(0) = 0$, this simplifies to the first-order condition for the

endogenous, firms may choose socially inefficient technologies in order to build up switching costs because of the implications for market equilibrium discussed in this paper.

26. Klemperer [1987b] examines a two-period model with new consumers in the second period.

27. Farrell and Saloner [1985, 1986] and Katz and Shapiro [1985, 1986a, 1986b] discuss incentives to standardize, emphasizing network-externality effects. Aghion and Bolton [1985] show how switching costs created by contracts can be used to deter new entry, and Klemperer [1987a] analyzes the effect of real switching costs on entry.

monopoly or collusive oligopoly equilibrium

$$f(2q^A) - \frac{\partial c^A}{\partial q^A} + 2q^A f'(2q^A) = 0,$$

or,

$$f(q) - ½(c^{A'}(q/2) + c^{B'}(q/2)) + qf'(q) = 0,$$

(this is (4) multiplied by $f'(q)$; as before $q = 2q^A$). If firms have constant, equal marginal costs, this result also holds for unequal market shares, $\sigma^A \neq \sigma^B$.

As $\gamma(0) \longrightarrow \infty$, the first-order condition approaches that for the symmetric Cournot equilibrium without switching costs:

$$f(2q^A) - \frac{\partial c^A}{\partial q^A} + q^A f'(2q^A) = 0.$$

As with price competition, we need to check whether the quantities given by the first-order conditions actually form an equilibrium. For linear demands and costs, $f(q) = \alpha - \beta q$, $c^A(q) = c^B(q) = Cq$, we can verify the following results.

If all consumers have a switching cost of at least $s > 0$, the noncooperative equilibrium is for each firm to choose $q^F = \sigma^F(\alpha - C)/2\beta$, if the first-order conditions define the equilibrium, which they do if

$$s \geq \left(\frac{\alpha - C}{2}\right)\left(\frac{1 - \sqrt{\sigma^F}}{1 + \sqrt{\sigma^F}}\right), \quad F = A, B.$$

Thus, for large enough s each firm chooses "its share" of the monopoly output, and the market price and total profits are the monopoly price and profits, respectively, in an otherwise identical market without switching costs.

If a proportion of consumers have no switching costs, and $\sigma^A = \sigma^B = ½$, the only possible symmetric pure-strategy equilibrium has each firm producing its Cournot output $q^A = q^B = (\alpha - C)/3\beta$, and this is an equilibrium if the proportion of consumers without switching costs is at least one-third.

With any distribution of switching costs with $\gamma(0) = 1/k$ that satisfies the second-order conditions, and $\sigma^A = \sigma^B = ½$, a unique symmetric pure-strategy equilibrium exists and has

$$\text{(A2)} \quad q^A = q^B = (1/6\beta)\{(\alpha - C) - k + \sqrt{k^2 + (\alpha - C)^2 + k(\alpha - C)}\},$$

hence

$$\text{(A3)} \quad p^A = p^B = ⅓\{k + 2\alpha + C - \sqrt{k^2 + (\alpha - C)^2 + k(\alpha - C)}\},$$

for all $k \in [0, \infty)$. Analogously to price competition, the market price and industry profits rise monotonically from the quantity-setting equilibrium without switching costs, $p = (\alpha + 2C)/3$, at $k = 0$, to the fully collusive outcome, $p = (\alpha + C)/2$, as $k \to \infty$. However, the price and profits are higher than with price competition for any given level of switching costs k. (Compare (A3) with (5): the intuition is the same as that for a market without switching costs, see Klemperer [1984].)

ST. CATHERINE'S COLLEGE AND INSTITUTE OF ECONOMICS AND STATISTICS, OXFORD

REFERENCES

Aghion, Philippe, and Patrick Bolton, "Entry Prevention through Contracts with Customers," Harvard University Working Paper, September 1985.

Clarke, Frank H., Masako N. Darrough, and John M. Heineke, "Optimal Pricing in the Presence of Experience Effects," *Journal of Business*, LV (October 1982), 517–30.

Dasgupta, Partha, and Eric Maskin, "The Existence of Equilibrium in Discontinuous Economic Games, I: Theory," *Review of Economic Studies*, LIII (January 1986), 1–26.

De Palma, A., V. Ginsburgh, Y. Y. Papageorgiou, and J.-F. Thisse, "The Principle of Minimum Differentiation Holds Under Sufficient Heterogeneity," *Econometrica*, LIII (July 1985), 767–81.

Farrell, Joseph von R., "Competition with Lock-in," GTE Laboratories Working Paper, November 1985a.

——, and ——, "Standardization, Compatibility, and Innovation," *Rand Journal of Economics*, XVI (Spring 1985), 70–83.

——, and Garth Saloner, "Economic Issues in Standardization," in *Telecommunications and Equity*, J. Miller, ed. (Amsterdam: North-Holland, 1986).

Green, Edward J., and Robert H. Porter, "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, LII (January 1984), 87–100.

Hotelling, Harold, "Stability in Competition," *Economic Journal*, XXXIX (March 1929), 41–57.

Katz, Michael L., and Carl Shapiro, "Network Externalities, Competition, and Compatibility," *American Economic Review*, LXXV (June 1985), 424–40.

——, and ——, "Product Compatibility Choice in a Market with Technological Progress," *Oxford Economic Papers*, XXXVIII (November 1986a), forthcoming.

——, and ——, "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy*, XCIV (August 1986b), 822–41.

Klemperer, Paul D., "Collusion via Switching Costs: How 'Frequent-Flyer' Programs, Trading Stamps, and Technology Choices Aid Collusion," Stanford Graduate School of Business Working Paper 786, May 1984.

——, "*Markets with Consumer Switching Costs*," Stanford University, Ph.D. Thesis, September 1986.

——, "Entry Deterrence in Markets with Consumer Switching Costs," *Economic Journal*, XCVII (1987a), forthcoming.

——, "The Competitiveness of Markets with Switching Costs," *Rand Journal of Economics*, XVIII (Spring 1987b), forthcoming.

——, and Margaret A. Meyer, "Price Competition vs. Quantity Competition: The Role of Uncertainty," *Rand Journal of Economics*, XVII (Winter 1986), forthcoming.

Nelson, Phillip, "Information and Consumer Behavior," *Journal of Political Economy*, LXXVIII (March/April 1970), 311–29.

Scotchmer, Suzanne, "Market Share Inertia with More Than Two Firms: An Existence Problem," Harvard University Working Paper, December 1985.

Shilony, Yuval, "Mixed Pricing in Oligopoly," *Journal of Economic Theory*, XIV (April 1977), 373–88.
Singh, Nirvikar, and Xavier Vives, "Price and Quantity Competition in a Differentiated Duopoly," *Rand Journal of Economics*, XV (Winter 1984), 546–54.
Spence, A. Michael, "The Learning Curve and Competition," *Bell Journal of Economics*, XII (Spring 1981), 49–70.
Stigler, George J., "A Theory of Oligopoly," *Journal of Political Economy*, LXXII (February 1964), 44–61.
Summers, Lawrence H., "On Frequent-Flyer Programs and other Loyalty-Inducing Economic Arrangements," *Economics Letters*, XXI (May 1986), 77–79.
von Weizsäcker, C. Christian, "The Costs of Substitution," *Econometrica*, LII (September 1984), 1085–1116.
Wernerfelt, Birger, "Nontatonnement Economies: Market Organization, Pricing, and General Equilibrium," Northwestern University Working Paper, October 1985.