RAND Journal of Economics
Vol. 22, No. 1, Spring 1991

Preemptive investment, toehold entry, and the mimicking principle

David A. Malueg*

and

Marius Schwartz**

This article explores the ability of firms to successfully enter against an initial monopolist by pursuing "toehold entry": sinking entry costs early and planning to produce only later. We endow the incumbent with strong first-mover advantages; for example, capital is the only input and is completely sunk, and in each period the incumbent invests before entrants. Nevertheless, the incumbent's profit can easily be negative if it tries to deter all entry when the market is growing. The incumbent may need prohibitive capital in place early on, because once toehold entry has occurred the incumbent's planned capital expansion might no longer be rational. Surprisingly, even if market conditions remain stationary, the incumbent may still allow some entry. In the model some firms are always deterred, showing that several firms can deter others more effectively than can a single firm. The difference arises because one firm cannot commit to mimicking the (aggregate) investment path that several independent firms would choose. This "mimicking principle" offers a general perspective for understanding when a monopolist finds deterrence optimal.

1. Introduction

■ There is an extensive literature discussing how an incumbent monopolist might invest preemptively in order to deter all future entry, including Spence (1977) and (1979), Dixit (1979) and (1980), Eaton and Lipsey (1980) and (1981), and Gilbert and Harris (1984). The basic idea is that a firm's capital can be quite inflexible downwards, so that by investing preemptively the monopolist commits itself to a higher capital stock and thus to a lower (short-run) marginal cost function, at least over some output range. The lowered marginal cost makes the incumbent a tougher competitor in any future interaction, thereby reducing expected profit from entry.

* Economic Analysis Group, Antitrust Division, U.S. Department of Justice.

** Georgetown University and Economic Analysis Group, Antitrust Division, U.S. Department of Justice.

The views expressed are not necessarily those of the U.S. Department of Justice. Helpful suggestions were received from Curtis Eaton, Maxim Engers, Dale Henderson, Bruce Kobayashi, Robert Reynolds, two referees and a co-editor of this Journal, and seminar participants at the University of Virginia, the Federal Trade Commission, the Antitrust Division of the U.S. Department of Justice, Simon Fraser University, and Georgetown University.

While preemptive investment therefore can credibly reduce the profitability of entry, important drawbacks of this deterrence strategy have been noted. One set of drawbacks is technological. The scope for preemptive investment to alter marginal cost irreversibly will be limited by the relative importance of capital costs and by the extent to which capital is durable and firm-specific (Eaton and Lipsey, 1981; Gilbert, 1986; Saloner, 1986). And even if high capital would yield the required low marginal cost, adjustment costs could prevent the monopolist from accumulating sufficient capital in time to deter all entry (Spence, 1979; Fudenberg and Tirole, 1983).

Another drawback to using preemptive investment to deter all entry (hereinafter, "deter entry" means deter *all* entry) is that a high capital stock (and hence low marginal cost) might not make the initial monopolist sufficiently aggressive to render a potential entrant's profit negative. Given any capital choices by the incumbent and entrant, the market outcome can be quite sensitive to the particular second-stage interaction assumed. For example, if the expected market interaction is Cournot, then an incumbent may be unable to deter entry, no matter how large the incumbent's capital, because a large capital does not commit the incumbent to supplying a correspondingly high output (Dixit, 1980).[1]

This article focuses on a different type of drawback, one that is inherent in the very idea of preemptive investment. Capital provides commitment because it is costly to reduce once it is in place. The flip side of this is that to provide commitment, capital must be in place—it is implausible to assume that a firm can commit itself to expand its future capital regardless of rivals' interim actions. This inability to commit to future expansion makes the incumbent vulnerable to "toehold entry": an entrant incurs the fixed costs of entry and temporarily remains small, planning to expand later. Once the entrant has sunk the fixed costs, the incumbent's future behavior changes fundamentally. Recognizing that the entrant will be active as long as price exceeds average production cost, the incumbent might not find its planned expansion optimal once the toehold has been established.

One would expect toehold entry to be particularly effective in growing markets. The main purpose of this article is to show that indeed preemptive investment cannot always profitably deter toehold entry into such markets. In our examples, to deter toehold entry the incumbent would have to install a prohibitively costly level of capital ahead of market expansion. A by-product of our discussion is that it highlights a unifying principle to help explain why some entry-deterrence models find deterrence to be optimal while others do not. The different findings can be traced to a "mimicking principle," which considers whether the incumbent can deter entry by duplicating the market outcome that would have emerged with entry.

The remainder of the article is organized as follows. Section 2 presents the model. In order to isolate the effect arising from the inability to commit to future capital expansion, we set aside the limitations of preemptive investment discussed earlier (technological constraints and the slack between capital and output levels). We do so by considering a technology that is favorable to deterrence and by endowing the incumbent monopolist with rather strong first-mover advantages. Section 3 shows that toehold entry nevertheless can overcome these first-mover advantages when the market expansion is sufficiently large. More surprisingly, Section 4 shows that deterring all entry can be suboptimal (though not strictly unprofitable) even when the market is stationary. Section 5 compares our work with studies that find deterrence through preemptive investment to be optimal and others that do not, showing how the differences can be understood in terms of the mimicking principle.

[1] This is easily seen. The incumbent's marginal cost is nonnegative whatever its capital. Under standard conditions, each firm's Cournot output is decreasing in the rival's output. Therefore, given any positive capital of the rival, the initial incumbent's Cournot-duopoly output will be less than the zero-cost monopoly output, and, for suitable parameters, even the latter might not suffice to deter.

2. The model

■ The model we use is an extension of the following familiar one-period model (e.g., Omori and Yarrow (1982), Gilbert (1986), and Vives (1988)). Consider a homogeneous and perishable good with stationary flow inverse demand $P(q)$. Production requires a single divisible input, capital (k), which is infinitely durable and has no resale value. Capital and output are measured in the same units; thus, one additional unit of capital can add up to one unit to the firm's output stream. In addition to capital expenses (which *ex ante* are variable), any entering firm sinks a positive fixed cost, F.[2] This technology is available to all firms.

Time ranges from 0 to ∞. Choices are made in an exogenous sequence before the market opens. The incumbent is identified as firm 1 and the potential entrants as firms 2, 3, ..., N_F. Firm 1 chooses its capital level and constant output rate; having observed firm 1's choices, firm 2 decides whether to sink F and, if so, chooses capital and output, and so on through firm N_F. The market price is determined by total output chosen. We assume that N_F is sufficiently large so that maximum industry revenue would not cover the fixed costs of all potential entrants. Therefore, for any F, there will always be some firms that remain idle.

Once capital is installed, marginal cost is zero up to capacity, since capital has no resale value. Long-run marginal cost, however, is positive and equals c, the instantaneous interest cost of a unit of capital (the sole input). Let r denote the interest rate and define $\bar{q}$ as the output that satisfies

$$\max_{q} \{P(\bar{q} + q) - c\} q = rF.$$

That is, $\bar{q}$ is the output that just deters an additional entrant. (Throughout we assume that a firm enters if and only if it expects to earn strictly positive profit.) This is the "limit output" associated with Bain and Sylos-Labini. (See Modigliani (1958) for an exposition of the Bain/Sylos-Labini model.)

It is easy to see that in this setting the incumbent, firm 1, will choose to deter entry. All that matters for the profit of a potential entrant is the total output chosen by all prior movers—the number of previous entrants is irrelevant. Given the assumption of many potential entrants, any total output below $\bar{q}$ would induce additional entry. Given the natural-monopoly cost structure, firm 1 prefers to supply at least $\bar{q}$ rather than allow further entry.

We now modify the above model so that there are *two* periods: the first consists of the time interval $[0, T)$; the second, $[T, \infty)$. All decisions are made at the beginning of a period, that is, at 0 and T. For simplicity, during a period demand and costs are stationary and output is sold continuously at the constant rate chosen for the period. An entering firm must sink F only once, at time of entry; thus, sinking F can be viewed as giving a firm the option to produce in the market, now or in the future. At date T, demand can increase or long-run marginal cost can decrease, and any such market expansion is foreseen at date 0. Firms move in the same order at date T as at date 0. The main purpose of this assumption is to give the incumbent a strong first-mover advantage. (It is less important that the other firms move in the same order. Our results do not hinge on this.)

At time T, a firm cannot reduce its capital (which is durable and sunk) but can increase it to any level without incurring adjustment costs. Firm j's capital and output choices at date t are denoted by k_{jt} and q_{jt}, respectively. In each period, a firm's output cannot exceed capacity but can be less: $q_{jt} \leq k_{jt}$, $t = 0, T$. (Maintaining excess capacity is never rational

[2] F can be the cost of learning the technology, or any other overhead cost. In the cable TV industry, for example, F includes the cost of a "headend" that receives and transmits over-the-air signals, while k includes the cost of cabling a unit distance (Smiley, 1986).

in the single-period model described earlier, since a firm only chooses capacity and output once; but in the two-period model, excess capacity might be rational in the first period if expansion is planned in the second.)

We assume that a firm's second-period *output* rate is no less than its first: $q_{jT} \geq q_{j0}$. Thus, a firm can commit to supplying no less than a given future output, regardless of other firms' subsequent choices, by choosing that output initially. One rationale for this "minimum-output" assumption is that a firm might be able to sign current customers to meeting-competition contracts that bind the firm to meeting a rival's price and bind the customers to continue purchasing at least the initial quantity from the firm at the matched price. Mainly, however, the minimum-output assumption is a modelling device. We already know that deterrence can be infeasible if a firm's high capital does not commit it to supplying a high output (see footnote 1), and we want to abstract from this possibility. In our model, the incumbent certainly *can* deter—by choosing sufficiently high capital and output at date 0.

Summarizing, we have modified the one-period model, in which deterrence is optimal for any fixed cost F, in a way that retains strong first-mover advantages for the incumbent: there are no adjustment costs, the incumbent moves first at both dates, and the incumbent can commit to supplying a high future output by producing the corresponding output initially before potential entrants can move. The incumbent, however, cannot threaten future expansion unconditionally; capital and output are expanded at T only if expansion is rational at that date given the observed prior choices. Will the incumbent still find it optimal to deter all entry? We know that new entry will not occur at date T, since then we are back in a single-period case for which it is never optimal to allow further entry. So if entry does occur, it must be at date 0. But what incentive does the incumbent have to allow entry at date 0, given that it also has the first-mover advantage in the future? The next two sections address this question, first when the market expands at time T, and second when the market remains stationary.

3. Market expansion

■ Let c_0 and c_T denote the long-run marginal cost of producing output using capital purchased at time 0 and T, respectively.[3] Let $\bar{q}_T$ denote the quantity that just deters entry at time T, i.e., $\bar{q}_T$ satisfies

$$\max_q \{P_T(\bar{q}_T + q) - c_T\}q = rF,$$

where $P_T(q)$ denotes inverse demand at time T. If no new entry has occurred at time 0 and if firm 1 wishes to remain a monopolist, then at time T its output must be at least $\bar{q}_T$. Firm 1 would ideally like to choose capital and output below $\bar{q}_T$ at time 0 (when demand is lower or cost is higher than at T) and expand to $\bar{q}_T$ at time T. But if it installs too little capital at time 0, then a second firm can enter on a small scale, planning to expand later. Such entry can reduce firm 1's own planned expansion. To prevent this prepositioning by a rival, firm 1 would have to install the requisite capital in the first period, before others have a chance to act. But installing such large capital ahead of market expansion can prove too costly.

□ **An example.** To illustrate the general ideas we begin with an example in which costs are stationary but demand increases at date T, and in which it is optimal not to carry excess capacity. The more general case is addressed later.

[3] Thus, $c_t = rA_t$, where A_t is the price at date t of the amount of capital needed to produce a unit output stream, $t = 0, T$. A drop in long-run marginal cost ($c_T < c_0$) could be due to a fall in the price of capital or to technical change that increases the output/capital ratio. The latter, in turn, might be embodied only in capital purchased at time T or might be transferable also to capital purchased at time 0. The proposition we shall prove covers all such cases of cost reduction, as well as growing demand.

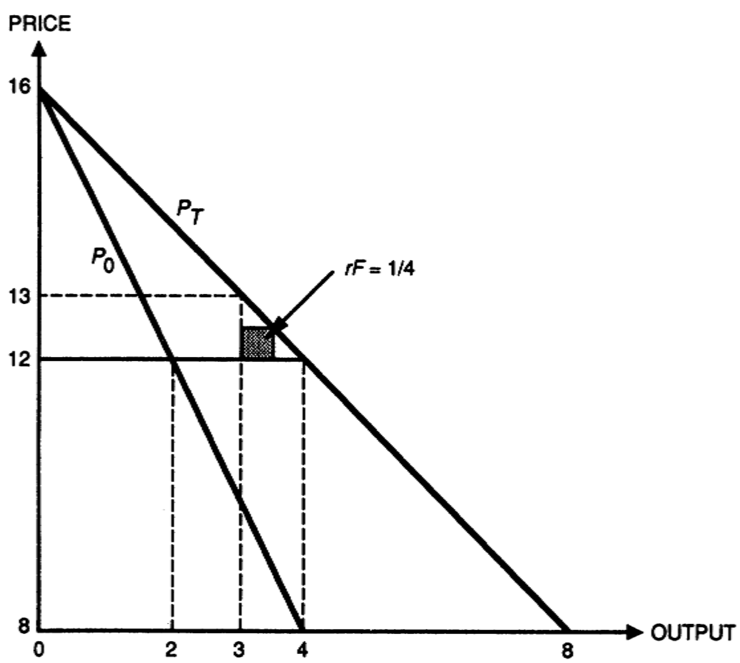
Referring to Figure 1: $P_0(q) = 16 - 2q$, $P_T(q) = 16 - q$, $c_K = c = 12$, and $rF = 1/4$. Thus, $\bar{q}_T = 3$. Observe that with two firms the Stackelberg leader output at date T is 2 and the follower output is 1, summing to 3 and sufficing to deter further entry.

A firm's value at date 0, our reference point from now on, can be expressed as a weighted sum of its (appropriately computed) constant profit streams in the first and second periods, the weights being $w_0 = (1 - e^{-rT})/r$ and $w_T = e^{-rT}/r$. Assume that the weights satisfy $w_0/w_T \in (1.2, 1.25)$. We first show that firm 1 would be unprofitable for $k_{10} = 2.5$ even if that capital did deter all potential entrants. We then show that $k_{10} = 2.5$ in fact is not sufficient to deter firm 2. Throughout we treat firm 1 as the incumbent that has already sunk F, and we consider only its gross profit (showing that even gross profit is negative if it attempts to deter all entry).

Suppose firm 1 were to choose $k_{10} = 2.5$ and all entry at time 0 were deterred. In the first period, firm 1 would utilize its capacity fully, $q_{10} = k_{10}$ (since marginal revenue is positive until $q_{10} = 4$ and marginal cost is zero once capital is installed). Thus, $p_0 - c = -1$, implying that first-period profit flow is -2.5. At time T, firm 1 would choose $k_{1T} = q_{1T} = 3$, yielding $p_T - c = 1$ and hence a second-period profit stream of 3. If $k_{10} = 2.5$ did deter all entry, firm 1's value at time 0 therefore would be $V_1 = -2.5w_0 + 3w_T$. Given $w_0/w_T > 1.2$, it follows that $V_1 < 0$.

Now suppose, given firm 1's choice of $k_{10} = q_{10} = 2.5$, that firm 2 pursues the strategy of "toehold entry": at date 0, sink F and choose zero capital and output; at date T, choose capital and output optimally given the history of the game. Firm 2's sinking F at time 0 fundamentally changes the game at time T.[4] Had no entry occurred at date 0, then at

FIGURE 1



[4] The assumption that all entry costs can be incurred well ahead of production is purely for convenience. Suppose that if a firm enters at date 0 but starts production at date T, it must incur at date T an additional fixed cost αF (e.g., because know-how depreciates if production does not start at entry). The example below shows that for $\alpha = 0$, firm 1 is strictly unprofitable if it tries to deter firm 2's toehold entry; by continuity, this holds also for some $\alpha > 0$.

time T firm 1 would have expanded capital and output to $\bar{q}_T\,(=3)$ in order to deter all entry in the last period. But given that firm 2 has sunk F, firm 1 will not expand beyond 2.5.

Rather, the game evolves as follows. Assuming that nobody other than firm 2 enters at time 0 (which we demonstrate shortly), at date T firm 1 maintains capital and output at 2.5 and firm 2 selects its "myopic" best response to this, namely .75. To see why, note that these outputs collectively suffice to deter further entry, and that 2.5 already exceeds the Stackelberg leader output for time T. These choices imply that firm 2's value at time 0 is

$$V_2 = -w_0 rF + w_T\{(16 - (2.5 + .75) - 12)(.75) - rF\} = (-4w_0 + 5w_T)/16.$$

Given $w_0/w_T < 1.25$, it follows that $V_2 > 0$.

Next we show that firm 3 (or any later mover) cannot enter profitably given any $k_{10} = q_{10} \geq 2$ and firm 2's toehold entry. If firm 3 enters, it will do so at time 0 and we can bound its profit by

$$V_3 \leq w_0(0 - rF) + w_T\{½ - rF\}.$$

The $-rF$ in each period represents the interest on the fixed cost. The bounds on gross profits, 0 and ½, emerge as follows. Firm 3's maximum gross-profit flow in the first period is at most 0 because firm 1 has chosen $q_{10} \geq 2$, implying $p_0 \leq 12 = c$. Now turn to the second period. Consider three firms choosing outputs sequentially at T, unconstrained by first-period choices. The outputs would be 2, 1, and ½ (which suffice to deter further entry), implying $p_T - c = ½$. The second-mover's gross profit in this one-period game would be ½. In the actual game, firm 1 is committed to $q_{1T} \geq q_{10} \geq 2$, so the second-mover's profit at T will not exceed ½. And firm 3 can do no better than this at date T (regardless of what choices firm 2 and firm 3 make at date 0). Since $rF = ¼$, the above bound is negative for $w_0 > w_T$. Thus firm 3 indeed stays out against the proposed strategies of firms 1 and 2, confirming that firm 2's entry is profitable.

Summarizing, for $w_0/w_T \in (1.2, 1.25)$ a capital stock $k_{10} = 2.5$ both fails to deter firm 2's toehold entry and makes firm 1 unprofitable. Observe the role played by the different bounds on the relative weights. The lower bound ensures that if firm 1 does deter, its first-period loss outweighs its second-period profit; the upper bound ensures that firm 2's first-period loss under toehold entry does not outweigh its second-period gain. It is easy to verify that any $k_{10} < 2.5$ would not deter, while any $k_{10} > 2.5$ would make firm 1 unprofitable. Therefore, firm 1 cannot profitably deter all entry.

To complete the example, note that firms 1 and 2 *can* profitably deter all others. Suppose $q_{10} = k_{10} = 2$ and firm 2 enters with $k_{20} = 0$. We saw that firm 3 cannot profitably enter and that second-period choices will be $q_{1T} = k_{1T} = 2$ and $q_{2T} = k_{2T} = 1$. Thus $V_2 = -w_0 rF + 3rF$, which is positive for $w_0 < 3w_T$. Firm 1 is even more profitable than firm 2 since its first-period gross profit equals firm 2's (zero), and its second-period profit is larger than firm 2's since its output is larger.

☐ **Large market expansion.** One might expect that monopolies will be less likely to persist when fixed entry costs are small relative to market expansion, since a monopolist's losses from installing capital ahead of market expansion would then be large relative to the later profit as a (deterring) monopolist. This reasoning is inconclusive, however, if an entrant's losses from early entry also would be large, since in that case deterrence might require only a moderate level of capital initially. In order for market expansion to favor entrants, we therefore require some asymmetry between the deterrence and entry strategies. This asymmetry is provided by toehold entry. The following proposition gives sufficient conditions for toehold entry to overcome an incumbent's first-mover advantage in installing capital.

Define q_{ct} as the quantity that satisfies $P_t(q) = c_t$, $t = 0, T$. Thus, q_{ct} is the "competitive output" given time-t conditions. Growing demand or falling marginal cost will typically yield $q_{cT} > q_{c0}$. This can be interpreted as market expansion. We assume that P_T is strictly decreasing at all outputs where P_T is strictly positive, and that $qP_t(q)$ is a strictly quasi-concave (hence single-peaked) function of q over the range where P_t is strictly positive, $t = 0, T$. For technical reasons we must also assume that in an open neighborhood of q_{cT}, the function P_T is differentiable and the derivative is continuous at q_{cT}.

Proposition. Suppose that (a) $0 < w_0 < w_T$; (b) $0 < c_0$, $c_T \leq c_0$; (c) $P_0(q) \leq P_T(q)$ for all q; and (d) $q_{cT} > q_{c0}$. For all sufficiently small F, if firm 1 deters all entry, then its value is negative.

The proof of the proposition is given in the Appendix, but we sketch it here. Fix some market expansion, $q_{cT} > q_{c0}$, and consider what happens as the entry cost F becomes small. (The following logic should make it clear that a similar proposition holds if instead we fix F and consider a sufficiently large market expansion. It is the *relative* size of these two parameters that is important.) As F goes to zero, $\bar{q}_T$ approaches q_{cT}, so firm 1's second-period profit as a deterring monopolist becomes arbitrarily small. To prevent toehold entry, at date 0 firm 1 must install a capital level of at least $\tilde{q}$ (defined in the proof). The quantity $\tilde{q}$ is less than $\bar{q}_T$, but as F goes to zero both $\bar{q}_T$ and $\tilde{q}$ go to q_{cT}. For F sufficiently small, it follows that $\tilde{q}$ exceeds q_{c0}, the "competitive output" for time-0 conditions. In the first period, therefore, firm 1 incurs a loss.[5] Given any positive weight on the first period, this loss can outweigh the second-period profit, since for sufficiently small F, second-period profit is arbitrarily small. It is true that a toehold entrant also earns only small profit in the second period; but unlike the incumbent, its first-period loss also is small: it is limited to the interest cost of F. As long as the present is not too important, therefore, toehold entry is profitable if firm 1 chooses first-period capital below $\tilde{q}$. A sufficient condition for toehold entry to be profitable when $k_{10} \leq \tilde{q}$ is $w_0 < w_T$.[6]

Two remarks are in order. First, firm 1's inability to deter all entry profitably does not hinge on the artificial assumption that firms can augment capacity only at dates 0 and T. Toehold entry simply involves sinking F at time 0 and waiting until T, when the market expands, to begin production. To foil this strategy, the incumbent must install sufficient capital already at time 0, for once entry has occurred at time 0 the incumbent's rational choice for time T is altered—making it suboptimal to choose capital and output high enough to make the entrant unprofitable. Allowing extra moves between 0 and T will not help the incumbent and cannot hurt the entrant.

Second, deterrence can be unprofitable under market expansion even if we modify the game to be still more favorable to deterrence. Suppose that instead of firms choosing capital and output sequentially in each period, capital is chosen sequentially but outputs are determined as follows (e.g., Spence (1977) or Dixit (1979)): if no other firm has chosen positive capital, the incumbent chooses its optimal output for the period (subject to its capacity constraint); if entry has occurred, each firm supplies output "competitively," along its marginal cost curve (up to its installed capacity in our model). Effectively, the incumbent can threaten to expand output up to capacity if entry occurs. In order to deter entry, the incumbent therefore needs only to install the requisite capital without having to supply a correspondingly high output while a monopolist. Nevertheless, under some additional restrictions on costs, the need to install sufficient capital at time 0 in order to prevent toehold entry still makes complete deterrence unprofitable (Malueg and Schwartz, 1989).

---

[5] Actually this explanation is "loose," because firm 1's output can be less than its first-period capital. But the proof shows that even for the cases in which maintaining excess capacity is optimal, the first-period losses from carrying the large capital are prohibitive.

[6] The need to foil *toehold* entry thus explains why deterrence is unprofitable if the future is relatively important. One might have expected the reverse, since deterrence sacrifices present profit for future profit. But toehold entry also entails an investment, and it is the entrant's calculations that are relevant when F is small.

4. Stationary environment

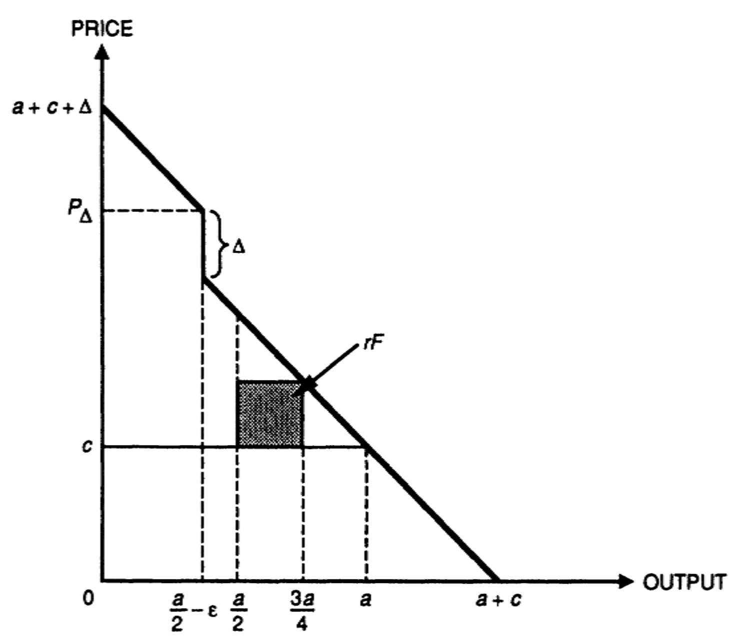
■ In the example and proposition above, some potential entrants are always deterred in equilibrium (recall the many-firms assumption). Why then can a subset of firms profitably deter others while the incumbent alone cannot? What does the incumbent achieve by allowing some entry that it cannot achieve if alone in the market—despite the natural-monopoly cost assumption and the assumption that the incumbent moves first in each period? The answer is that a lower *industry* capital and output suffice to deter further entry in the first period if there are multiple firms in the industry than if the incumbent alone deters all. In the earlier example, a first-period capital of 2.5 does not enable firm 1 to deter firm 2; but with firm 2 also in the market (with zero capital), a first-period capital of 2 by firm 1 suffices to deter firm 3 and all others. The differential ability to deter arises because an industry with several firms sometimes can credibly signal a greater future expansion than could a monopolist.

From the perspective of the initial incumbent, therefore, allowing some early entry enhances the industry's ability to commit to future expansion. As one might expect, such ability is valuable when the market is growing. This raises the question of whether complete deterrence is optimal if instead costs and demand remain stationary. Recalling that deterrence *is* optimal in the one-period model, and that in the two-period model the incumbent moves first at both dates, one might conjecture that deterrence also is optimal in the two-period model if the market is stationary. The following example shows this conjecture to be false. In the example, the incumbent can deter all entry and be profitable (unlike the market-expansion example and proposition, in which deterrence made the incumbent's value negative), but it can do better still by allowing some entry.

Consider the flow demand curve in Figure 2. For simplicity, assume that marginal cost, c, equals zero. (If outputs are determined "competitively," as described in Section 3, then a similar example can be constructed provided $c > 0$.) Letting $rF = a^2/16$, in this example $\bar{q}_0 = \bar{q}_T = a/2$. We show that for judicious choices of the parameters ϵ and Δ, firm 1's value is higher if it produces $q_{10} = (a/2) - 2\epsilon$ at time 0 and induces firm 2 to enter, than if firm 1 deters all entry.

In the second period, industry output will be at least $a/2$, otherwise further entry would

FIGURE 2



occur. So whatever the number of firms entering at time 0, at time T they will behave as though they were facing the demand curve $p = a + c - q$ (the lower segment of the curve in Figure 2). Note that the Stackelberg leader and follower outputs for this demand curve are $a/2$ and $a/4$, respectively.

If firm 1 chooses any first-period output $q_{10} < a/2$, then at date 0 firm 2 can profitably enter with (for example) the following strategy:

$$\text{if} \quad q_{10} \leq a/4, \quad \text{then} \quad q_{20} = k_{20} = a/2;$$
$$\text{if} \quad a/4 < q_{10} < a/2, \quad \text{then} \quad q_{20} = k_{20} = a/4.$$

This strategy by firm 2 deters further entry, since total output is at least $a/2$, and it yields firm 2 a first-period profit exceeding rF and a second-period profit of at least rF. (Recall that $rF = a^2/16$, which equals the profit of the Stackelberg follower in a one-period game. Firm 2 does at least as well as the Stackelberg follower in the second period, given the above strategy.) Thus, to deter all entry firm 1 must choose throughout an output of at least $a/2$, thereby receiving a price substantially less than P_Δ.

Firms 1 and 2 together, however, *can* deter further entry while enjoying the high price P_Δ in the first period. Specifically, the parameters ϵ and Δ can be chosen so that if firm 1 initially produces $q_{10} = (a/2) - 2\epsilon$, then:

(i) firm 2 enters with output $q_{20} = \epsilon$, given firm 3 stays out;
(ii) firm 1 prefers this path over keeping firm 2 out; and
(iii) firm 3 does not enter, given $q_{10} = (a/2) - 2\epsilon$ and $q_{20} = \epsilon$.

It is clear that for any ϵ, conditions (i) and (ii) will be satisfied by choosing Δ sufficiently large. Furthermore, for sufficiently small ϵ, condition (iii) is met. This can be seen as follows (an alternative proof of condition (iii) can be found in Malueg and Schwartz (1989)). Along the proposed path, $q_{10} + q_{20} = a/2 - \epsilon$, so for ϵ small, firm 3's maximum first-period profit is greater than, but arbitrarily close to, rF. For ϵ sufficiently small, firm 3's second-period profit is bounded by a number above but arbitrarily close to $rF/2$. (This is the profit of the second mover in a three-firm Stackelberg game beginning at date T. The argument for this bound is the same one used in the example of Section 3; it is irrelevant that now demand jumps at $a/2 - \epsilon$, since to deter further entry second-period output must exceed this level.) Finally, profit flows of rF and $rF/2$ would yield firm 3 negative value. Hence, given any positive weights on both periods, for ϵ sufficiently small and Δ sufficiently large, all three conditions are satisfied.

While the demand curve in this stationary example is a bit unusual, the underlying force that makes deterrence suboptimal is the same as under market expansion. In the first period, firms 1 and 2 together can deter all further entry while producing a lower total output than firm 1 alone would need to if it wished to deter all others. The differential deterrence ability arises because in the second period, two firms maximizing individual profits will supply a higher total output than would one firm. Therefore, expecting tougher treatment in the second period, an entrant is deterred by a lower first-period output. Given the jump in the demand curve, the ability to reduce first-period industry output—made possible by allowing firm 2's entry—can be sufficiently valuable to firm 1 to outweigh the decrease in its second-period profit. Of course, if firm 1 could commit at date 0 to expand output at date T, it would forestall entry by firm 2. Allowing entry is useful only because the presence of multiple firms enhances the industry's ability to threaten future expansion.

5. Related work and conclusion

■ Our finding that complete deterrence can be suboptimal for the incumbent contrasts with those of, for example, Eaton and Lipsey (1979, 1980), Gilbert and Harris (1984), and Schwartz and Thompson (1986). While the details of these models differ, there is a unifying principle that helps explain why those studies find deterrence to be optimal while other studies (discussed below) do not.

Deterrence will be optimal if *industry* profit with entry is no higher than without it. This condition is sufficient, not necessary (since the incumbent collects the full industry profit if it deters), but it is used in most demonstrations that deterrence is optimal. To show that this condition holds, one typically shows that the incumbent has the option of mimicking the market outcome that would have prevailed with entry (e.g., in terms of price or capacity)—while deterring all entry.

The mimicking option seems innocuous, but it need not be available. For example, if the number of potential entrants is small, it may be feasible to let all enter, and doing so may allow the industry capital level to be smaller than that in a deterrence equilibrium. If the number of potential entrants is large, inevitably some will be deterred. But the investment that an incumbent would need to deter all can exceed what a subset of firms would need to deter the rest. The reason by now is clear: independent firms might use a given capacity more aggressively than would a single firm, or might add capacity more aggressively in the future. Whether the mimicking option exists must therefore be checked in specific cases.

In Schwartz and Thompson (1986) the mimicking option arises because incumbent firms can establish independent competing divisions, a la General Motors, that will behave in all future interactions as new entrants would have. In Eaton and Lipsey (1979, 1980) and Gilbert and Harris (1984) the mimicking option arises endogenously, given the demand and cost conditions assumed.

Eaton and Lipsey (1979) consider a growing spatial market in which *one* additional plant can profitably be added at some future date (plant size is discrete). The incumbent monopolist preempts entry since it could locate the plant where the entrant would have located, but it can do even better by internalizing pricing and location externalities between plants. Eaton and Lipsey (1980) consider a nonspatial market with stationary demand but repeated investment decisions. By spending a fixed cost, a firm obtains a plant capable of producing at constant marginal cost for some fixed duration. Importantly, maximum industry revenue cannot support duopoly. Eaton and Lipsey show that the incumbent deters entry by building a new plant sufficiently before the old one would need to be replaced. The mimicking option exists in these two models because, by assumption, the market cannot support two independently operated plants. Thus, by adding just earlier the plant that the entrant would have added, the incumbent deters entry.

Gilbert and Harris (1984) consider a market with continuously growing demand. There is a fixed cost for installing a plant that is infinitely durable and produces at zero marginal cost a maximum (flow) of one unit of output. Demand is always such that maintaining excess capacity would be suboptimal in a monopoly or Cournot duopoly (e.g., a demand function with elasticity everywhere greater than one). In the subgame perfect equilibrium they examine (propositions 6–8), the incumbent preempts all entry by always adding a plant just before an entrant would. The mimicking option arises because the incumbent and entrant would use the added plant identically (i.e., up to capacity) in any subsequent output interaction. So if allowing one entrant suffices to deter further entry, then duplicating that entrant's investment will suffice to deter it.

In our model, the incumbent generally is unable to deter entry simply by duplicating (slightly earlier) the initial investment decision that an entrant would make. For instance, in the example and proposition in Section 3, which show that deterrence can be unprofitable under market expansion, the entrant sinks the entry cost but chooses zero capital in the first period; so (trivially) duplicating the entrant's capital choice obviously fails to deter entry.

The mimicking option is absent also in Gelman and Salop (1983). There the scale of entry is variable, and the entrant commits to low capacity. To deter such an entrant, the incumbent would had to have set a lower price than the entrant would have charged in equilibrium. This "judo economics" makes deterrence unprofitable. In Judd's (1985) spatial model, a multiproduct incumbent cannot deter a single-product entrant because, faced with entry, the incumbent would withdraw the product closest to the entrant's to avoid negative externalities on its other products. Thus, the incumbent would use an added product "less aggressively" than would an entrant, voiding the mimicking option. If withdrawing a product is prohibitively costly, as implicitly assumed in Schmalensee (1978), the mimicking option is restored and with it preemption.

Finally, the mimicking perspective helps explain a distinction between preemption incentives for a single innovation versus multiple innovations (Gilbert and Newbery, 1982; Dasgupta, 1986). With a single innovation looming, the incumbent outbids an entrant to preserve its monopoly: the incumbent could, as an option, use the innovation as the entrant would have done, thereby denying the innovation to the entrant while mimicking the entry outcome. With multiple innovations, complete deterrence is generally not optimal (Dasgupta, 1986; Gilbert and Newbery, 1982, also hint at this). To see this, consider multiple identical innovations. If the incumbent is expected to purchase them all, the price for each will be the profit an entrant could earn in duopoly. With enough innovations, the incumbent's total outlay will invariably exceed the monopoly profit. Thus, complete deterrence is sub-optimal because the incumbent must pay more for the same innovations than the entrants. The reason, again, is that with multiple firms active in the industry, expected future output will be higher and the equilibrium price of an "entry key" correspondingly lower.

Summarizing, one can classify entry-deterrence models into two groups. In the first the mimicking option is available: if the incumbent makes the same investment decision that an entrant would make (in capacity or innovation), all relevant parties view the market as identical to the previous two-firm market. In the second mimicking is not available: in such models, allowing some entry may be beneficial to the incumbent in future interactions with input suppliers or subsequent potential entrants.

Appendix

■ To prove the proposition some additional notation is helpful:

$$q^*(q') \equiv \underset{q}{\operatorname{argmax}} \{P_T(q' + q) - c_T\} \cdot q;$$

$$q^{**}(q') \equiv \underset{q}{\operatorname{argmax}} \{P_T(q' + q + q^*(q' + q)) - c_T\} \cdot q;$$

$$\tilde{q} \text{ satisfies } rF = \{P_T(\tilde{q} + q^{**}(\tilde{q}) + q^*(\tilde{q} + q^{**}(\tilde{q}))) - c_T\} \cdot q^{**}(\tilde{q}); \quad \text{and}$$

$$\Pi_T \equiv \{P_T(\tilde{q} + q^*(\tilde{q})) - c_T\} \cdot q^*(\tilde{q}).$$

Thus, $q^*(q')$ denotes a firm's best output at time T if it chooses to enter, given that prior movers have chosen total output q' and that the firm expects no further entry, while $q^{**}(q')$ is the firm's best output at T if it expects one more firm to choose output after it. The output $\tilde{q}$ is such that in a three-firm single-period game starting at T, if the first mover chooses output $\tilde{q}$, the profit to the second mover is rF. The significance of $\tilde{q}$ is that if at date 0 firm 1 produces $\tilde{q}$ and only firm N_F enters, with toehold entry, then further entry at T is just unprofitable. Finally, Π_T is the gross profit stream to firm N_F at T if it enters with toehold entry and if prior movers have chosen at T total output $\tilde{q}$.

The proof of the proposition involves showing that to deter all entry, firm 1 must choose at date 0 capital exceeding $\tilde{q}$, and that such a capital choice makes firm 1's value negative. The proof uses Lemmas 1 and 2.

Lemma 1. $\lim_{F \downarrow 0} \tilde{q} = q_{cT}$.

Lemma 2. $\liminf_{F \downarrow 0} \Pi_T / rF \geq 2$.

Lemma 1 is intuitive: if $\tilde{q}$ were bounded away from q_{cT}, then the second mover in a three-firm game starting at T would earn (gross) profit exceeding rF, for F sufficiently small. The proof of Lemma 2, which relies on our assumption that P_T has continuous derivative at q_{cT}, uses a linear approximation of P_T near q_{cT} (see Malueg and Schwartz, 1989).

Proof of the proposition. The proof proceeds in three steps.

Step 1. For all sufficiently small F, to deter all entry, firm 1 must choose $k_{10} > \tilde{q}$.

Suppose that at date 0 firm 1 chooses $k_{10} \leq \tilde{q}$ and that none of the firms 2, 3, ..., $N_F - 1$ enters. We show that firm N_F can profitably enter using the strategy of toehold entry: at time 0, sink F and choose zero capital and output; at time T, choose capital and output optimally given the history of the game.

Under these strategies, at date T firm 1 will choose $q_{1T} = k_{1T} = \tilde{q}$. By definition of $\tilde{q}$, this choice by firm 1, together with firm N_F's presence in the market, just deters all further entry. Any output below $\tilde{q}$ would allow further entry, which cannot be optimal for firm 1 at time T. Any higher output is less profitable for firm 1, since for sufficiently small F, $\tilde{q}$ already exceeds the time-T Stackelberg leader output. Firm N_F's optimal output and capital levels at date T are both simply $q^*(\tilde{q})$, since it is the last mover.

Let V^* denote the value of firm N_F's profit under the above path. Then

$$\begin{aligned} V^* &= -F + w_T\Pi_T \\ &= -w_0 rF + w_T\{\Pi_T - rF\} \\ &= rF\{-w_0 + w_T(\Pi_T/rF) - w_T\}. \end{aligned}$$

The bracketed term following the last equality has a limit infimum that is at least as large as $w_T - w_0$, which is strictly positive (Lemma 2 and condition (a) in the proposition). Therefore, for all (positive) F sufficiently small, if at time 0 firm 1 chooses $k_{10} \leq \tilde{q}$ and if nobody else enters, then firm N_F can profitably enter.

Step 2. For all sufficiently small F, firm 1's value if it deters all entry can be bounded as follows:

$$V_1 \leq \begin{cases} w_0 R_0 - w_0 c_0 \tilde{q} + w_T\{P_T(\bar{q}_T) - c_T\}\bar{q}_T & \text{if} \quad q_{cT} > q_{R0} \quad \text{and} \quad q_{cT} \leq q_{RT} \\ w_0 P_0(\tilde{q}) q_{cT} - w_0 c_0 \tilde{q} + w_T\{P_T(\bar{q}_T) - c_T\}\bar{q}_T & \text{if} \quad q_{cT} \leq q_{R0} \quad \text{or} \quad q_{cT} > q_{RT}, \end{cases}$$

where q_{R0} and q_{RT} denote the outputs that maximize industry revenue, *given demands at time 0 and T, respectively, and R_0 denotes maximum revenue for time-0 demand.*

The two cases differ only regarding the profitability of maintaining excess capacity in the first period, which in turn affects the bounds on first-period revenue, R_0 versus $P_0(\tilde{q})q_{cT}$. The other, common, terms are understood as follows. To deter all entry, firm 1's capital at date 0 must be at least $\tilde{q}$ (Step 1 above), and at date T its capital and output must be least $\bar{q}_T$. Since $\bar{q}_T$ exceeds the second-period monopoly output (for all sufficiently small F), and since $qP_T(q)$ is strictly quasi-concave in the relevant range, if firm 1 deters all entry it earns a second-period revenue stream whose time-0 value is at most $w_T\bar{q}_T P_T(\bar{q}_T)$. Given the assumption $c_T \leq c_0$, if firm 1 deters, its total cost is at least $w_0 c_0 \tilde{q} + w_T c_T \bar{q}_T$. (Observe that this cost bound is valid whether cost falls due to a decrease in the price of capital or to technical change that increases the output/capital ratio. The best case for the incumbent is the one of technical progress that is applicable also to capital purchased at time 0, but even in that case the given bound applies.)

Now consider first-period revenue. The bound used in the first case is maximum industry revenue, R_0. If either $q_{cT} \leq q_{R0}$ or $q_{cT} > q_{RT}$, then firm 1's first-period revenue, given entry is deterred, can be bounded more stringently. Along any optimal path for firm 1, $q_{10} \leq q_{cT}$ (since $q_{10} = q_{cT}$ suffices to deter). To justify the revenue bound $P_0(\tilde{q})q_{cT}$ it only remains to show that first-period price will not exceed $P_0(\tilde{q})$, i.e., that $q_{10} \geq \tilde{q}$. We know from Step 1 that deterring entry requires $k_{10} \geq \tilde{q}$. If $q_{cT} \leq q_{R0}$, it is profit maximizing to use fully such capital once installed (recall that q_{R0} maximizes first-period revenue and that the latter is strictly quasi-concave). If $q_{cT} > q_{RT}$, then in order to deter toehold entry firm 1's first-period *output* (not just capital) must satisfy $q_{10} \geq \tilde{q}$. (Given $q_{cT} > q_{RT}$, for sufficiently small F the time-T Stackelberg leader output is below $\tilde{q}$, since the former is below q_{cT} and is independent of F, while the latter goes to q_{cT} as F goes to 0. In this case, firm 1 would not expand output at time T to $\tilde{q}$ once toehold entry by firm N_F has occurred, and toehold entry by firm N_F indeed would be profitable if nobody else had entered, given $q_{10} < \tilde{q}$.) Hence, if either $q_{cT} \leq q_{R0}$ or $q_{cT} > q_{RT}$, then firm 1's optimal first-period output along any deterring path satisfies $q_{10} \geq \tilde{q}$.

Step 3. $\limsup_{F \downarrow 0} V_1 < 0$, which establishes the proposition.

From Step 2 it suffices to consider two cases. First, suppose $q_{cT} \leq q_{R0}$ or $q_{RT} < q_{cT}$. Then,

$$\limsup_{F \downarrow 0} V_1 \leq w_0\{P_0(q_{cT}) - c_0\} q_{cT} < 0,$$

where the first inequality follows from Lemma 1 and the fact that $\lim_{F \downarrow 0} \{P_T(\bar{q}_T) - c_T\}\bar{q}_T = 0$; and the second inequality follows from the assumptions that $P_0'(q_{c0}) < 0$, $c_0 > 0$, and $q_{c0} < q_{cT}$, which imply that $P_0(q_{cT}) < c_0$.

Next, suppose $q_{R0} < q_{cT}$ and $q_{cT} \leq q_{RT}$. Then

$$\begin{aligned} R_0 &\equiv P_0(q_{R0})q_{R0} \\ &\leq \max P_T(q)q \qquad \text{such that} \qquad q \leq q_{R0} \qquad (\text{since } P_0(q) \leq P_T(q)) \end{aligned}$$

$< \max P_T(q)q$ such that $q \leq q_{cT}$ (since $q_{R0} < q_{cT} \leq q_{RT}$ and $qP_T(q)$ is strictly quasi-concave)

$= P_T(q_{cT})q_{cT}$ (since $q_{cT} \leq q_{RT}$ and $qP_T(q)$ is strictly quasi-concave)

$= c_T q_{cT}$ (by definition of q_{cT})

$\leq c_0 q_{cT}$ (by the assumption that $c_T \leq c_0$).

Hence, by Step 2,

$$\limsup_{F \downarrow 0} V_1 \leq w_0\{R_0 - c_0 q_{cT}\} < 0. \quad Q.E.D.$$

References

DASGUPTA, P. "The Theory of Technological Competition." In J. Stiglitz and G.F. Mathewson, eds. *Recent Developments in the Analysis of Market Structure.* Cambridge, Mass.: MIT Press, 1986.

DIXIT, A. "A Model of Duopoly Suggesting a Theory of Entry Barriers." *Bell Journal of Economics,* Vol. 10 (1979), pp. 20–32.

———. "The Role of Investment in Entry Deterrence." *Economic Journal,* Vol. 90 (1980), pp. 95–106.

EATON, B.C. AND LIPSEY, R.G. "The Theory of Market Preemption: The Persistence of Excess Capacity and Monopoly in Growing Spatial Markets." *Economica,* Vol. 46 (1979), pp. 149–158.

——— AND ———. "Exit Barriers Are Entry Barriers: The Durability of Capital as a Barrier to Entry." *Bell Journal of Economics,* Vol. 11 (1980), pp. 721–729.

——— AND ———. "Capital, Commitment and Entry Equilibrium." *Bell Journal of Economics,* Vol. 12 (1981), pp. 593–604.

FUDENBERG, D. AND TIROLE, J. "Capital as a Commitment: Strategic Investment to Deter Mobility." *Journal of Economic Theory,* Vol. 31 (1983), pp. 227–250.

GELMAN, J.R. AND SALOP, S.C. "Judo Economics: Capacity Limitation and Coupon Competition." *Bell Journal of Economics,* Vol. 14 (1983), pp. 315–325.

GILBERT, R.J. "Preemptive Competition." In J. Stiglitz and G.F. Mathewson, eds. *Recent Developments in the Analysis of Market Structure.* Cambridge, Mass.: MIT Press, 1986.

——— AND HARRIS, R.G. "Competition with Lumpy Investment." *RAND Journal of Economics,* Vol. 15 (1984), pp. 197–212.

——— AND NEWBERY, D.M.G. "Preemptive Patenting and the Persistence of Monopoly." *American Economic Review,* Vol. 72 (1982), pp. 514–526.

JUDD, K.L. "Credible Spatial Preemption." *RAND Journal of Economics,* Vol. 16 (1985), pp. 153–166.

MALUEG, D.A. AND SCHWARTZ, M. "Preemptive Investment, Toehold Entry, and the Mimicking Principle." Economic Analysis Group Discussion Paper EAG 89-8, Antitrust Division, U.S. Department of Justice, 1989.

MODIGLIANI, F. "New Developments on the Oligopoly Front." *Journal of Political Economy,* Vol. 66 (1958), pp. 215–232.

OMORI, T. AND YARROW, G. "Product Diversification, Entry Prevention and Limit Pricing." *Bell Journal of Economics,* Vol. 13 (1982), pp. 242–248.

SALONER, G. "The Role of Obsolescence and Inventory Costs in Providing Commitment." *International Journal of Industrial Organization,* Vol. 4 (1986), pp. 333–345.

SCHMALENSEE, R. "Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry." *Bell Journal of Economics,* Vol. 9 (1978), pp. 378–393.

SCHWARTZ, M. AND THOMPSON, E.A. "Divisionalization and Entry Deterrence." *Quarterly Journal of Economics,* Vol. 101 (1986), pp. 307–321.

SMILEY, A. "Direct Competition Among Cable Television Systems." Economic Analysis Group Discussion Paper EAG 86-9, Antitrust Division, U.S. Department of Justice, June 1986.

SPENCE, A.M. "Entry, Capacity, Investment and Oligopolistic Pricing." *Bell Journal of Economics,* Vol. 8 (1977), pp. 534–544.

———. "Investment Strategy and Growth in a New Market." *Bell Journal of Economics,* Vol. 10 (1979), pp. 1–19.

VIVES, X. "Sequential Entry, Industry Structure and Welfare." *European Economic Review,* Vol. 32 (1988), pp. 1671–1687.