The Economic Journal, **97** (*Conference* 1987), 99–117
Printed in Great Britain

ENTRY DETERRENCE IN MARKETS WITH CONSUMER SWITCHING COSTS*

Paul Klemperer

Consumers that have previously bought from one company may incur substantial costs if they switch to buying the identical product from a competitor. There are, for example, high transaction costs incurred in closing an account with one bank and opening another with a competitor, or in returning rented equipment to one company and renting identical equipment from another. Consumers may also face switching costs between brands that, although not strictly identical, are functionally identical. Two computer manufacturers may make machines that are known to be of the same quality, but having learnt to use a firm's product line one has a strong incentive to continue to buy from that firm. Similarly, when choosing a cake mix it is easiest to buy the brand that one already knows exactly how to prepare. In all these markets, rational consumers display brand loyalty when faced with a choice between functionally identical products.[1] Products that are ex ante homogeneous become, after the purchase of one of them, ex post heterogeneous.

Switching costs suggest natural explanations for several aspects of firms' commonly-observed behaviour, as well as some more surprising results. This paper examines how the threat of new entry affects an incumbent's behaviour in a market with switching costs, and thus provides a simple explanation of limit pricing behaviour:[2] A firm may cut price prior to the date in which new entry is threatened, to build up its customer base and thus deny customers to, and so reduce the profits of, any entrant. In contrast to the classical models of Bain (1949), Gaskins (1971), and others, this form of limit pricing deters *rational* potential entrants. In contrast to the models of Milgrom and Roberts (1962), Matthews and Mirman (1983), and Saloner (1984), among others, a firm in this model does not directly dissipate first-period profits to signal information. If signalling the information is important, there are probably far cheaper ways to do this. As in the latter group of models, however, rational limit pricing may

* This project was begun at Stanford University in the summer of 1983 and supported there by the Stanford Graduate School of Business and the Center for Economic Policy Research. I would like to thank colleagues at Oxford and Stanford, especially Tim Bresnahan, Jeremy Bulow, Meg Meyer, Jim Mirrlees and John Roberts, and the editors of this volume for helpful suggestions.

[1] Our model also applies to cases in which, for no obvious economically rational reason, consumers continue to buy a brand from simple habit or loyalty, that is, cases in which there is a psychological cost of switching brands. It does not, however, apply directly to cases in which 'artificial' switching costs are created by contracts, repeat-purchase coupons, etc. in which there are no real (social) costs of brand switching (see Klemperer (1984, 1986)).

[2] In Klemperer (1986) we showed how switching costs can lead to price wars as equilibrium behaviour with pure strategies. Klemperer (1984, 1986), von Weizsäcker (1984), Farrell (1985) and Summers (1985) have studied the effect of switching costs on markets protected from entry. Schmalensee (1982) emphasises the advantage that the first entrant into a market has when consumers are reluctant (in his model because of uncertainty about quality) to switch to a subsequent entrant's product. However, he artificially restricts firms' strategies and assumes away the incumbent's strategic behaviour upon which we focus.

actually be followed by new entry – in our case because having a larger number of customers 'buy in' to its product may help an incumbent even if entry does occur.

It might seem that firms facing potential entrants always have an incentive to produce more than they would otherwise, in order to lock in customers. However, this is *not* always the case. Especially when demand is growing, or when some consumers leave the market each period and are replaced by new consumers, a firm may produce less, and so price higher if it faces entry than if it does not: 'limit over-pricing' or 'under-investment' by the firm in building its customer base may occur. The reason is that a low output makes the incumbent more likely to price lower in the next period to attract new customers (the profits thereby foregone on 'bought-in' customers are less), and this helps to scare off entry. Thus very low customer bases and very high customer bases are the most effective entry deterrents.

Similarly, higher switching costs typically help an incumbent and deter or weaken new entry, but this also is not always true. In addition to locking in customers to buying from the incumbent, high switching costs may in effect lock in the incumbent to selling only to repeat customers. Higher switching costs reduce the incentive for the incumbent to fight aggressively for new customers in later periods, and so can encourage rather than discourage entry. Entry may be more deterred by either very high or very low switching costs than by switching costs in a middle range.[1]

Finally, it might seem that the first entrant into a market with switching costs always has a major advantage – consumers have no other supplier to turn to while the first entrant is a monopolist and, having bought from it are locked in by switching costs even after other firms have entered the market. In a growing market, however, even without uncertainty (uncertainty about growth rates would hurt an incumbent), the incumbent's first-mover advantage must be set against a loss of flexibility. Having built up a customer base as a monopolist, the incumbent is on the horns of a dilemma. Should it charge a high price that capitalises on its repeat purchasers but restricts it to a small share of the growing market, or should it charge a low price that attracts new customers but does not take advantage of its current customer base? (We are assuming that firms cannot easily price discriminate between new and repeat purchasers.) The incumbent may even be worse off than the new entrant. The entrant has no choice but to charge a low price to go after new customers and, knowing this, the incumbent may do better to 'skim' the market with a high price which milks its repeat purchasers than to compete head-to-head with its opponent. However this leaves it worse off than the entrant which thereby gets a monopoly of the bottom end of the market. A company with the ability to enter a market at an early stage may therefore be at a disadvantage relative to (i.e. make less total profit over time than) an otherwise identical company that can enter the market only at a later stage.

[1] These results are reminiscent of, but contrast with, Spence's (1981) result that entry is most likely with a very fast or a very slow learning curve in an industry, and that entry barriers are highest with moderate rates of learning.

Section I sets out the model in a general framework. Section II illustrates the results for the case of linear demand and costs, and thus shows that they do not depend on unusual assumptions. Section III concludes.

I. THE MODEL

We consider a market that operates for two periods $t = 1, 2$ with complete and perfect information. An incumbent firm I is present in the market in both periods, and in each period, t, has costs $C_t^I(q_t^I)$ of producing output q_t^I. A potential entrant E observes I's period-one output (or observes the period-one market price, p_1^I) and decides whether or not to incur a fixed cost F of entering the market. If it does, its costs of producing output q_2^E in period two are $C_2^E(q_2^E)$. The firms' products are functionally identical, that is, we assume they are undifferentiated except by switching costs. Demand in period t is $f_t(q)$, to be interpreted as the qth consumer having reservation price $f_t(q)$ for one unit of either firm's product in period t, net of any switching cost. Each consumer has a 'switching cost' (or start-up cost) s, which we take as given, of buying either firm's product for the first time. Products cannot be stored between periods.

If entry occurs, we assume Cournot equilibrium in period two leading to market prices p_2^I and p_2^E for the incumbent's and entrant's products respectively.[1] Thus in the second period

$$p_2^E = f_2(q_2^I + q_2^E) - s,$$

$$p_2^I = f_2(q_2^I + q_2^E) \quad \text{if} \quad q_2^I \leqslant q_1^I,$$

and

$$p_2^I = f_2(q_2^I + q_2^E) - s \quad \text{if} \quad q_2^I > q_1^I.$$

(We are assuming that I cannot price discriminate against its repeat customers. However, similar results to those we will obtain would emerge if firms could price discriminate, but not by the full amount of the switching cost s, or if there were a cost to price discrimination.) Firms and consumers discount second-period revenues and costs by a factor λ in first-period terms, and I's first-period price p_1^I is computed from its choice of output q_1^I, assuming consumers have rational expectations about future prices.[2] We will write I's first-period profits and second-period profits as $\Pi_1^I(q_1^I)$ and $\Pi_2^I(q_2^I, q_2^E; q_1^I)$, respectively. (Note that I's second-period profits depend on its first-period output as well as on both firms' second-period outputs.) We will write E's second-period profits as $\Pi_2^E(q_2^I, q_2^E)$. I's total profits in first-period terms are $\Pi^I = \Pi_1^I + \lambda \Pi_2^I$.

The analysis is simplified by considering firms' second-period reaction curves, when both firms are in the market. (A firm's reaction curve is the locus of its optimal output as a function of the other firm's output in that period.) We write

[1] That is, each firm simultaneously chooses a quantity, and the firms' prices are those that clear the market. When $q_2^I = q_1^I$ an interval of prices clears the market for I's product, so we assume p_2^I is the highest price that clears the market.

[2] Thus $p_1^I = f_1(q_1^I) - s$ if consumers rationally expect $q_2^I \leqslant q_1^I$ and so also $p_2^I = p_2^E + s$, and $p_1^I = f_1(q_1^I) - s + \lambda s$ otherwise, since if $q_2^I > q_1^I$, and so also $p_2^I = p_2^E$, then consumers who bought from I in period one are better off by s in period two than if they had not bought in period one. However, the assumption of rational expectations is not important for the results.

$R^I_0(.)$ and $R^E_0(.)$ for what would be I's and E's reaction curves if consumers had no switching costs, and $R^I_s(.)$ and $R^E_s(.)$ for I's and E's reaction curves when consumers have a switching cost s and $q^I_1 = 0$ (so that neither firm has any customer base). $R^I_s(.)$ is everywhere lower than $R^I_0(.)$, and $R^E_s(.)$ is everywhere lower than $R^E_0(.)$, since the existence of the switching costs makes each firm's price and marginal revenue lower by s – see Fig. 1. For convenience we assume that on the demand curve $f_2(q)$ each firm's marginal revenue is everywhere decreasing in the other's output (what Bulow *et al.* (1985) call the 'strategic substitutes' assumption), which implies that $R^I_0(.)$, $R^E_0(.)$, $R^I_s(.)$ and $R^E_s(.)$ are all downward sloping. We also assume that on the axes drawn $R^I_0(.)$ and $R^I_s(.)$ are steeper than $R^E_0(.)$ and $R^E_s(.)$, to ensure that equilibria are stable.

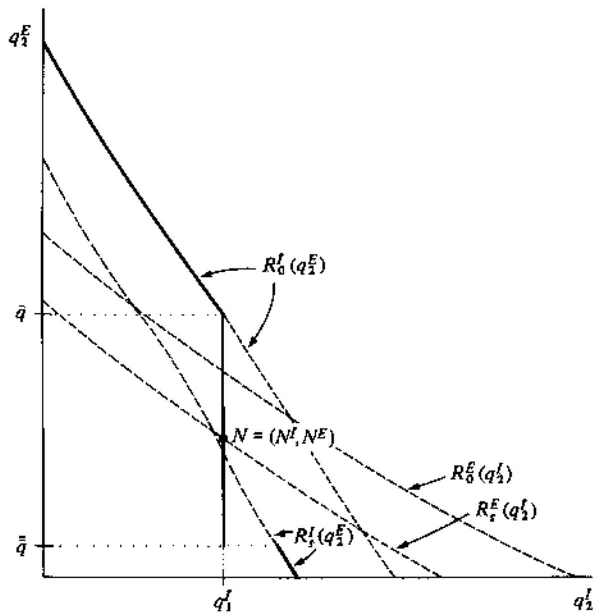

Fig. 1. Reaction curves in the second period.

The heavy line in Fig. 1 is I's reaction curve given $q^I_1 > 0$. To derive it, we first recall that for $q^I_2 \leqslant q^I_1$, I's residual demand is $f_2(q^I_2 + q^E_2)$, since its consumers have no switching costs, whereas for $q^I_2 > q^I_1$, I's residual demand is $f_2(q^I_2 + q^E_2) - s$, as if all its consumers had to pay a switching cost s. Define $\bar{q}$ by $R^I_0(\bar{q}) = q^I_1$. I's optimal response to $q^E_2 \geqslant \bar{q}$ would be $R^I_0(q^E_2) \leqslant q^I_1$ if I faced demand $f_2(q^I_2 + q^E_2)$ everywhere, so it must remain $R^I_0(q^E_2)$ here, since I faces this demand at $R^I_0(q^E_2)$ and weakly lower demand everywhere else. I's optimal response to $q^E_2 < \bar{q}$ would be $R^I_s(q^E_2)$ if I faced demand $f_2(q^I_2 + q^E_2) - s$ everywhere, so any response greater than q^I_1 is dominated by $R^I_s(q^E_2)$. Further, any response $\tilde{q} < q^I_1$ is dominated by q^I_1 since the 'strategic substitutes' assumption implies

$$\Pi^I_2(q^I_1, q^E_2; q^I_1) - \Pi^I_2(\tilde{q}, q^E_2; q^I_1) > \Pi^I_2(q^I_1, \bar{q}; q^I_1) - \Pi^I_2(\tilde{q}, \bar{q}; q^I_1)$$

which is positive by the definition of $\bar{q}$. Thus I's optimal response to $q^E_2 < \bar{q}$ is either q_1^I with profits

$$\Pi_2^I(q_1^I, q_2^E; q_1^I) = [f_2(q_1^I + q_2^E)\, q_1^I - C_2^I(q_1^I)]$$

or $R_s^I(q_2^E)$ earning profits

$$\Pi_2^I(R_s^I(q_2^E), q_2^E; q_1^I) = [f_2(R_s^I(q_2^E) + q_2^E)\, R_s^I(q_2^E) - C_2^I(R_s^I(q_2^E)) - R_s^I(q_2^E)\, s].$$

Observe that for $R_s^I(q_2^E)$ close to q_1^I the former gives higher profit. Let $\bar{\bar{q}}$ be the largest value of q_2^E satisfying $\Pi_2^I(R_s^I(q_2^E), q_2^E; q_1^I) \geqslant \Pi_2^I(q_1^I, q_2^E; q_1^I)$.

For $\tilde{\tilde{q}} < \bar{\bar{q}}$ we have

$$\begin{aligned}\Pi_2^I(R_s^I(\tilde{\tilde{q}}), \tilde{\tilde{q}}; q_1^I) - \Pi_2^I(q_1^I, \tilde{\tilde{q}}; q_1^I) &> \Pi_2^I(R_s^I(\bar{\bar{q}}), \tilde{\tilde{q}}; q_1^I) - \Pi_2^I(q_1^I, \tilde{\tilde{q}}; q_1^I) \\ &> \Pi_2^I(R_s^I(\bar{\bar{q}}), \bar{\bar{q}}; q_1^I) - \Pi_2^I(q_1^I, \bar{\bar{q}}; q_1^I) \geqslant 0,\end{aligned}$$

in which the first and second inequalities follow from the definition of $R_s^I(\,.\,)$ and the 'strategic substitutes' assumption, respectively. So for all $q_2^E \in (\bar{\bar{q}}, \bar{q}]$ I's optimal response is q_1^I, and for $q_2^E \leqslant \bar{\bar{q}}$ I's optimal response is $R_s^I(q_2^E)$. Hence I's second-period reaction curve is as drawn: For high q_2^E I sells to fewer than all its first-period customers. For a range of q_2^E I sells to all its first-period customers and to no others. However increasing output slightly above q_1^I loses I a lot of revenue (more than $q_1^I\,.\,s$) from repeat purchasers and is never profitable. Only if q_2^E is very low does I find it worthwhile to 'hold a sale' in which a large number of sales to new customers compensates for the loss of revenue from old customers.

Limit Pricing, or Over-investment in Building Customer Base

If E enters the market, E's second-period reaction curve is just $R_s^E(q_2^I)$ and the second-period Cournot–Nash equilibrium is at the intersection (N^I, N^E) of the second-period reaction curves.[1] For $N^E < \bar{q}$ or $N^E > \bar{\bar{q}}$, a small change in q_1^I would have no effect on the second-period equilibrium. If, however, E's reaction curve cuts I's reaction curve in its vertical section, that is $N^E \in (\bar{q}, \bar{q})$, as drawn, a small increase (decrease) in q_1^I increases (decreases) I's second-period output and decreases (increases) E's second-period output, that is,

$$\frac{dq_2^I}{dq_1^I} > 0 \quad \text{and} \quad \frac{dq_2^E}{dq_1^I} < 0.$$

Decreasing q_2^E raises I's second-period residual demand everywhere and so increases I's second-period profits. I therefore chooses q_1^I at a higher level than if it simply maximised its long-run profits ignoring the effect of q_1^I on q_2^E. Since q_1^I is higher, p_1^I is lower than if I behaved non-strategically, taking q_2^E as given

[1] The reaction curves may intersect twice in which case there are two pure-strategy Nash equilibria, and an additional assumption must be made to choose between them. All our analysis and results hold whether we adopt the convention that the equilibrium with higher q_2^I (and lower q_2^E) would be played or the convention that the equilibrium with lower q_2^I (and higher q_2^E) would be played. We adopt the former convention throughout since this both raises I's second-period profits (q_2^E is lower) and raises I's first-period profits, because the lower second-period price of I's product raises the amount consumers will pay for I's product in the first period.

at its equilibrium value.[1] It also follows that I chooses q_1^I higher than in an open loop equilibrium in which q_1^I, q_2^I and q_2^E are all chosen simultaneously so that I cannot influence E's behaviour.[2] In both these senses, I 'over-invests' in its future market share (or current customer base), or 'limit prices', in the first period in order to influence E's behaviour in the second period.[3]

Thus far we have assumed that E's second-period entry is inevitable. In fact, E will enter the market only if its potential second-period profits, Π_2^E, exceed its fixed cost of entry, F. Since in the range of Nash equilibria discussed above, $dq_2^I/dq_1^I > 0$ we have $d\Pi_2^E/dq_1^I < 0$ so increasing q_1^I may prevent E's entry completely, giving a discontinuous jump in I's second-period profits. Thus I might 'limit price' to altogether prevent entry that would occur if it chose its output ignoring strategic considerations. In cases of strategically blocking entry, and in some cases of strategically reducing entry, I chooses a higher output and lower price in period one (and so also in period two) than a monopolist not facing entry would choose in either period.

'Limit Over-pricing', or Under-investment in Building Customer Base

The above shows that in a market with consumer switching costs incumbents may limit price or over-invest in building up a customer base in order to gain a strategic advantage in the second-period market. However, it is possible to obtain the opposite conclusion.

Consider Fig. 2 in which $q_1^{I'}$ is the first-period output I would choose ignoring

[1] Note that even with non-strategic behaviour, I chooses a greater q_1^I than the value at which first-period marginal revenue equals first-period marginal cost, so $\partial\Pi_1^I/\partial q_1^I < 0$, because higher output in the first period increases second-period demand by increasing the number of consumers 'bought-in' to I's product. Our point is that I chooses a higher q_1^I than would be optimal if I allowed for this effect but took q_2^E as given. Formally, acting strategically, I chooses q_1^I such that

$$0 = \frac{d\Pi^I}{dq_1^I} = \frac{\partial\Pi_1^I}{\partial q_1^I} + \lambda\left(\frac{\partial\Pi_2^I}{\partial q_2^I}\frac{dq_2^I}{dq_1^I} + \frac{\partial\Pi_2^I}{\partial q_2^E}\frac{dq_2^E}{dq_1^I}\right),$$

hence

$$\frac{\partial\Pi_1^I}{\partial q_1^I} + \lambda\frac{\partial\Pi_2^I}{\partial q_2^I}\frac{dq_2^I}{dq_1^I} < 0,$$

whereas with non-strategic behaviour I would choose a lower q_1^I to set

$$\frac{\partial\Pi_1^I}{\partial q_1^I} + \lambda\frac{\partial\Pi_2^I}{\partial q_2^I}\frac{dq_2^I}{dq_1^I} = 0,$$

in which expressions $\Pi_2^I(q_2^I, q_2^E; q_1^I) = q_2^I f(q_2^I + q_2^E) - C_2^I(q_2^I)$ and $dq_2^I/dq_1^I = 1$ in the relevant range. We assume that the second-order conditions for I's two-period maximisation problem are satisfied.

[2] In an open loop equilibrium (OLE) q_1^I and q_2^I are optimal for I given q_2^E, and q_2^E is optimal for E given q_1^I and q_2^I. In our (closed loop) equilibrium (CLE) q_1^I is chosen first and q_2^I and q_2^E are then optimal for I and E respectively given q_1^I and given the other player's second-period strategic variable. I must be better off in CLE than in (pure-strategy) OLE since in CLE it has the option of choosing its OLE output in the first period and so inducing the OLE outcome. Therefore q_2^E must be lower in CLE than in OLE because otherwise I would be worse off both because of the higher q_2^E and because its outputs would no longer (in general) be optimal given q_2^E. Therefore if $dq_2^E/dq_1^I < 0$, q_1^I is higher in CLE than in OLE.

[3] Fudenberg and Tirole (1984) say that a firm over-invests if it invests more in closed loop equilibrium than in open loop equilibrium. An alternative definition follows Bulow *et al.* (1985) and says that a firm over-invests if it invests more than would be optimal if it took its opponent's subsequent behaviour as given at its CLE value. In our context the two definitions are equivalent (see note 2 on p. 106).

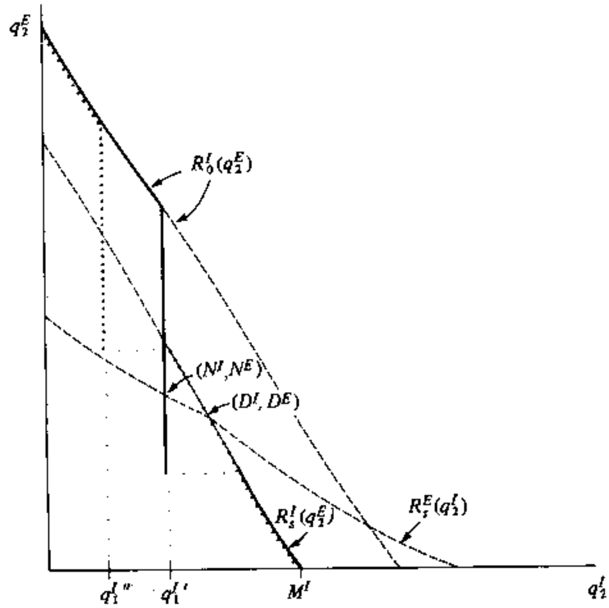

Fig. 2. Reaction curves illustrating under-investment.

strategic considerations. (That is, $q_1^{I'}$ maximises I's long-run profits ignoring its effect on q_2^E, which is taken to be optimal given $q_1^{I'}$; thus $q_1^{I'}$ and q_2^E are as in an open loop equilibrium.) The heavy line is I's second-period reaction curve given $q_1^I = q_1^{I'}$, and the second-period equilibrium is at (N^I, N^E) assuming that $\Pi_2^E(N^I, N^E) \geqslant F$ so that E does enter the market. As before, slightly raising q_1^I leads to a second-period equilibrium in which Π_2^I is higher and Π_2^E lower so limit pricing is a possibility. However, *reducing* q_1^I may be a more attractive option. Reducing q_1^I always raises the vertical position of the horizontal gap in I's second-period reaction curve, and as $q_1^I \to 0$ the gap moves towards the intersection of $R_s^I(.)$ and the vertical axis.[1] Consider the effects of reducing q_1^I to the output $q_1^{I''}$, for which I's resulting second-period reaction curve is shown by the dotted line. The second-period Nash equilibrium is now at (D^I, D^E) – the duopoly equilibrium in a market in which neither firm has any customer base.[2] I's second-period profits in this equilibrium may be either

[1] To confirm this, let $\bar{\bar{q}}$ be the value of q_2^E at which I's second-period reaction curve has its horizontal gap when $q_1^I = q_1^{I'}$, let $\bar{q}$ be such that $R_0^I(\bar{q}) = q_1^{I'}$ so $\bar{q} > \bar{\bar{q}}$, let $q_1^{I''} < q_1^{I'}$ and consider any $\tilde{q} \leqslant \bar{\bar{q}}$. 'Strategic substitutes' implies that $\Pi_2^I(q_1^{I'}, \tilde{q}; q_1^{I'}) - \Pi_2^I(q_1^{I''}, \tilde{q}; q_1^{I'}) > \Pi_2^I(q_1^{I'}, \bar{q}; q_1^{I'}) - \Pi_2^I(q_1^{I''}, \bar{q}; q_1^{I'})$ which is positive by definition of $\bar{q}$. Thus $\Pi_2^I[R_s^I(\tilde{q}), \tilde{q}; q_1^{I''}] = \Pi_2^I[R_s^I(\tilde{q}), \tilde{q}; q_1^{I'}] \geqslant \Pi_2^I(q_1^{I'}, \tilde{q}, q_1^{I'}) > \Pi_2^I(q_1^{I''}, \tilde{q}; q_1^{I'}) = \Pi_2^I(q_1^{I''}, \tilde{q}; q_1^{I''})$. Since, therefore, $\Pi_2^I(R_s^I(\tilde{q}), \tilde{q}; q_1^{I''})$ strictly exceeds $\Pi_2^I(q_1^{I''}, \tilde{q}; q_1^{I''})$ for all $\tilde{q} \leqslant \bar{\bar{q}}$, it follows that the vertical position of the horizontal gap in I's second-period reaction curve is raised by reducing q_1^I.

[2] A choice of q_1^I such that the segment of I's reaction curve that coincides with $R_s^I(.)$ just touches (D^I, D^E) is sufficient for the second-period equilibrium to be at (D^I, D^E). However, I's and E's reaction curves then cross twice so that we rely on the assumption about equilibrium selection made in note 1 on p. 103. Fig. 2 illustrates a smaller choice of q_1^I, $q_1^I = q_1^{I''}$, for which the reaction curves cross only once so that the second-period equilibrium is unambiguously at (D^I, D^E).

more or less than its profits in the equilibrium resulting from $q_1^I = q_1^{I'}$, since I's second-period residual demand is raised by q_2^E being lower ($D^E < N^E$) but lowered by the fact that its customer base is smaller.[1] In addition, however, $\Pi_2^E(D^I, D^E) < \Pi_2^E(N^I, N^E)$ since $D^I > N^I$. If $\Pi_2^E(D^I, D^E) < F$, then E will not enter, and I will be a monopolist in the second-period with equilibrium at $(M^I, 0)$. In this case I is likely to be better off than at (N^I, N^E), even taking into account any reduction in first-period profits resulting from reducing its output below $q_1^{I'}$.

Thus it is possible that I may 'under-invest' in building up its customer base or 'limit over-price' in order to deter entry.[2] This cannot occur if firms' costs and the market demands are the same in both periods and if (as we assumed) there is no turnover of consumers between periods – in this case choosing $q_1^I = M^I$ always dominates choosing $q_1^I < D^I$ since M^I is both the monopoly output on the first-period demand curve and is at least as effective in deterring entry.[3] However, if demand is growing over time or if I's costs are different in the two periods, then the second-period market is, in effect, larger than the first-period market, so I's monopoly output on its first-period demand may be less than D^I. In these cases a high first-period output (relative to first-period demand and costs) may be less cost-effective in deterring entry than a smaller first-period output that guarantees I will 'hold a sale' and produce D^I in the second period. Under-investment can then arise even with linear demand and constant marginal costs in both periods, as we show in Section II. Under-investment can also arise if a fraction of first-period consumers leave the market

[1] If $qu_1^{I'}$ is just above the value of q_1^I at which the second-period equilibrium changes to (D^I, D^E) and $q^{I''}$ is just below this value, then I's second-period profits will be higher at $q^{I''}$ because the equilibrium value of q_2^E (and hence I's residual demand) changes discontinuously here. (We cannot ascertain whether or not $q^{I''}$ is a better strategy than an over-investment strategy without knowing the details of demand and costs.)

[2] As with over-investment, there are two natural definitions of under-investment. We could say that a firm under-invests if (*a*) it invests less in closed loop equilibrium than in open loop equilibrium (as in Fudenberg and Tirole (1984)), or if (*b*) it invests less than would be optimal if it took its opponent's subsequent behaviour as given at its CLE value (which is similar to Bulow *et al*'s (1985) definition). In our model the definitions are equivalent: write $q_1^I(CLE)$, $q_2^E(CLE)$, $q_1^I(OLE)$ and $q_2^E(OLE)$ for I's choices of first-period output and E's choices of second-period output in CLE and in OLE respectively. Write $q_1^{I\ddagger}$ for the first-period output that would be optimal for I taking E's CLE output as given, and write $q_2^{E\ddagger}$ for the second-period output that E would choose if I chose $q_1^{I\ddagger}$. Observe (*i*) $q_2^E(CLE) < q_2^E(OLE)$ unless $q_1^I(CLE) = q_1^I(OLE)$ (shown in note 2 on p. 104), and (*ii*) $q_2^E(CLE) < q_2^{E\ddagger}$ unless $q_1^I(CLE) = q_1^{I\ddagger}$ (since otherwise I would prefer choosing $q_1^{I\ddagger}$ to choosing $q_1^I(CLE)$ the former is more profitable even against $q_2^E(CLE)$). Now note that there exists $\hat{q}_1^I$ such that for all $q_1^I \leqslant \hat{q}_1^I$, $q_2^E = D^E$ (because the vertical position of the horizontal gap in I's second-period reaction curve is raised by reducing q_1^I - shown in note 1 on p. 105) and for all $q_1^I > \hat{q}_1^I$, $dq_2^E/dq_1^I < 0$ (because E's reaction curve is downward sloping). So E's equilibrium choice of q_2^E is weakly quasi-concave as a function of q_1^I. It therefore follows from (*i*) and (*ii*) that $q_1^I(CLE)$ cannot lie between $q_1^I(OLE)$ and $q_1^{I\ddagger}$. Since $q_1^I(CLE) = q_1^I(OLE)$ if and only if $q_1^I(CLE) = q_1^{I\ddagger}$, it follows that the two definitions of under-investment, (*a*) and (*b*), are equivalent, and also that the two definitions of over-investment given in note 3 on p. 104 are equivalent.

[3] I's long-run profits from choosing $q_1^I = M^I$ are, if entry is not completely blocked, $\Pi^I = [M^I f(M^I) - C(M^I) - M^I s] + \lambda[\tilde{M}^I f(\tilde{M}^I + \tilde{M}^E) - C(\tilde{M}^I)]$ where $f(\cdot) = f_1(\cdot) = f_2(\cdot)$, $C(\cdot) = C_1^I(\cdot) = C_2^I(\cdot)$, $\tilde{M}^I$ is the smaller of M^I and the I^{th} co-ordinate of the intersection of $R_0^I(\cdot)$ and $R_s^E(\cdot)$ so $\tilde{M}^I > D^I$, and $\tilde{M}^E = R_s^E(\tilde{M}^I)$ so $\tilde{M}^E < D^E$. I's long-run profits from an under-investment strategy with $q_1^I < D^I$ are, in the case that entry is not completely blocked, $\Pi^I = [q_1^I f(q_1^I) - C(q_1^I) - q_1^I(1-\lambda) s] + \lambda[D^I f(D^I + D^E) - C(D^I) - D^I s] < [q_1^I f(q_1^I) - C(q_1^I) - q_1^I s] + \lambda[D^I f(D^I + D^E) - C(D^I)] < [M^I f(M^I) - C(M^I) - M^I s] + \lambda[D^I f(D^I + \tilde{M}^E) - C(D^I)] < [M^I f(M^I) - C(M^I) - M^I s] + \lambda[\tilde{M}^I f(\tilde{M}^I + \tilde{M}^E) - C(\tilde{M}^I)]$. Similar arguments take care of the case in which entry is completely blocked by M^I but not by q_1^I, and of the case in which entry is completely blocked by both M^I and q_1^I. (It is not possible for entry to be completely blocked by q_1^I but not by M^I.)

and are replaced by new consumers in period two, because in this case, too, a relatively high first-period output can still leave I with a customer base below D^I on entering the second period.[1]

We have seen that more entry is deterred by small customer bases than by customer bases in a middle range. The intuition is that a very small customer base makes it rational for an incumbent to 'hold a sale' – increase output – in the next period. With a slightly larger customer base the incumbent will do better to 'milk' its current customers for a higher price. Unfortunately for the incumbent, an entrant which takes advantage of the fact that it is rational for the incumbent to show restraint is in a stronger position than it would be if the incumbent had a smaller customer base. Any customer base smaller than a critical size deters entrants whose fixed costs of entry exceed the entrant's duopoly profits in a market in which neither firm has an established customer base. A marginally larger incumbent's customer base discontinuously changes the post-entry equilibrium and so discontinuously increases the maximum cost a firm will be willing to incur to enter the market. Above this size of customer base, however, increasing the incumbent's customer base always reduces the profitability of entry, until the customer base reaches the point corresponding to the intersection of the reaction curves $R_0^I(\,.\,)$ and $R_s^E(\,.\,)$, beyond which point further increases have no effect on the second-period equilibrium.

More generally, because E's second-period reaction curve is (we assumed) downward sloping, I reaps strategic benefits from any actions that raise its second-period reaction curve. When there are consumer switching costs, increasing I's first-period sales typically raises I's second-period reaction curve in the relevant range because the marginal profitability of selling an additional unit to a repeat customer is higher than to a new customer; hence the likelihood of over-investment in the first period. However, we showed that higher first-period sales also lower part of I's second-period reaction curve by reducing the range of q_2^E for which 'holding a sale' is rational; hence the possibility of under-investment in the first period.[2]

The Size of Switching Costs

We might expect that larger switching costs reduced the probability of entry by reducing consumers' willingness to buy from a new entrant. Consideration of Fig. 3, however, shows that this need not be the case.

First take I's period-one output as given as $q_1^I = q_1^{I*}$. When there are no

[1] Let $q_1^{I'}$ be the first-period output that would maximise I's long-run profits ignoring the effect on E's second-period output, and let r be the fraction of first-period consumers leaving the market. Then $(1-r)\,q_1^{I'}$ is the customer base with which I would enter the second period after a non-strategic choice of q_1^I. If this customer base would lead to a second-period equilibrium in which I sold only to that customer base, that is, produced $(1-r)\,q_1^{I'}$, and if $(1-r)\,q_1^{I'} < D^I$, and if a slightly smaller customer base would lead to an equilibrium in which I 'held a sale' and sold D^I, then I will typically under-invest, in equilibrium (see note 1 on p. 106).

[2] If E's reaction curve sloped upward around equilibrium, I would get strategic benefits from any actions that lowered its second-period reaction curve and under-investment results would be easy to obtain (see Fudenberg and Tirole (1984) and Bulow *et al.* (1985)). Similar under-investment results to ours have been obtained in models in which advertising rather than switching costs alters firms' demand curves by Baldani (1983) and Schmalensee (1983), who show that lower advertising can provide strategic benefits for a firm by making it more likely to behave aggressively in a future period.

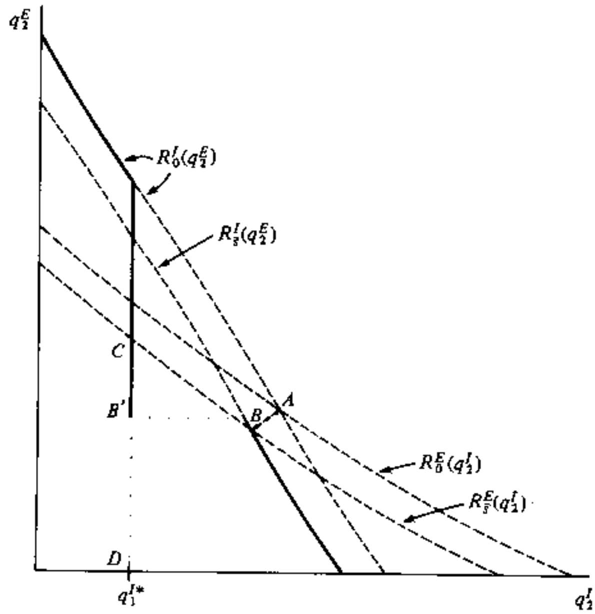

Fig. 3. Effect of size of switching cost on second-period equilibrium.

switching costs, the second-period equilibrium is at the intersection of $R_0^I(.)$ and $R_0^E(.)$, point A. If q_1^{I*} is less than I's output at A, then as we perform the experiment of raising switching costs slightly, each firm's reaction curve is moved in around the equilibrium point so that the equilibrium point moves in, in the direction of the arrows. (With no switching costs, or low switching costs, I will always increase output ('hold a sale') in equilibrium rather than restrict itself to its previous customer base q_1^{I*}.) Raising consumers' switching costs (for purchases both from the entrant and from the incumbent) is in this range formally equivalent to raising each firm's marginal costs so that, in general, each firm earns lower second-period profits.[1] As switching costs rise, however, the profitability for the incumbent of restricting itself to its previous customer base increases relative to the profitability for it of 'holding a sale', for any given entrant's output. Thus the horizontal gap in I's reaction curve falls vertically. Let $\bar{s}$ be the largest switching cost for which I's reaction curve with customer base q_1^{I*} includes the intersection of $R_{\bar{s}}^I(.)$ and $R_{\bar{s}}^E(.)$, that is, includes the intersection of the zero-customer-base reaction curves with switching cost $\bar{s}$.[2]

[1] See Seade (forthcoming) for conditions under which raising each firm's costs does not lower the firms' profits.

[2] The condition for I's second-period reaction curve to include the intersection of $R_s^I(.)$ and $R_s^E(.)$ is that $\Pi_2^I(q^I(s), q^E(s), q_1^{I*}) \geqslant \Pi_2^I(q_1^{I*}, q^E(s); q_1^{I*})$, in which $q^I(s)$ and $q^E(s)$ are I's and E's respective outputs in the second-period equilibrium that would result if neither firm had a customer base ($q_1^I = 0$). If, as is natural, both the entrant's output and the incumbent's second-period profits are decreasing in s in such a second-period equilibrium, then the left-hand side of the expression is decreasing in s and the right-hand side of the expression is increasing in s (for $q^I(s) > q_1^{I*}$). Under these conditions (which hold, for example, with linear demand and costs) there is a unique value of s below which I's second-period reaction curve always includes the intersection of $R_s^I(.)$ and $R_s^E(.)$ (and so second-period equilibrium is at this intersection) and above which I's reaction curve does not include this intersection. More generally, a largest switching cost $\bar{s}$ for which I's reaction curve includes this intersection always exists if q_1^{I*} is not too large.

Then the second-period equilibrium with switching cost $\bar{s}$ will be at this intersection, point B in Fig. 3.[1] For slightly higher switching costs s, however, the unique second-period Nash equilibrium is slightly below point C, where $R_s^E(.)$ intersects the line CB'. There is a discontinuous decrease in q_2^I and a discontinuous increase in q_2^E, hence also a discontinuous decrease in I's second-period profits and a discontinuous increase in E's profits. I's second-period output is lower, and both E's profits and E's output may be either higher or lower than in the equilibrium without switching costs, A. (With linear demand and costs E's profits are higher if and only if E's output is higher.[2]) The reduction in q_2^I helps E's profitability but the increase in switching costs hurts it. When switching costs are increased still further, the second-period equilibrium moves vertically down the line $CB'D$ towards the point D as I sticks to its first-period output q_1^{I*} and E steadily reduces its output. Further increases in switching costs beyond $\bar{s}$ thus reduce E's profitability and so reduce the value of fixed cost above which E will not enter.

The above analysis assumed for simplicity that q_1^I was determined independently of s and independently of the second-period equilibrium. If I's optimal first-period output with switching costs just above $\bar{s}$ but ignoring strategic considerations gave it a second-period customer base q_1^{I*} (say because of growing demand, falling costs, or turnover of customers), I would typically have an incentive to under-invest to restore the second-period equilibrium to the neighbourhood of point B. However, there would then be a level of switching costs at which it would no longer be worthwhile for I to under-invest to achieve an equilibrium like B, and at which the second-period equilibrium would move discontinuously to a point like C. Typically E's second-period profits would jump discontinuously upward at this level of switching costs, and it is not hard to construct examples in which E's second-period profits would then be higher than in a regime of no switching costs.

Thus, even accounting for the incumbent's strategic behaviour in period one, if switching costs are increased then an entrant's post-entry output and profits and so the fixed cost it is prepared to pay to enter the market may be increased, while the incumbent's output and profits may be decreased. These results can all arise even if demand is linear in both periods and firms' costs are linear and equal.[3] Examples are given in Section II. Though very high switching costs are the best deterrent to entry, switching costs in a middle range may be more favourable to entry than no switching costs or low switching costs. The intuition

[1] We are using the assumption that of the two second-period Nash equilibria in pure strategies, B and C, the one more profitable to I is played. (See note 1 on p. 103.) Changing the convention about which Nash equilibrium is played lowers the value of s at which the equilibrium changes discontinuously in s, but does not change the nature of the results.

[2] The reason is that with linear demand both q_2^I and s affect the intercept but not the slope of E's residual demand so that if in equilibrium E produces more then E's residual demand and hence E's profits must be higher.

[3] However these results cannot arise with linear demand and costs if firms' costs and the market demands are the same in both periods, and there is no turnover of consumers between the periods. The explanation is the same as for the under-investment results: for given s the incumbent will choose q_1^I greater than the I^{th} coordinate of the intersection of $R_s^I(.)$ and $R_s^E(.)$. The results in the text arise when a first-period output that is large relative to I's first-period demand and costs leads to I entering the second-period with a customer base that is small relative to second-period demand and costs.

is that the incumbent has a choice in the second period between a higher price that capitalises on repeat purchases but restricts it to a small share of the market, and a lower price that attracts new customers but fails to take advantage of its current customer base. Increasing switching costs makes the former strategy relatively more profitable, if entry occurs, but helps the entrant even more.

A Second-Mover Advantage

Consideration of Fig. 3 also shows the possibility of a 'second-mover advantage'. In a second-period equilibrium around point C, E produces more and may earn greater second-period profits than I even if the two firms' costs are identical. With demand growing over time, or costs falling over time, it is even possible that E's discounted second-period profits may exceed I's total profits over both periods, though the firms are identical in all respects except for I's ability to enter the market in the first period.[1] This case can arise with linear demand and costs and is exhibited in the next section. Firm I has the option of not entering the market in the first period in which case the firms produce identical second-period outputs and so earn identical profits. However I earns more profits by entering the market in the first period and building up custom that can be milked with a high second-period price, than it earns by waiting until the second period to enter and then competing head-to-head with E. Although I gains by its early entry, E gains even more from the reduced competition in the second period that leaves it with the low end of the market to itself. Thus the second mover – the firm unable to enter in the first period – earns more than the firm with the ability to choose whether or not to enter in the first period.

II. ILLUSTRATION: LINEAR DEMANDS AND COSTS

We use the case of linear demand $f_t(q) = \alpha_t - \beta q, \alpha_t > 0, \beta > 0$, and constant marginal costs that are the same for both firms, $C_1^I(q) = c_1 q$, $C_2^I(q) = C_2^E(q) = c_2 q$, to demonstrate that some of the less obvious outcomes can arise even with this natural demand and cost structure. For simplicity we assume no discounting, that is, $\lambda = 1$, and no fixed costs of entering the market, so $F = 0$. For reasons of space we will omit detailed numerical calculations; they can be found in Klemperer (1986).

We begin by deriving the firms' second-period reaction curves, and so solving for the second-period equilibrium resulting from any given first-period choice of q_1^I: If

$$q_1^I \leqslant \hat{q}_1^I(\alpha_2, \beta, c_2, s) = \{2(\alpha_2 - c_2) + s - \sqrt{[12(\alpha_2 - c_2)s - 3s^2]}\}/6\beta,$$

then

$$q_2^I = q_2^E = \left(\frac{\alpha_2 - c_2 - s}{3\beta}\right),$$

[1] If demands and costs are the same in period two as in period one and I and E have the same costs, then I's total profits must exceed E's profits, since I's profits cannot be less than a monopolist's profits in the first period and E's profits cannot be more than a monopolist's profits in the second period.

and so
$$\Pi_2^I = \Pi_2^E = \frac{(\alpha_2 - c_2 - s)^2}{9\beta}.$$

If
$$\hat{q}_1^I < q_1^I \leqslant \left(\frac{\alpha_2 - c_2 + s}{3\beta}\right),$$

then
$$q_2^I = q_1^I \quad \text{and} \quad q_2^E = \left(\frac{\alpha_2 - \beta q_1^I - c_2 - s}{2\beta}\right),$$

and so
$$\Pi_2^I = q_1^I\left(\frac{\alpha_2 - \beta q_1^I - c_2 + s}{2}\right) \quad \text{and} \quad \Pi_2^E = \frac{(\alpha_2 - \beta q_1^I - c_2 - s)^2}{4\beta}.$$

If
$$q_1^I \geqslant \left(\frac{\alpha_2 - c_2 + s}{3\beta}\right),$$

then
$$q_2^I = \left(\frac{\alpha_2 - c_2 + s}{3\beta}\right) \quad \text{and} \quad q_2^E = \left(\frac{\alpha_2 - c_2 - 2s}{3\beta}\right),$$ [1]

and so
$$\Pi_2^I = \frac{(\alpha_2 - c_2 + s)^2}{9\beta} \quad \text{and} \quad \Pi_2^E = \frac{(\alpha_2 - c_2 - 2s)^2}{9\beta}.$$

Thus $\hat{q}_1^I$ is the critical value of q_1^I below which I will 'hold a sale' in period two (choose q_2^I greater than q_1^I).

We note that $\Pi_2^I(q_1^I)$ falls discontinuously and $\Pi_2^E(q_1^I)$ rises discontinuously at $q_1^I = \hat{q}_1^I$, since q_2^E rises discontinuously and q_2^I falls discontinuously at this value. Thus for a given s, an increase in I's customer base, q_1^I, may increase the entrant's post-entry profits and decrease the incumbent's second-period profits – see Fig. 4. We also note that a larger s means a smaller $\hat{q}_1^I$, but that always $\hat{q}_1^I \leqslant (\alpha_2 - c_2)/3\beta$. Therefore for a given $q_1^I < (\alpha_2 - c_2)/3\beta$ an increase in s may increase the entrant's post-entry profits and decrease the incumbent's second-period profits – see Fig. 5.[2]

Given our knowledge of how I's second-period profitability depends on q_1^I, it is now easy to solve for the q_1^I that maximises I's total profits over both periods.[3] If both demands and costs are the same in period two as in period one ($\alpha_1 = \alpha_2 = \alpha, c_1 = c_2 = c$) we have limit pricing in the entire range
$$s \in \left(\frac{\alpha - c}{5}, \frac{2(\alpha - c)}{3}\right),$$

[1] These cases hold only if the specified q_2^E is greater than or equal to zero. If not, then in equilibrium $q_2^E = 0$. If the reaction curves cross twice, then we assume that the equilibrium with larger q_2^I obtains (see note 1 on p. 103). (Our results would be qualitatively unchanged by making the alternative assumption. In fact the less intuitive results, for example E making higher profits than I, would then hold for wider ranges of parameter values.)

[2] If $q_1^I \geqslant (\alpha_2 - c_2)/3\beta$ an increase in s always (weakly) decreases the entrant's post-entry profits and (weakly) increases the incumbent's second-period profits.

[3] Firm I's total profits are $(\alpha_1 - \beta q_1^I - c_1)\, q_1^I + [\alpha_2 - \beta(q_2^I + q_2^E) - s - c_2]\, q_2^I$ if $q_1^I \leqslant \hat{q}_1^I$, and $(\alpha_1 - \beta q_1^I - s - c_1)\, q_1^I + [\alpha_2 - \beta(q_2^I + q_2^E) - c_2]\, q_2^I$ if $q_1^I > \hat{q}_1^I$.

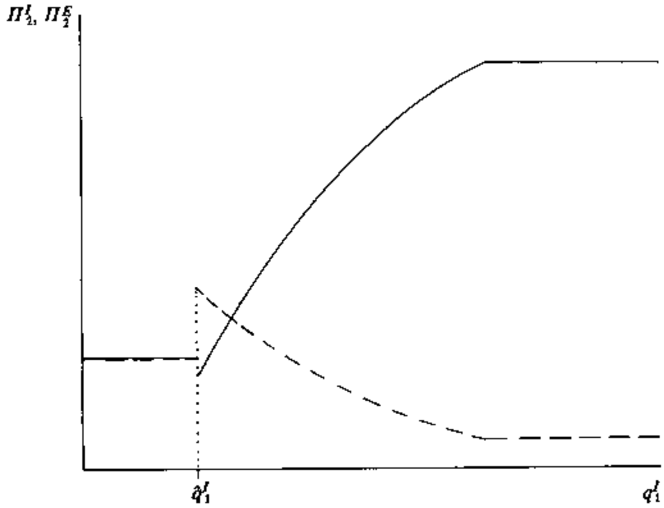

Fig. 4. Second-period profits as functions of incumbent's customer base (s given). [To scale — linear demand and costs.]

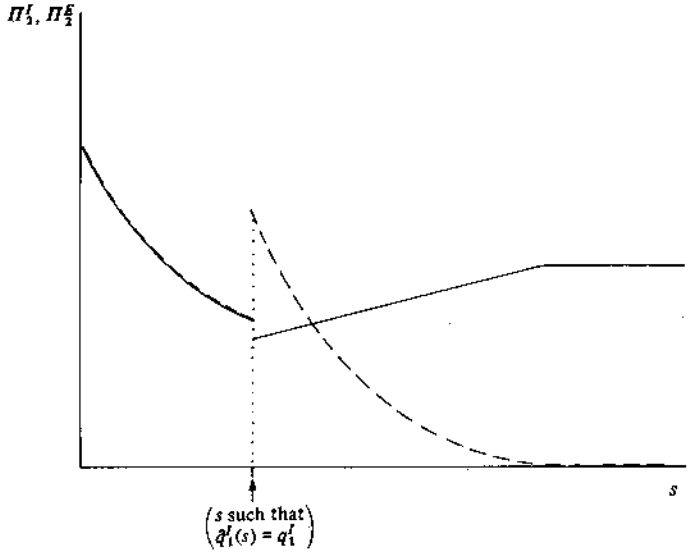

Fig. 5. Second-period profits as functions of switching cost (q_1^I given). [To scale — linear demand and costs.]

but as noted in Section I, I never under-invests.[1] We therefore focus on a case in which demands and costs are linear in both periods but *either* demand is higher in the second period than in the first *or* both firms' second-period costs are lower than the incumbent's first-period costs. We let $(\alpha_2-c_2)=9(\alpha_1-c_1)$. We can then compute that if $s<0{\cdot}39(\alpha_2-c_2)$, then

$$q_1^I=\frac{(\alpha_2-c_2)}{18\beta}<\hat{q}_1^I \quad \text{and} \quad q_2^I=q_2^E=\frac{(\alpha_2-c_2)-s}{3\beta}$$

in equilibrium. If $s\in(0{\cdot}39(\alpha_2-c_2),\ 0{\cdot}69(\alpha_2-c_2))$, then

$$q_1^I=\hat{q}_1^I \quad \text{and} \quad q_2^I=q_2^E=\frac{(\alpha_2-c_2)-s}{3\beta}$$

in equilibrium. If $s\in(0{\cdot}69(\alpha_2-c_2),\ 0{\cdot}96(\alpha_2-c_2))$, then

$$q_1^I=q_2^I=\left(\frac{11(\alpha_2-c_2)-9s}{54\beta}\right) \quad \text{and} \quad q_2^E=\left(\frac{43(\alpha_2-c_2)-45s}{54\beta}\right)$$

in equilibrium.[2] Figs. 6 and 7 show the two firms' outputs and profits as functions of s. We can see that increasing s above the critical value of $0{\cdot}69(\alpha_2-c_2)$ increases E's equilibrium profits and output while reducing I's equilibrium total profits and second-period output. Firm I's profits over both periods together exceed E's if $s<0{\cdot}69(\alpha_2-c_2)=6{\cdot}23(\alpha_1-c_1)$ or if $s>0{\cdot}70(\alpha_2-c_2)=6{\cdot}30(\alpha_1-c_1)$. However there is a range of values of s for which I's profits over time are lower than E's.

We also note that for a range of s below $0{\cdot}69(\alpha_2-c_2)$, I chooses its first-period output as low as $\hat{q}_1^I$ only for the strategic reason of credibly committing to a 'sale' $(q_2^I>q_1^I)$ in the next period and thereby reducing E's second-period output. Thus (in equilibrium) I is setting a lower output and higher price than it would if it ignored the strategic implications of its decisions, that is, it is under-investing in its first-period customer base or limit over-pricing.[3] For $s>0{\cdot}69(\alpha_2-c_2)$, on the other hand, I is setting a higher output and lower price than it would if it were ignoring the strategic implications of its behaviour, that is limit pricing, as is more usual.

[1] For $s\leqslant(\alpha-c)/5$, I's equilibrium second-period output is below q_1^I. For $s\geqslant 2(\alpha-c)/3$, entry is blockaded in the sense of Bain (1956), so I chooses the q_1^I that would be optimal if there were no threat of entry and E finds entry unattractive. In the range $s\in(2(\alpha-c)/7,\ 2(\alpha-c)/3)$ I is also setting a higher output and lower price than would a monopolist unthreatened by entry in the second period.

[2] If $s\geqslant 43(\alpha_2-c_2)/45\approx 0{\cdot}96(\alpha_2-c_2)$, then E will not enter the market in equilibrium. It is always optimal for I to enter the market in the first period if $s<(\alpha_2-c_2)$. Even if $(\alpha_1-c_1)<s<(\alpha_2-c_2)$, choosing $q_1^I=0$ is dominated by choosing $q_1^I=\min\ ((\alpha_1-c_1)/2\beta,\ \hat{q}_1^I)$ which leads to the same second-period equilibrium, $q_2^I=q_2^E=(\alpha_2-c_2-s)/3\beta$, and also earns positive profits in the first period since consumers expect a period-two 'sale' and so ignore their switching costs in computing their first-period willingness to pay (see note 2 on p. 101; $\lambda=1$ here).

[3] Taking $q_2^E=(\alpha_2-c_2-s)/3\beta$ as given, I would choose $q_1^I=q_2^I=(7(\alpha_2-c_2)-6s)/36\beta>\hat{q}_1^I(s)$ for s not too much less than $0{\cdot}69(\alpha_2-c_2)$. In open loop equilibrium, $q_1^I=q_2^I=(11(\alpha_2-c_2)-9s)/63\beta>\hat{q}_1^I(s)$, $q_2^E=(26(\alpha_2-c_2)-27s)/63\beta$.

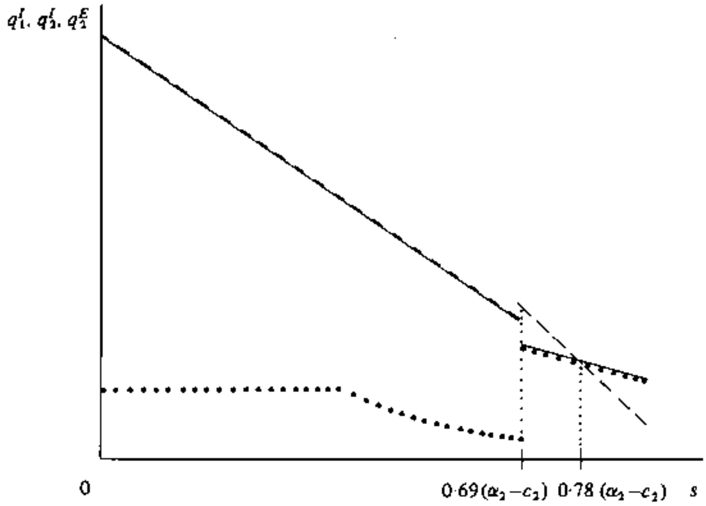

Fig. 6. Equilibrium outputs as functions of switching cost.
[To scale for example with linear demands and costs.]
Incumbent's first-period output, in equilibrium, q_1^I ....
Incumbent's second-period output, in equilibrium, q_2^I ——
Entrant's second-period output, in equilibrium, q_2^E - - - - -

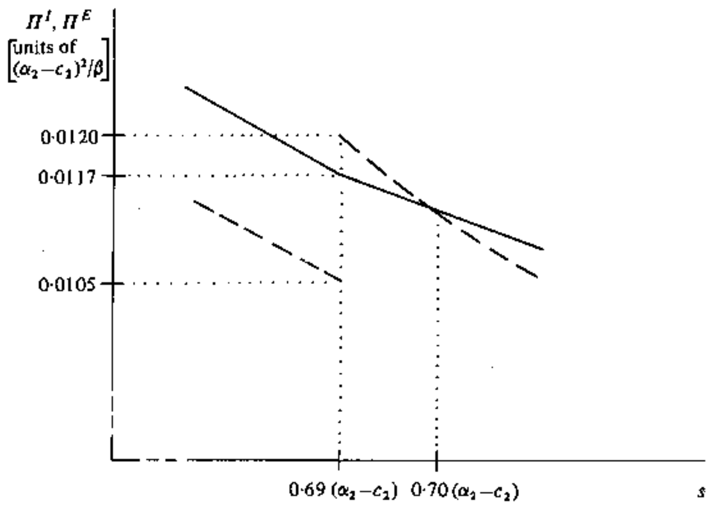

Fig. 7. Equilibrium profits over time as functions of switching cost.
[To scale for example with linear demands and costs.]
Incumbent's total profits over time, in equilibrium, Π^I ——
Entrant's total profits over time, in equilibrium, Π^E - - - -

So far we have assumed E has no fixed cost of entering the market. We can also illustrate how an incumbent may use either very high or very low customer bases to deter entry by considering the case in which E will not enter the market if its potential profits are less than a fixed cost, F. Fig. 8 shows I's equilibrium

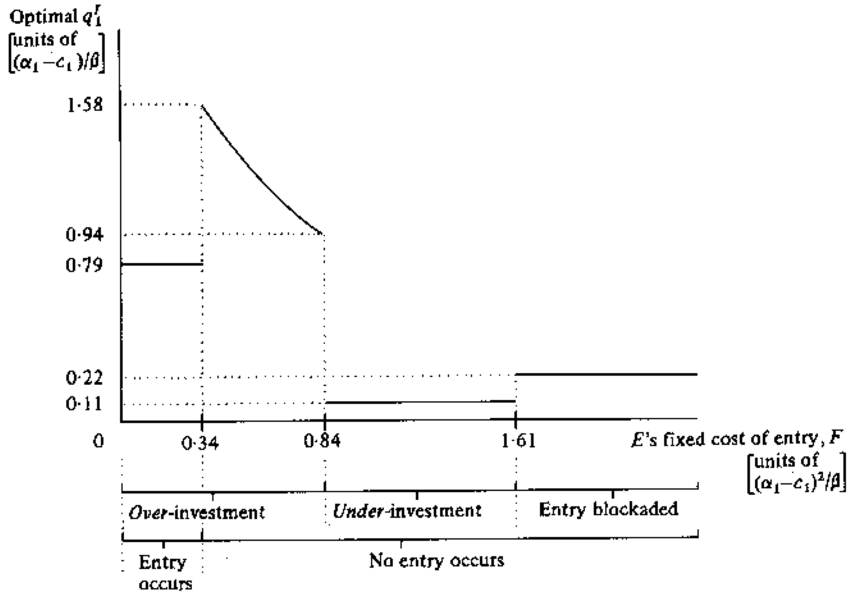

Fig. 8. Incumbent's first-period output as function of potential entrant's fixed cost of entry. [To scale for example with linear demands and costs.]

first-period output as a function of F for our example with $(\alpha_2-c_2) = 9(\alpha_1-c_1)$ and taking $s = 6{\cdot}25(\alpha_1-c_1)$.[1]

III. CONCLUSION

Consumer switching costs can explain the phenomenon of limit pricing: an incumbent monopolist may charge lower prices than would a monopolist unthreatened by entry, in order to persuade more customers to 'buy in' to its product and so deter entry. A similar logic can also lead an incumbent faced by certain entry to limit price and over-invest in its pre-entry customer base because of the strategic advantage that this confers in post-entry competition.[2] Thus far the story is formally identical to Dixit's (1980) model in which an

[1] For $F > 1{\cdot}61\ (\alpha_1-c_1)^2/\beta$, entry is blockaded and I chooses q_1^I as would a monopolist facing no threat of entry. For smaller F, under-investing to block entry dissipates less of I's profits than over-investing to block entry (in this example) but cannot reduce the entrant's potential profits below a duopolist's profits in a market in which neither firm has a customer base. For $F < 0{\cdot}84\ (\alpha_1-c_1)^2/\beta$, I can block entry only by over-investing, and the lower is F the more I must over-invest in the first period to block entry, since $q_2^I = q_1^I$ in an over-investment equilibrium. For $F < 0{\cdot}34\ (\alpha_1-c_1)^2/\beta$, it is not worthwhile for I to block E's entry, but even in this case I is over-investing to reduce the scale of E's entry. (If I took E's second-period output as given at its equilibrium value, I would choose $q_1^I = 0{\cdot}69\ (\alpha_1-c_1)/\beta$; in open loop equilibrium I would choose $q_1^I = 0{\cdot}68\ (\alpha_1-c_1)/\beta$.)

[2] The same logic suggests that predatory pricing may also be rational after entry. The basic point, again, is that stealing customers in one period also steals them for future periods, in contrast to the classical situation in which a firm that loses customers (or is driven from the market) in one period can easily recapture those customers (or re-enter the market) later when the predator raises its price. We could also develop more formally the view (see, for example, Fogg (1974)) often expressed by businessmen that it is easier for a firm to increase its market share in a growing market than in a static one in which new business can come only from poaching another firm's custom.

incumbent builds capacity in advance of threatened entry to reduce the extent of or block that entry. Capacity that reduces second-period costs of production by s per unit up to the amount of the capacity in place has a similar effect to switching costs that increase consumers' second-period willingness-to-pay by s per unit up to the amount of the previous period's output. Whether the incumbent's incentive to maintain output derives from a saving on the cost side or from a bonus on the demand side makes little difference. Both stories at least partially justify Sylos-Labini's (1962) original explanation that limit pricing works because potential entrants assume an incumbent will maintain its output after entry. Within the range in which limit pricing is effective (and, in the switching costs story, if there is no turnover of customers in the market), the incumbent's post-entry output will be its pre-entry output.

However, in Dixit's model of precommitment on the cost side, increasing second-period output beyond first-period capacity does not reduce the cost savings realised up to capacity. In a switching-cost model, increasing second-period output beyond first-period sales does reduce a firm's demand-side bonus because repeat customers must be charged the same price as new customers. (If the incumbent could costlessly price discriminate by the entire switching cost s between new and repeat customers, our model would be formally equivalent to Dixit's. However, if the maximum amount of costless price discrimination possible is less than s, the results of this paper would still obtain.) Thus precommitment on the demand side, unlike precommitment on the cost side, is a double-edged sword. Switching costs not only lock in customers but may also lock in an incumbent to its current customer base. Higher switching costs, or more bought-in customers, make it less likely that it will be rational for the incumbent to react aggressively to new entry by competing for new customers rather than confining itself to 'skimming the market', that is, milking its repeat customers. This in turn makes entry more likely, so that the incumbent may wish to reduce its first-period customer base or 'limit over-price' to deter entry. Sufficiently high switching costs, or a sufficiently large customer base, however, must make entry less profitable, so entry is more deterred by very large or very small switching costs and by very large or very small pre-entry outputs than by switching costs and pre-entry outputs in a middle range. With demand growing over time, building a large customer base relative to post-entry demand is very costly, and building a very small customer base is wasteful of early profits. In this case the firm that is able to enter the market first may actually earn less discounted total profits than a firm that is not able to enter the market until a later stage.

This work has considered the effects on entry of exogenously given switching costs; further work should more directly address firms' incentives to create them.

St. Catherine's College and Institute of Economics and Statistics, Oxford

References

Bain, Joe S. (1949). 'A note on pricing in monopoly and oligopoly.' *American Economic Review*, vol. 39 (March), pp. 448–64.

—— (1956). *Barriers to Entry of New Competition*. Harvard: Harvard University Press.

Baldani, Jeffrey. (1983). 'Strategic advertising and credible entry deterrence policies.' Working Paper, Colgate University, (November).

Bulow, Jeremy I., Geanakoplos, John D. and Klemperer, Paul D. (1985). 'Multimarket oligopoly: strategic substitutes and complements.' *Journal of Political Economy*, vol. 93 (June), pp. 488–511.

Dixit, Avinash. (1980). 'The role of investment in entry-deterrence.' ECONOMIC JOURNAL, Vol. 90 (March), pp. 95–106.

Farrell, Joseph von R. (1985). 'Competition with lock-in.' Working Paper GTE Laboratories, (November).

Fogg, C. Davis. (1974). 'Planning gains in market share.' *Journal of Marketing*, vol. 38 (July), pp. 30–8.

Fudenberg, Drew and Tirole, Jean. (1984). 'The fat-cat effect, the puppy-dog ploy, and the lean and hungry look.' *American Economic Review*, vol. 74 (May), pp. 361–6.

Gaskins, Darius W. (1971). 'Dynamic limit pricing: optimal pricing under threat of entry.' *Journal of Economic Theory*, vol. 3 (September), pp. 306–22.

Klemperer, Paul, D. (1984). 'Collusion via switching costs.' Research Paper 786, Graduate School of Business, Stanford University, (May).

—— (1986). 'Markets with consumer switching costs.' PhD Thesis, Stanford University, (September).

Matthews, Steven A. and Mirman, Leonard J. (1983). 'Equilibrium limit pricing: the effects of private information and stochastic demand.' *Econometrica*, vol. 51 (July), pp. 981–96.

Milgrom, Paul and Roberts, John. (1982). 'Limit pricing and entry under incomplete information: an equilibrium analysis.' *Econometrica*, vol. 50 (March), pp. 443–59.

Saloner, Garth. (1984). 'Dynamic equilibrium limit-pricing in an uncertain environment.' Working Paper 342, Massachusetts Institute of Technology, (February).

Schmalensee, Richard. (1982). 'Product differentiation advantages of pioneering brands.' *American Economic Review*, vol. 72 (June), pp. 349–65.

—— (1983). 'Advertising and entry deterrence: an exploratory model.' *Journal of Political Economy*, vol. 91, pp. 636–53.

Seade, Jesus (forthcoming). 'Profitable cost increases and the shifting of taxation: equilibrium responses of markets in oligopoly.' *Journal of Public Economics*.

Spence, A. Michael. (1981). 'The learning curve and competition. *Bell Journal of Economics*, vol. 12 (Autumn), pp. 49–70.

Summers, Lawrence, H. (1985). 'On frequent-flyer programs and other loyalty-inducing economic arrangements.' Working Paper, Harvard University, (August).

Sylos-Labini, Paolo (1962). *Oligopoly and Technical Progress*. Harvard: Harvard University Press.

Weizsäcker, C., Christian von. (1984). 'The costs of substitution.' *Econometrica*, vol. 52 (September), pp. 1085–116.