Nº 210
Febrero 2002

Documento de Trabajo

ISSN (edición impresa) **0716-7334**
ISSN (edición electrónica) **0717-7593**

Pricing and marketing rules with brand loyalty.

Salvador Valdés

www.economia.puc.cl

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
Version of February 11, 2002 **Comments welcome**

Pricing and marketing rules with brand loyalty*

Salvador Valdés-Prieto**

Abstract

Many firms face a dynamic trade-off: if price is reduced, the firm attracts new customers who will yield profits in the future, but it also forgoes the opportunity to squeeze profits now from loyal customers. This paper identifies a rule that represents the optimal resolution of this trade-off, in terms of an intuitive modification to the static Lerner rule. We find that the "effective" price elasticity depends on the discount rate used by the firm, on the rate of depreciation of the clientele through exit from the stock of repeat purchasers, and on a weighted sum of the price elasticities of the flow of entries into the stock of repeat-purchasers and the flow of exits from the stock of repeat-purchasers. None of these factors enter the optimal pricing strategy for a firm facing a conventional demand function with instantaneous adjustment, i.e. where consumers are "fast switchers", rather than repeat-purchasers. We also find optimal rules for marketing investment and for quality of service, which are extensions of the Dorfman-Steiner conditions. The paper shows that our rules, with suitable modifications, are valid for many market structures, including monopolistic competition, pure monopoly and strong cartels, dominant firms and oligopolists that have full commitment ability. In the case of dominant firms and oligopolists that cannot commit to their strategy paths, these simple optimal pricing and marketing rules do not apply.

JEL classification numbers: D4, L1, M3

* This work received support from the Research Fund of the School of the Economics and Business Administration of the Catholic University of Chile. I am grateful for the comments of Birger Wernerfelt, Felipe Zurita, Jorge Tarziján, Edgardo Barandarián, Ronald Fischer, José Miguel Sánchez, to participants at the seminar at Termas del Corazón, January 22-24, 2001, at the Latin American Economic Association meetings in Montevideo, Uruguay, October 18-20, 2001, and at the seminar at Termas del Corazón, January 18-19, 2002. Of course, the only one responsible for remaining errors and ommissions is the author.
** Professor at the Instituto de Economía, Pontificia Universidad Católica de Chile, Santiago, Chile. svaldes@faceapuc.cl

The cornerstone of our knowledge remains missing: how (should) a firm exploit brand loyalty once it has it? (Wernerfelter,1991)

1. Introduction

Consumers in many markets do not respond to abrupt price changes with an equally abrupt change in the flow of orders of the same size that will prevail in the long run. We describe this fact as "brand loyalty". An example is given by the business model where the firm builds a clientele with contracts for continuous service into the indefinite future. Clients sign contracts in which they commit to pay regularly for the service received, while keeping the right to switch suppliers, possibly giving some notice. This business model dominates industries such as telephone subscriptions, magazine and newspaper subscriptions, most of the consumer end of the financial services industry (mutual funds, life insurance, credit cards, pension fund management), health insurance and health maintenance, educational services, cable TV, auditing services, legal services, tax accounting, security services and many more.

The brand loyalty concept applies also to firms that do not provide continuous service. Clients that sign a service contract may be seen as repeat-purchasers that have formalized the relationship with their supplier. Informal relationships that develop through repeat purchases can be as strong as service contracts. Examples are frequent flyer programs by airlines, addictive consumer goods and firms that sell capital goods with a relatively low margin while selling supplies, spare parts, and maintenance at a relatively large margin.

We provide now a brief summary of the competing explanations for brand loyalty. One explanation is the presence of "consumer switching costs" - those sunk costs incurred by clients when changing supplier[1]. Indeed, it is rational for a consumer confronted with an abrupt price change that is smaller than the switching cost, not to respond with an abrupt change in the flow of orders. In fact, when the price change is smaller it is best not to respond at all. For example, the gain from getting cell-phone service from the best rather than the current provider for a single month may be swamped by the sum of the search cost of identifying the best provider for the month plus the sunk costs of disconnection and reconnection.

However, a forward-looking consumer would take into account expected future relative prices before sinking a switching cost. Thus, the mathematical complexity of consumer response in the presence of switching costs can be high. To keep the consumer problem simple, most of the consumer switching cost literature has been developed in a two-period setting (Padilla, 1995)[2]. A smaller literature has explored consumer switching costs in multi-period settings, starting with Farrell and Shapiro (1988). However, to achieve tractability these models keep each consumer's horizon limited to two periods (as in Padilla, 1995) or assume that a consumer that stops buying for a single period must exit the market (Beggs and Klemperer, 1992)[3]. One problem is that two-period models fail to capture the essence of brand loyalty: in two-period contexts the set of consumers that face a

[1] The seminal work is in Klemperer, 1987; see Klemperer 1995 for an excellent survey.

[2] See the survey by Klemperer 1995, and the papers by Stahl 1996, Chen 1997, and Bernstein and Micco 2001.

[3] In addition, multiperiod models limit the shape of individual consumer demand to a reservation price for single unit of the good, which is completely inelastic.

price differential larger than their switching cost respond *as fast* as in traditional models of demand, while the rest does not respond at all. In two-period settings, switching costs reduce the size of consumer response only, not the speed of response, and thus fail to generate brand loyalty as usually understood.[4]

A second explanation of brand loyalty is "consumer inertia". In this explanation, consumers keep on buying the current brand passively without searching until a stimulus arrives that induces a switch (Wernerfelter, 1991). The stimulus may be information on price differentials received by word of mouth as a by-product of other activities, or the pressure exerted by a salesperson. The difference with the switching cost approach is that inert consumers switch even in the absence of price differentials. A rational customer may switch even if her price *increases* by a limited amount in order to respond to various stimuli, including among the latter getting rid of an oppressive salesperson.

A third approach to brand loyalty would be to define it as a result of slow information revelation, which may be in part be due to the time it takes a consumer to search for alternatives and compare prices and quality. The search literature models that keep track of the time that it takes consumers to reach decisions, such as XXX, could provide a solid foundation for the behavior we label as brand loyalty.

As the controversy on how to model brand loyalty continues, Wernerfelter's (1991) question remains unanswered: Given a known degree of brand loyalty, how much should a firm reduce price to attract new customers, knowing that it is giving up the opportunity to squeeze profits from loyal customers? Up to now, the only answer provided by the literature has necessitated the "sticky-price" assumption. In the tradition of Cournot oligopoly, this literature assumes that within every instant firms announce first the quantities they will auction and then consumers bid and set the price. A sort of "loyalty" is added by postulating that the price bid by consumers in each instant adjusts slowly (is "sticky") to the "long run" willingness to pay[5]. However, these assumptions seem unrealistic for many actual markets where pricing decisions are made by firms and not by consumers, while consumers choose quantities in response to those prices. In other words, the applicability of the sticky price approach seems limited to only a few cases[6].

We propose a "new" approach: to represent consumer behavior through general response functions. This is the classical approach followed by Abba Lerner (1934) to analyze a firm's pricing strategy, which was to postulate a general demand function for the consumer segment under analysis, without delving into the details of consumer demand. Nerlove and Arrow (1962) contributed a dynamic model of marketing "goodwill" in the same spirit of using general response functions. The contribution of this paper is to analyze the pricing decision when there is brand loyalty. To our knowledge, there is only one paper in the literature that attacked this problem with a comparable model (Wernerfelt, 1991), but it was devoted to analyze the existence of a steady-state industry equilibrium, rather than the first-order conditions for the firm's problem and their interpretation, as we do here.

---

[4] Consumer switching costs may still generate brand loyalty in more general settings.

[5] See Evans (1924) for the monopoly version, and Simaan and Takayama (1978) for duopoly. Fershtman and Kamien (1987) and others summarized by Dockner et al (2000, section 10.1) show further developments.

[6] Following Friedman (1983, section 2.2), the sticky-price model does appear applicable to a fishing duopoly where two firms send fresh fish to an auction market in continuous time, while partially inert consumers take their time to reevaluate their true willingness to pay for a given amount of fish through introspection, and meanwhile go on paying the previous price.

The limitation of this paper's approach is that the postulated consumer response functions may not be general enough to represent consumers' response to dynamic pricing strategies, chiefly because future expected prices are absent from the response function. However, this assumption would be valid for forward-looking consumers when the current price is a good predictor of the future level of the price charged by the firm. The alternative assumption that consumers have two-period lives may be even more restrictive.

The gains from our modest approach appear substantial in terms of clarifying the optimal strategy for the firm. In the unique optimal path, the pricing rule can be presented as a modified Lerner Rule[7]: the effective price elasticity depends on the discount rate, on the rate of depreciation of the clientele through exit, and on a weighted sum of the price elasticities of entry and exit flows. As the model allows also separate marketing investments both to attract new clients and to improve the quality of service, modified Dorfman-Steiner conditions obtain: optimal marketing investment depends on the previously mentioned factors as well. These rules allow an intuitive representation of the dynamic trade-off faced by firms when it comes to exploit the loyalty to their brand.

Section 2 proposes a simple model for decisions of a firm facing a continuous service demand or repeat-purchasers. Section 3 discusses the first order conditions for the steady state strategies for a single firm within a monopolistically competitive market. Section 4 extends the model to pure monopoly, strong cartels and dominant firms. Section 5 explores the pricing and marketing rules for dominant firms and for oligopolists, under the polar cases of full and no commitment. Section 6 concludes.

2. A model of brand loyalty as seen by the firm

The conventional economic model of a firm that sells a differentiated product to a market segment where it charges a uniform price and where it *does not* enjoy brand loyalty, takes the form of program (A). This is a special case where marketing is kept to a simple form[8]. Note that consumers respond to abrupt price changes with an equally abrupt change in the flow of orders, and that this change has the same size that prevails in the long run. Therefore, demand in program (A) represents the case where consumers are "fast switchers", i.e. consumers respond swiftly to any change in price differences.

$$\text{(A)} \quad \underset{\{p(t);M(t);q(t)\}}{Max} \left\{ V = \int_{t=0}^{\infty} \left[p \cdot D - F - c_a \cdot D - c_m \cdot M - c_q \cdot q \cdot D \right] e^{-rt} dt \right\}$$

$$\textit{subject to}: \qquad D = D(p, p_{-i}, M, \Sigma M_{-i}, q, q_{-i})$$

where V = value to the firm of the segment of consumers under analysis.

---

[7] We also specialize to the case where the firm has chosen to use or is compelled to use uniform prices and uniform marketing strategies for each segment of the clientele, and the sizes of segments are not influenced by consumer decisions. Thus, the case of second-degree price discrimination is left for future research.

[8] With several firms present in the market, this type of demand is not simple, as shown by Schmalensee (1976). Starting with Nerlove and Arrow (1962), the literature has developed the case where demand depends of a stock S of goodwill, rather than of the flow of new messages M as in this paper. In that case, demand is modeled as D = D(p,S), where goodwill S evolves according to $dS/dt = M - \delta \cdot S$, and δ = depreciation rate of the goodwill stock, in the mind of consumers. We choose the simpler model because the focus of our paper is on pricing decisions.

p = the uniform price charged at time t to this segment of fast switchers ($/unit).

p_{-i} = prices charges by rivals at time t.

M = number of marketing messages issued in t to attract new sales, at a marginal cost of c_m each. These pre-sale messages do not raise production cost $F + (c_a + c_q \cdot q) \cdot D$.

SM_{-i} = sum of marketing messages issued by rivals to attract sales at time t.

q = quality of service provided to those that purchase at time t, which is uniform in the segment. The marginal cost of raising quality by one unit is denoted by c_q.

q_{-i} = quality of service offered by rivals at time t.

In program (A), the firm maximizes the present value V of its profits from the segment, choosing time paths for a uniform price p, for the number of marketing messages M and for quality of service q. Expenditures include a fixed cost F that accounts for economies of scale both in the back office and in marketing (say, in sales-force training), a variable production cost $c_a \cdot D(p,M)$ and the marketing investment $c_m \cdot M$. The variable production cost has been simplified by assuming constant returns to scale[9]. Although all units are sold at the uniform price p, price discrimination between new and old consumers in this segment is allowed implicitly, since marketing messages may consist in coupons or gifts to new consumers.

It is well known[10] that the optimal pricing strategy for this type of demand fulfills in every instant the Lerner Rule $(p^*-c)/p^* = 1/\eta_{D,p}$ and that the optimal marketing investment strategy meets in every instant the Dorfman-Steiner conditions $(p-c) \cdot \partial D / \partial M = c_m$ and $(p-c) \cdot \partial D / \partial q = c_q \cdot q$. (where $c = (c_a + c_q \cdot q)$ is marginal production cost). These rules show that when there is no brand loyalty, pricing and marketing decisions do not depend on the discount rate r, or on the rate of depreciation of the clientele through contract severance or loss of loyalty. We show below that this result is not general, as it is due entirely to the assumed swift response of consumers: when price is reduced, volume increases immediately and thus simultaneously. As the volume increase is contemporaneous with the loss of margin, discount rates and other dynamic considerations are excluded by this polar case.

The model we postulate has the same cost structure as program (A), but the demand is of a "stock" nature due to brand loyalty. We specialize by assuming that the firm is free to change its price p over time, provided the price is uniform for all clients in the segment[11]. To facilitate comparison with program (A), we also keep marketing to its simplest form[12]. The model is:

---

[9] This simplification may be unrealistic in applications, but our model can be easily expanded to consider more general cost functions. We make this assumption because the focus of the paper is on pricing.

[10] See Berndt (1991) for a summary.

[11] Price uniformity is standard in the mutual fund management market, for example. In the U.S., the Investment Company Act of 1940 requires mutual funds to apply price uniformity regarding both expenses and management fees. More recently this requirement was relaxed somewhat by allowing each fund to issue several series of shares, whose commissions and expenses might differ. Our model would represent the pricing and marketing decisions for an individual series of shares.

[12] That is, we do not present the extension where demand is modeled as $E = E(p,S,q)$ and $\phi = \phi(p,S,q)$, where S evolves according to $dS/dt = M - \delta \cdot S$, and δ = depreciation rate of the marketing stock.

$$(B) \quad \underset{\{p(t);\, M(t);\, q(t)\}}{Max} \left\{ V = \int_{t=0}^{\infty} \left[p \cdot A - F - c_a \cdot A - c_m \cdot M - c_q \cdot q \cdot A \right] \cdot e^{-rt} dt \right\}$$

$$s.\ to: \quad 1) \quad \frac{dA}{dt} = E(p, M, q, p_{-i}, q_{-i}, \Sigma M_{-i}, A_{-i}) - A \cdot f(p, q, p_{-i}, \Sigma M_{-i}, q_{-i})$$

$$2) \quad A(t) \geq 0$$

$$3) \quad M(t) \geq 0 \qquad and \quad 4) \quad A(0) = A_o$$

where A_t = stock of clients at this firm, i.e. members of the repeat-purchase clientele.
E_t = positive transfers (number of new members) into the clientele during instant t.
ϕ_t = exit *rate* from the clientele of repeat purchasers, as of date t.
$S_t = \phi(p,q) \cdot A$ = negative transfers (number of exits from the clientele) in instant t.
A_{-i} = stock of the repeat-purchase clientele at rival firms (a vector), as of date t.

Brand loyalty is modeled by the first ("dynamic") restriction in program (B), which indicates that the stock of repeat-purchasers - or clients with a service contract outstanding - evolves slowly in response to entries E and exits S. We do not provide micro foundations for the functions E and S, but rather use general functions which may be estimated empirically. Functions E and S can have values as large as required, so the full range of consumer response speeds can be represented.

It is natural to assume that entries (E) depend negatively on the current uniform price chosen by the firm for the segment ($\partial E/\partial p < 0$). This assumption would also be valid for forward-looking consumers whenever the current price is thought to be a good predictor of the future level of the uniform price charged by the firm to the segment. Entries (E) also depend positively of marketing expenses in messages ($\partial E/\partial M > 0$)[13]. Potential repeat-purchasers that may come to this firm are assumed not to be sensitive to past marketing expenditures through a stock of "goodwill". The essential distinction between program (B) and the "goodwill" extension of program (A) is that in the former the net flow of new clients does not depend on past marketing expenditures[14], while in the latter the net flow of new clients (dD/dt) depends positively on past marketing expenditures. However, in both cases current revenue depends on past marketing expenditures: in the "goodwill" extension of program (A), those expenditures created goodwill, which explains current demand and revenue. In program (B) those expenditures attracted clients in the past, formed the current

[13] In practical cases, identifying messages that attract new clients may require some effort. For example, if a mutual fund manager pays a salesperson a commission in proportion to increments in funds attracted from his or her client list, then the salesperson may see an incentive to churn the funds of its existing client list, as this increases the increments in funds - at the expense of increasing decrements as well - with little attraction of new clients. But when a mutual fund manager pays commission on the basis of the number of new clients and their funds, this investment may be represented safely by M.

[14] In oligopolies, the net flow of new clients may depend negatively on past marketing expenditures: these expenditures reduced the clientele of rivals, so quits from rivals' clienteles are smaller. Thus, this firm's current marketing expenditures are less effective and fewer entries result for given M. In addition past marketing expenditures may have a further strategic effect: when this firm's clientele is larger and its total exits are larger, the productivity of rivals' marketing messages increases, inducing rivals to spend more in marketing, further increasing exits from this firm.

repeat-purchase clientele (denoted by A) and thus contribute to current revenue. For this reason, the firm in program (B) must also regard the current flow of messages as an investment, because the additional future revenue generated by an increment in the clientele justifies the current expense.

We also assume that entries (E) depend positively of the firm's current quality of service ($\partial E/\partial q > 0$)[15]. Forward-looking consumers may use the current level of quality (q_t) to project the level that service quality will attain in the future. The firm also views expenditure in service quality (or in "maintenance" marketing) as an investment, because it attracts entries, raises the clientele and generates a flow of revenue in the future.

Entries E depend negatively on the level of rivals' marketing expenditures (SM_{-i}), because a higher level makes it less likely for the firm to attract those customers that are exiting any given rival's clientele. Competition with other rivals' marketing efforts "dilute" this firm's efforts when the former are higher. The same happens with potential clients that currently do not have contracts with any firm, who become harder to attract to this firm.

Last but not least, entries E depend positively on the size of the rivals' clientele (A_{-i}). This is because given an overall exit rate from the rivals' clientele, the number of those exits increases in proportion to the size of that clientele, and the number of entries into this firm should increase as the number of exits from rivals rise, for a given marketing and pricing effort. Note that entries E *do not depend* directly on the stock of clients in this firm (A). This does not prevent a firm from choosing to increase its marketing investment M in proportion to its stock of clients. However, that would be a particular strategy -whose optimality is questioned below- and is not a datum for the firm.

Exits S from the firm are proportional to the stock of members served by the firm (A). This is a natural assumption because when some factor makes it less convenient for the consumer to remain with this firm, this factor acts on the whole stock of repeat-purchasers, and the number of exits should increase in proportion to the size of that stock. This is true both for the arrival of stimuli that induce switches when there is "consumer inertia" and for economic factors that affect the exit rate ϕ such as the level of the firm's price ($\partial\phi/\partial p > 0$) and the quality of service ($\partial\phi/\partial q < 0$).

An advantage of our modeling approach is that brand loyalty can be measured using econometric methods. When consumers are "fast switchers", as in program (A), it is natural to estimate demand using the functional form of D in that program[16]. A natural extension followed in the literature has been to estimate the same demand function plus a partial adjustment mechanism. Partial adjustment proposes that demand response - in quantities - will occur slowly over time, and that if current conditions remain constant, a new steady state will be eventually achieved by the flow of customer orders[17]. With partial adjustment, the consumer is assumed to adjust her stock demand for a given period by only θ % of the

---

[15] In program (B), quality of service can take a different meaning. When post-sale marketing messages are distributed uniformly to all members to "maintain" current members, then the number of marketing messages issued in t for this purpose, and expressed as a rate per member, can be modeled as part of quality of service. In financial firms, a critical dimension of service quality is financial performance, i.e. the risk-adjusted return earned by the funds owned by clients and managed by the firm.

[16] In applications, D is augmented to allow for the influence of a stock of "goodwill". Recent empirical studies that have used program (A) with this extension are Roberts and Samuelson (1988) and Rizzo (1999).

[17] This approach is the conceptual equivalent to the "sticky price" approach, the difference being that here quantities adjust slowly, rather than prices.

difference between past stock demand and current "ideal" or "long run" stock demand[18]. One limitation of partial adjustment is that only "ideal" demand depends of pricing and marketing strategies, while the coefficient of adjustment is assumed to be exogenous and fixed, regardless of the firm's pricing and marketing strategies[19].

The econometrics suggested by the "dynamic restriction" in program (B) are different: estimate two "flow" functions, one for entries, i.e. new additions to the stock of repeat-purchasers, and another for the exits from the stock of repeat-purchasers of the firm. The data required to estimate these functions is usually available in firms that sign continuous service contracts, but is harder to obtain for firms that cater to repeat-purchasers through informal relationships. In the informal case, the date may be obtained from a panel of surveys.

An advantage of this econometric modelling approach is that it acknowledges that the firm can choose its exit and entry rates over time, and thus can choose the speed of adjustment to the steady state. As the temptation to squeeze more profits out of current customers is an important determinant of the steady state, and the cost of yielding to that temptation is to increase the speed of consumer response, methods that recognize the endogeneity of the speed of adjustment should be more useful to firms.

Monopolistic Competition

We use program (B) to describe the decisions of a monopolistically competitive firm. A monopolistically competitive firm is defined as one whose consumer response functions are not influenced by the actions (prices, marketing, quality of service) or states (clientele) of any individual rival acting *alone*. This requires that no rival is big enough to influence industry averages by acting alone[20]. Monopolistic competition still allows for a significant influence of the actions of groups of competitors acting together. If the average actions of groups of competitors acting together change over time, the consumer response functions for entries and exits should include that information. The analysis that follows focuses on the special but interesting case where the aggregate of a firm's rivals have achieved a steady state, in the sense that its actions remain constant over time. In this setting, the arguments of entries and exits functions given by rivals' actions and states are set to constants and can be ignored. This simplifies program (B) considerably.

Another standard component of monopolistic competition is free entry. Free entry changes the level of profit that can be obtained with optimal actions, but does not change the need to optimize, which is our focus. Thus, we do not analize its implications.

Let us consider the existence of a solution to program (B). The existence of a solution to optimal control problems is usually handled by assuming that the functions E, ϕ and the profit function (the integrand in (B)) are sufficiently smooth, satisfy certain boundedness conditions and that the integral converges. Dockner et al (2000, p. 40) point

---

[18] For example, $D^{shortrun} = A(t) = A(t-1) + \theta[D^{longrun} - A(t)]$, with θ = adjustment coefficient in [0,1]. θ is an exogenous parameter.

[19] Another limitation arises when the price of a firm's product follows (say) a rising path relative to competitor's prices: with partial adjustment, this leads actual demand to be always behind ideal demand.

[20] Monopolistic competition is compatible also with a perceptible influence on the consumer response functions faced by an individual firm of sudden entry or exit of a single competitor at a non-negligible scale. However, we do not analyze the impact of entry and exit of rivals in this paper.

out that in an oligopoly game a smoothness or boundedness assumption may implicitly restrict the strategies of rivals, because these functions depend on the strategies chosen by rivals, possibly affecting the outcome of the game. This is not the case in monopolistic competition, however, because in this setting any strategy change adopted by a single competitor has a negligible impact on the consumer response functions faced by the firm. Thus, the consumer response functions can be required to meet boundedness and smoothness assumptions in the market structure analyzed in this section.

According to Dockner et al (2000, see Theorem 3.2 in p. 49), a solution exists under the following conditions: (i) the state space is a convex set. In program (B), the state space is the set containing all possible values of the clientele A, which is the set $[0, A^U]$, which is a convex set. However, we assumed monopolistic competition and ignored the restriction $A < A^U$, effectively expanding the state space to [0,8). As this is also a convex set, this condition is satisfied; (ii) the optimized Hamiltonian function is concave with respect to the state variable. Equation (A.1) in Appendix 1 defines the Hamiltonian for program (B). It is linear in the state variable, which is the clientele A, and therefore this condition is met; (iii) the Hamiltonian function is continuously differentiable with respect to the state variable A. A Hamiltonian that is linear in the state variable meets this condition also; and (iv) the integral that defines the objective functional converges as the horizon grows to infinity. In the long run, the integrand grows in proportion to the product $e^{-rt}\cdot\lambda(t)\cdot A(t)$ (see Dockner et al, 2000, Lemma 3.1), where λ is the value of each client (see section 3 and Appendix 1). The term e^{-rt} certainly helps to achieve convergence if $r > 0$, and it is appropriate to recall that the discount rate in an economy must not only be positive, but must also surpass the GDP growth rate in the long run, due to arbitrage in the capital market. In the opposite direction, the integrand increases in proportion to the economic value for the firm of each client (λ). This value should not increase indefinitely at a positive rate, because it is unlikely that the barriers to entry to the industry can rise at a rate large enough to increase the profit margin indefinitely. This suggests that $\lambda(t)$ does not grow in the long run.

We must consider also the rate of increase in the clientele (A). Although this rate must eventually converge to the population growth rate in any viable path, monopolistic competition assumes that the firm will always remain small as compared to the overall market, so we cannot use this argument to bound the growth rate of the clientele. However, Appendix 1 shows that when the Hamiltonian is concave in price, at least one steady state exists for the optimal actions of a single firm, with the local dynamics shown in figure 1. This means that price increments raise exit rates and reduce entries fast enough to put a limit on the size of the clientele. Therefore, provided the discount rate is positive and that the Hamiltonian is concave in price, requirement (iv) is met as well and a solution to program (B) exists[21].

When there is a single steady state, the laws of movement in figure 1 obtain (as proved in Appendix 1). Although the optimal path is unique near the steady state, figure 1 shows two optimal paths. This is because parameter values determine whether the optimal path is such that prices rise or fall as the clientele approaches the steady state from below.

---

[21] Note that the Hamiltonian should be concave in current prices even if the consumer takes into account future relative prices as well.

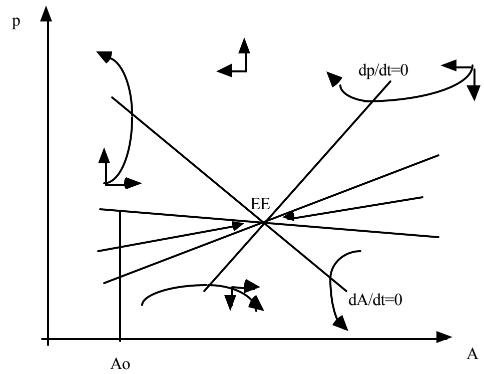

Figure 1: Phase diagram for problem (2)

3. Optimal strategies for a monopolistically competitive firm

This section presents the solution of program (B) in a familiar and intuitive format. Recall that a monopolistically competitive firm operating in a stable industry takes rivals' actions as given.

3.1 Optimal pricing strategy with brand loyalty

Recall the trade-off faced by a firm with brand loyalty: if price is reduced, the firm attracts new customers who will yield a flow of profits in the future, but it also forgoes the opportunity to extract profits now from loyal customers. If program (B) is applicable, this trade-off is resolved optimally in the steady state by the following new rule (see Appendix 1):

(1) $$\frac{p^* - (c_a + c_q \cdot q^*)}{p^*} = \frac{1 + (r/\phi^*)}{\eta^E + \eta^\phi}$$

where: $\eta^E \equiv -[\partial E/\partial p]\cdot(p/E)$ = *elasticity of entries to price*. Note that by definition $\eta^E > 0$ because $-\partial E/\partial p > 0$.

$\eta^\phi \equiv [\partial\phi/\partial p]\cdot(p/\phi)$ = *elasticity of exits to price*. In this case $\eta^\phi > 0$ as well.

ϕ^* = exit rate in the steady state (where $dA/dt = 0$).

r = discount rate used by the firm.

Rule (1) can be understood as a modified Lerner Rule for products with brand loyalty. When the "effective price elasticity" defined in equation (2) is replaced in (1), and it is recognized that marginal cost is the sum of administration marginal cost (c_a) and the marginal cost of maintenance marketing ($c_q \cdot q$), our rule takes the Lerner format.

$$(2) \qquad h^{effective} \equiv (h^E + h^f)\cdot\left(\frac{f^*}{f^* + r}\right)$$

With brand loyalty, the effective price elasticity is proportional to *the sum* of the price elasticities of the two flow branches of the stock demand, ($\eta^E + \eta^\phi$). If one branch is price elastic and the other one is not, overall demand is still price elastic. When at least one flow price elasticity is large, the firm looses the freedom to chose its price and is driven towards perfect competition.

Equation (2) also shows that with brand loyalty, the discount rate r has a direct influence over pricing strategy: at a higher discount rate r, the firm is relatively more interested in squeezing profits from currently loyal customers by raising price, knowing full well that market share is being sacrificed and future profits will suffer[22].

The exit rate is simply the ratio between the flow of exits (equal to entries in a steady state) and the stock of clients. Thus, equation (2) can be interpreted as saying that the effective price elasticity depends of the size of the flow of exits for a given stock of clients. From the point of view of competition policy, (2) says that as the equilibrium flow of exits falls relative to the stock of clients, each firm obtains a "captive demand" because the effective price elasticity falls to zero. The profit-maximizing price rises without bound, despite the presence of several firms in the industry. For the same reason, if some regulation reduces churning, a monopolistically competitive firm with brand loyalty would react taking advantage of a clientele that is captive to a fuller extent and would raise prices.

The inverse of the exit rate $1/\phi^*$ can also be seen as an operational measure of brand loyalty, because it is the average residence time of clients in the firm in a steady state[23]. The average residence time is intuitively interpreted as the average time during which clients are "captive" or loyal to the firm. A higher degree of brand loyalty pushes toward the future the reductions in profits caused by a price increase, and for this reason they become discounted more heavily. To go through, this argument needs the discount rate r to be positive. This point is confirmed by (2), because as r falls towards zero, the exit rate - i.e. brand loyalty - cancels away and ceases to influence the effective price elasticity of demand[24]. Thus, when brand loyalty increases, i.e. when there is less "churning" ϕ of the clientele between firms, a monopolistically competitive firm becomes relatively more interested in squeezing profits from current customers, even though that strategy sacrifices some market share and future profits fall in present value.

In a number of industries where brand loyalty is prevalent, firms charge commission *rates* such as a percentage of assets under administration, rather than prices measured in

[22] This direct influence of the discount rate on prices has a host of implications that we do not develop in this paper. Those implications are summarized in Klemperer's (1995) overview.

[23] In any steady state the entries and exits are equal ($E = \phi \cdot A$) to assure that $dA/dt = 0$.

[24] In any case, if the discount rate r fails to be positive, the infinite-horizon integral in the objective functional of the firm diverges and one condition for existence of a solution to program (B) is not met.

dollars. For example, in mutual fund management the price is the product of a commission rate τ and a base B, so that $p \equiv \tau \cdot B$. In this case, rule (1) also asserts that when an exogenous trend or shock raises the base B, the optimal commission rate τ^* falls inversely proportionately. Moreover, rule (1) says that the fall in the commission rate should be instantaneous, even if the pricing strategy has not achieved the steady state, because program (B) fixes an optimal path for the product $\tau(t) \cdot B(t)$, without specifying how it is distributed between τ and B.

This prediction is contradicted by the experience of the equity mutual fund industry in the U.S., because those firms did not reduce commission rates (in fact, increased from 1.25% in 1985 to 1.53% in 1998), even though the extended stock market boom of the 1990's quadrupled the assets under management[25]. This contradiction is interesting because it reveals a lot about the workings of this industry. The fact that rule (1) was not followed suggests that those firms were not free to choose their price to maximize their profits. It may be postulated that the commission rates of those firms were effectively regulated by the judicial standards on "reasonable" commission rates, which mutual fund directors must meet under their fiduciary obligation[26].

To conclude this section, we discuss the optimal pricing strategy during the transition path, i.e. before reaching the steady state described by rule (1). To do that, it is necessary to determine first the economic value of a client (denoted by λ) at time t. Condition (A.2.a) in Appendix 1 shows that in an optimal path the value of a client is:

$$(3) \qquad \lambda(t) = \frac{\lfloor p(t) - (c_a + c_q \cdot q(t)) \rfloor + d\lambda / dt}{r + \phi(t)}$$

This equation says that the economic value of a client is the present discounted value of the "dividend" it yields while loyal (i.e. the gross margin of price over direct production marginal costs), plus capital gains (i.e. $d\lambda/dt$). The appropriate discount rate to determine this present value is the sum of the discount rate and the exit rate ϕ. The latter rate must be added to take into account that brand loyalty is limited.

The economic value of a client is always positive as long as the firm charges a price above marginal production cost, regardless of whether it covers marketing expenses and fixed costs. Given that the firm can always close down, this value is bound to be positive.

In the transition path to the steady state, the trade-off between attracting new customers who will be valuable in the future and squeezing profits now from loyal customers, is resolved optimally by the following new rule (see Appendix 1)[27]:

---

[25] Source: *Business Week*, November 30, 1998, article "Why Fund fees are so high? A series of lawsuits points a finger at "co-opted" directors". I thank Peter Diamond for raising this example.

[26] I thank Paul Joskow for raising this possibility.

[27] Equation (6) is obtained from the first order conditions (A.2a) and (A.2b) *alone*. The fact that equation (A.2e) is not needed for result (6) is important in the oligopoly case. In the steady state - the long run - the conditions $d\lambda/dt = 0$ and $dA/dt = 0$ (i.e. $\phi = E/A$) hold, so (6) settles into equation (3).

(4) $$\frac{p^* - (c_a + c_q \cdot q) + d\lambda / dt}{p^*} = \frac{1}{\left[h^E \cdot \left(\frac{E}{A \cdot \phi}\right) + h^\phi\right] \cdot \left(\frac{\phi}{r + \phi}\right)}$$

Rule (4) shows that in the short run the expectation of a positive capital gain (i.e. $d\lambda/dt > 0$) makes it optimal to *reduce* prices, with the aim of speeding up entry and reducing exit in order to capture the coming capital gain on additional clients. On the other hand, if the authorities announce a new regulation that reduces the exit rate, equation (3) tells us that the economic value of each client increases instantaneously, because the firm will find it desirable to raise prices *in the future*, once loyalty increases. Rule (4) tells us that a competitive firm reacts to the announcement of this new regulation by lowering *current* prices. This announcement generates a J-curve in the time path of prices.

Rule (4) also shows that in the transition path the effective elasticity of demand depends also on whether the firm's stock of repeat-purchasers is growing or falling. If the firm is growing, the ratio between the "entry rate" (E/A) and the exit rate ϕ, given by the term $(E/A\cdot\phi)$, is above 1 and the effective price elasticity is larger. The intuition for this result is that in a growing firm more of its value lies in the future rather than the present, so an optimal resolution of the tension between investing in market share, and squeezing profits from current customers, should give an additional weight to investing in market share.

However, rule (4) also establishes that this influence of future growth vanishes as the price elasticity of entries falls towards zero. This is intuitive, because if entries increase more slowly in response to price reductions, the additional weight given to investing in market share does not justify the price reduction. Other tools, such as marketing, may obtain growth with a smaller sacrifice of profits.

3.2 Optimal marketing strategies with brand loyalty

In the steady state, the optimal level of messages M* meets the following rule (see equations (A.2 c) and (A.2 a) in Appendix 1):

(5) $$\frac{(p - c_a - c_q \cdot q)}{(r + \phi^*)} \cdot \frac{\partial E}{\partial M} = \cdot c_m$$

where: $\partial E/\partial M$ = sensitivity of entries to commercial messages that attract new clients.

Rule (5) is the analogue of the Dorfman-Steiner (1954) condition for program (A), for an environment with brand loyalty. Rule (5) should be interpreted as balancing the *present value* of the benefits earned by a message (which is the profit margin earned on the stock of repeat-purchasers, times the number of new members attracted by the marginal message M), with the investment in the message (which is cost c_m). The novelty introduced by brand loyalty is that the appropriate discount rate includes the rate of depreciation of the new clientele earned by message M, which is ϕ^* and thus is inversely proportional to the degree of brand loyalty[28]. The optimal scale of marketing investment in dollar terms is $c_m \cdot M^*$.

Rule (5) implies, as its analogue for program (A) does, that a higher level of prices creates a direct economic incentive to invest more in marketing. Consider an example where a mutual fund manager experiences an increase in the funds under management, because the prices of the securities held by the fund increased. Taking as given the commission rate τ, this should be interpreted as an increase in B, the average base of the commission per customer. The result is an increase in the firm's price p ($\equiv \tau \cdot B$), and rule (5) tells us that the manager has a higher incentive to invest in marketing. However, rule (1) suggests that the commission rate should be reduced, as discussed before, so if after that rate reduction the price p has not increased, the firm should *not* invest more in marketing.

In the steady state, quality of service q* (or the level of maintenance messages) meets the following rule (see equations (A.2 d) and (A.2 a) in Appendix 1):

$$(6) \qquad \frac{(p - c_a - c_q \cdot q)}{(r + f^*)} \cdot \left(-\frac{\partial f}{\partial q} + \frac{1}{A}\frac{\partial E}{\partial q} \right) = c_q$$

The left hand side of rule (6) shows the marginal benefit of increasing the quality of service, which is the marginal reduction in the rate of exits, plus the increase in the rate of entries, times the present value of the margin earned on the members attracted or retained. The right hand side reports the cost of investment in one more unit of service quality per client (which is c_q). Again, the novelty is that the appropriate discount rate is inversely proportional to the degree of brand loyalty.

3.3 Combined first order conditions and the optimal degree of loyalty

Combining (5) and (6) with (1) it can be shown that the steady state optimal strategy meets the following conditions, when M* and q* are interior solutions:

$$(7a) \qquad \frac{c_m \cdot M^*}{p^* \cdot A} = \frac{e_{E,M}}{h^E + h^f} \quad ; \qquad (7b) \qquad \frac{c_q \cdot q^* \cdot A}{p^* \cdot A} = \frac{e_{E,q} + e_{f,q}}{h^E + h^f}$$

where $\varepsilon_{E,M} = (\partial E/\partial M)(M/E) > 0$ is the elasticity of entries to marketing messages that attract new members, $\varepsilon_{E,q} = (\partial E/\partial q)(q/E) > 0$ is the elasticity of entries to service quality (including maintenance expenditures), and $\varepsilon_{\phi,q} = -(\partial \phi/\partial q)(q/E) > 0$ is the elasticity of the exit rate to service quality.

The left-hand side of (7a) is the optimal ratio of marketing expenditure to sales, in the steady state. The right hand side shows in the numerator, the elasticities of entries and

[28] When the present value of the new clients obtained with the first message is below c_m, then the optimal marketing strategy is zero investment ($M^*= 0$). This may happen when messages are unproductive, when the operational margin ($p - c_a - k_m m$) is relatively low as compared to the cost of messages.

exits to messages, and in the numerator, the price elasticities. Equation (7a) is analogous to the optimal marketing sales ratio for model without brand loyalty like (1) (see Berndt, 1991 for a summary). Likewise, the left-hand side of (7b) is the optimal ratio of service quality expenditure to sales, in the steady state.

An important question is whether an increase in marketing investment should *cause* an increase in price. For example, when the new private pension fund management industries of Argentina and Chile increased marketing investment and churning of members in the mid 1990's, some observers asserted that commissions were forced to increase by this cost increase[29]. However, rule (1) shows that the number of messages (M) does not influence the optimal level of price directly. This is because expenditure in messages that attract new clients (M) *are not* part of production cost[30], but rather is an investment in expanding the clientele whose magnitude is chosen by the firm according to the prospects of that investment. Just as the fixed cost F does not appear in rule (1), investment $c_m \cdot M$ does not and should not appear.[31]

At the same time, equation (7) implies that, provided all elasticities are constant, it is a profit maximizing strategy to increase investment in marketing when the number of clients or repeat-purchasers grows. This *does not* imply that an increase in marketing investment would force firms to raise prices on cost grounds. It means that it is profitable to increase marketing investment in order to increase the number of clients. In sum, the causality is demand-based, not cost-based.

We can also identify the optimal "degree of brand loyalty", which is defined as the inverse of the optimal churning rate, for the steady state. The concept of an optimal degree of brand loyalty may be thought of as the dynamic equivalent of the conventional rule that "price should be raised until demand becomes optimally elastic"[32], for a firm that enjoys brand loyalty. Combining (1) and (5) to eliminate p*, we obtain:

$$(8) \qquad f^* = \frac{r}{(h^E + h^f) - 1} + \left(\frac{c_a + c_q \cdot q^*}{c_m} \right) \cdot \frac{\partial E / \partial M}{(h^E + h^f) - 1}$$

Equation (8) states that it is optimal to increase the churning rate of the firm to make it unprofitable to go on raising the investment in marketing that attracts entry. It is optimal for the degree of brand loyalty (average residence time $1/\phi^*$) of new members to fall to the level where it is too small to justify further increases in the marketing investment rate. In equation (8), the firm takes into account that as brand loyalty falls, the effective price elasticity increases and it becomes desirable to reduce price, which in turn reduces the benefit of investing in marketing and reduces further the optimal degree of brand loyalty.

[29] For example, the Chilean Superintendent of AFP for 1982-90 proposed to introduce regulations "to reduce churning, and reduce marketing investments, allowing in that way a fall in commissions" (Ariztía, 1997).
[30] Dorfman and Steiner (1954) pointed this out clearly.
[31] An increase in maintenance marketing expenditures that raises total marginal cost should increase price, according to rule (3). The situation differs in that a higher quality of service raises the cost of serving current customers.
[32] This rule is the inverted Lerner Rule: $\eta^* = p/(p-c)$.

4. Pure monopoly and strong cartels

This section shows that the rules presented in the previous section are valid for pure monopoly and strong cartels [33], after a suitable modification. The first obvious modification is to recognize that rivals do not exist in this market structure. Thus those arguments must be eliminated from the consumer response functions in program (B).

Its was argued in section 2 that exits S are proportional to the stock of members served by the firm (A), because when some factor makes it less convenient for the consumer to remain with this firm, this factor acts on the whole stock of repeat-purchasers. By symmetry with this reasoning, entries should fall when the stock of potential clients *not* served by the firm falls. Let us define:

A^U = universe of potential consumers, larger than the clientele of the whole industry.

We have argued that when the set of potential repeat-purchasers that are not served by a firm (A^U-A_t) falls, entries should also fall proportionately. Of course, as long as the firm had a small market share, the effect on the entry rate (E) of changes in (A^U-A_t) should be negligible, and thus could be safely ignored as in program (B). In a pure monopoly or strong cartel, however, this effect must be taken into account, as argued by the marketing literature based on the Lanchester model[34]. A monopoly should take into account that as its "participation" (A/A^U) grows, the effectiveness of a given level of marketing messages in attracting new repeat-purchasers should fall, because of the growing proximity of the ceiling placed on the monopoly's stock of repeat-purchasers by the universe of potential clients. For a pure monopoly we replace the dynamic restriction in program (B) with:

(9a) $$\frac{dA}{dt} = C(p,M,q)\cdot\left(1-\frac{A}{A^U}\right) - f(p,q)\cdot A$$

(9b) $$\equiv C(p,M,q) - \left(f(p,q) + \frac{C(p,M,q)}{A^U}\right)\cdot A$$

where: C(p,M,q) = flow of contacts with potential new clients.

The amendment introduced by (9a) is that entries depend negatively on the current participation of the monopoly (A/A^U). The idea behind is that any given marketing and pricing effort generates C contacts with repeat-purchasers, but a proportion of those contacts fail to bear fruit, because the client is already a client of the monopoly. Equation

[33] A "strong" cartel is a set of firms that is able to act as a pure monopoly. Both pure monopoly and strong cartels are assumed to enjoy blocked entry into their market. The implications of the threat of entry for pricing strategy have been analyzed by Fudenberg and Tirole (1984) and Klemperer (1995).

[34] The application of the military combat model by Lanchester to marketing is reviewed by Dockner at al (2000, section 11.1). Equation (10) is a version of this approach. See the model by Gould (1970), cited by Tu (1984).

(9a) postulates that this proportion is equal to the participation of the monopoly [35]. Equation (9b) reinterprets this amendment as an increment to the depreciation rate ϕ, because in this setting any contact that fails is equivalent to an additional exit. In this interpretation, the depreciation rate is increased by the "contact rate" (C/A^U).

Now we make precise the definition of size for firms that enjoy brand loyalty, distinguishing between "participation" and "market share". Consider the case of a monopoly of cellular phone service. It is intuitive that at any point in time the number of clients of the firm that have a service contract outstanding is smaller than the number of potential customers, because a number of the latter have not signed a contract. The number of clients of such a monopoly would evolve over time slowly because of brand loyalty and yet the firm would still try to attract new customers that have not signed such a contract yet. On the other hand, this firm has a market share of 100%, by definition of monopoly. This example shows that the total number of repeat-purchasers for the monopoly is smaller than the number of potential customers (denoted by A^U) because some of them have not developed explicit or implicit loyalties to any particular firm. In the general case of several firms:

$$(10) \qquad A_i + \sum_{j \neq i} A_j < A^U \qquad \Rightarrow \qquad s_i \equiv \frac{A_i}{\sum A_k} < \frac{A_i}{A^U} \equiv \alpha_i$$

where: s_i = market share of firm i according to outstanding service contracts.

α_i = participation of firm i in total potential customers ($\alpha_i < 1$ for all i).

Equation (10) makes clear that a pure monopoly or a cartel (where s = 1) does not enjoy total participation of all potential customers, except in special cases [36].

The size of the universe of potential repeat-purchasers, A^U, responds to the actions of a pure monopoly. We expect a lower average price and higher total marketing investment to increase the size of the market (A^U). Note that potential customers are not "repeat-purchasers", because they are *not* loyal to any firm, even if there is only one firm available. For this reason, we model potential clients with the approach of program (A), rather than the approach of program (B). In this spirit, we describe the size of the market (A^U) with the following consumer response function, which is valid for the general case with several firms [37]:

---

[35] Gould (1970) analyzed the case where entries are proportional to the product $A \cdot (A^U - A)$, which may be justified when the contacts are made mainly by recommendation of current clients, not by marketing messages paid by the firm. Gould only solved the simplified case where price p is exogenous. See a summary in Tu (1984) p. 334-7.

[36] One such case is the one of privatized social security, where the overall market for fund management services is mandatory, implying that $\partial A^U / \partial \overline{p} = 0$, $\partial A^U / \partial \overline{q} = 0$ and $\partial A^U / \partial \sum M_j = 0$. In the setting, a cartel or a pure private monopoly could extort very high profits from consumers.

[37] This is the third consumer response function that the firm must tajke into account. This function is free of brand loyalty because it does not include any "dynamic restriction" to represent the net accumulation of potential customers.

$$(11a) \qquad A^U = A^U(p, M, q) \qquad \textit{for a pure monopoly}$$

$$(11b) \qquad A^U = A^U\left(s_i p_i + (1-s_i)\overline{p}_{j\neq i};\ \ M_i + \sum_{j\neq i} M_j;\ \ s_i q_i + (1-s_i)\overline{q}_{j\neq i} \right) \qquad (\textit{oligopoly})$$

where $s_i = A_i/A^U$ is the market share of firm i in the oligopoly case.

We assume that $\partial A^U / \partial \overline{p} < 0$, that $\partial A^U / \partial \sum M_j > 0$, and $\partial A^U / \partial \overline{q} > 0$.

As a monopoly expands, its participation remains below 100% and the term (1-A/A^U) remains positive. This means that the effectiveness of marketing messages does not fall to zero, although the marginal cost of increasing participation may rise to very high levels. Because of this cost, in most cases the monopoly would not want to increase its participation to 100%. Thus, the appropriate program for a pure monopoly is (C):

$$(C) \qquad \underset{\{p(t); M(t); q(t)\}}{Max} \left\{ V^D = \int_{t=0}^{\infty} \left[p \cdot A - F - c_a \cdot A - c_m \cdot M - c_q \cdot q \cdot A \right] e^{-rt} dt \right\}$$

subject to:

$$1) \quad \frac{dA}{dt} = \left(1 - \frac{A}{A^U(p, M, q)} \right) \cdot C(p, M, q) - A \cdot f(p, q)$$

$$2) \quad A(t) \geq 0$$

$$3) \quad M(t) \geq 0 \qquad \textit{and} \quad 4) \quad A(0) = A_o$$

In what follows, we present the steady state pricing and marketing rules for a pure monopoly, taking into account the amendments included in equations (9) and (11). Appendix 2 shows that the optimal pricing strategy for a pure monopoly or cartel is :

$$(1') \qquad \frac{p^* - (c_a + c_q \cdot q^*)}{p^*} = \frac{1}{\left((1-a)\left[h^C + h^f\right] + a \cdot h^{AU,p} \right) \cdot \left(\dfrac{f}{r(1-a) + f} \right)}$$

where $\eta^{AU,p} \equiv -[\partial A^U / \partial p] \cdot (p/A^U)$ = price elasticity of potential market size.

$\eta^C \equiv -[\partial C / \partial p] \cdot (p/C)$ = price elasticity of successful contacts.

Rule (1') shows that as the participation of the monopoly approaches 100%, the standard Lerner Rule gives the profit-maximizing margin[38]. The optimal margin is inversely proportional to the elasticity of potential market size to monopoly price. The

[38] The influence of the discount rate vanishes because it is multiplied by (1-α), and then the churning rate cancels out.

influence of brand loyalty vanishes as participation increases, because potential customers are not brand loyal. Regarding potential customers, the trade-off between attracting more customers and forgoing the opportunity to squeeze profits from current customers occurs within the same instant and is the one modeled by Lerner (1934).

Regarding optimal marketing strategy, (5) changes to:

$$(5') \qquad \frac{(p - c_a - c_q \cdot q)}{(r \cdot (1-\alpha) + f)} \cdot \left[(1-\alpha)^2 \cdot \frac{\partial C}{\partial M} \; + \; \alpha^2 \cdot f \cdot \frac{\partial A^U}{\partial M} \right] = c_m$$

where $[\partial A^U/\partial M]$ = sensitivity of potential market size to marketing effort.

Rule (5') shows that as the monopoly raises its participation rate α, the response that dominates is the one of potential demand, and the Dorfman-Steiner rule is recovered[39]. Regarding optimal service strategy, or marketing maintenance, (6) changes to:

$$(6') \qquad \frac{(p - c_a - c_q \cdot q)}{(r(1-\alpha) + f)} \cdot \left((1-\alpha)^2 \left[\frac{1}{A} \frac{\partial C}{\partial q} \right] + \alpha^2 \frac{f}{A} \frac{\partial A^U}{\partial q} + (1-\alpha) \left[\frac{-\partial f}{\partial q} \right] \right) = c_q$$

where $\partial A^U / \partial q$ = sensitivity of potential market to the monopoly's quality of service.

Again, rule (6') shows that as a monopoly raises its participation, the potential demand response becomes the main consideration for setting service quality.

5. Dominant firms and oligopolies

In this section we consider the impact of a firm's actions on rivals' actions, and identify the set of cases where this consideration does not void the rules found in sections 3 and 4. We analyze only two market structures: the first exhibits a single large firm (the "dominant" firm) and a "fringe" made up of many small firms, all of which enjoy brand loyalty. The second structure exhibits at least two "large" firms who act independently, i.e. do not cooperate as in a cartel, where no fringe is present. This is the standard pure oligopoly case with no collusion. Unlike pure monopolies or monopolistically competitive firms, a dominant firm or an oligopolist would take into account the influence of its actions on the actions of the fringe or of rivals, respectively.

There are many ways to model an oligopoly game over time, but here we discuss only two: In the first polar case, each firm in the fringe (each rival) cannot observe in real time the actions of the dominant firm (of the oligopolist). If the dominant firm changes an action at some point in time, the fringe ignores it and goes ahead with its initial plan. An alternative assumption that leads to this same behavior is that each fringe firm is able to commit to follow the strategy initially planned, and does so. We call this situation "full commitment". We only analyze the case where all participants commit to their action paths

[39] In the limit when participation rises to 100%, violating (9), the influence of the discount rate vanishes because it is multiplied by (1-α), and then the churning rate cancels out.

simultaneously, and all anticipate the chosen paths of the others (this is the Nash assumption).

In the other polar situation, each firm in the fringe (each rival) observes in real time the actions of the dominant firm (the oligopolist) and cannot commit not to respond to changes in its actions. The literature calls this situation "no commitment". In it, the dominant firm can influence the fringe's actions by changing its own actions after the start of the game, as the game is played, while the fringe's plans are being implemented. The dominant firm can anticipate the fringe's reactions. However, the fringe firms can also anticipate the dominant firm's attempts to take advantage of its position (the Nash assumption again). When the fringe anticipates the dominant firm's inability to restrain itself later in time, it may modify its actions and may constrain the dominant firm further.

Other ways to model an oligopoly game over time, not considered here, allow firms to consider strategies that are a function of other factors apart from the current state, such as the past history of the game (past prices, past marketing investments, past stocks of repeat-purchasers). Therefore, these approaches take into account reprisals (punishments) and the strategies based on them, such as threats to rivals or to defectors on agreed target paths. This is the basis of strategies such as reputation, wars of attrition and tacit collusion. Although the two equilibrium concepts used below are not the only ways to model an oligopoly game over time, they do cover substantial ground for the case where oligopolists do not collude tacitly.

5.1 Full commitment

The full-commitment situation is modeled with the open-loop Nash equilibrium, which assumes that each firm implements its chosen dynamic strategy by looking at the clock only (see Dockner at al, 2000, section 3.5). In the case of the market structure with a dominant firm, each member of the fringe behaves as described in sections 2 and 3, i.e. each one is monopolistically competitive. Each of them knows that it cannot affect the strategies or the state of the dominant nor the average firm, and thus it is rational for it to take those strategies and states as given. However, we lift the assumption made in section 2 that the strategies and the state of the rest of the industry are constant over time. Due to full commitment, the fringe's actions are expressed as a function of time alone.

The dominant firm takes into account that it may influence the number of potential customers, as shown by equation (11b). It also knows that it can affect the actions and the state of the fringe. In the open loop equilibrium, the fringe actions are fully committed, so they will follow the planned time path and it is rational for the dominant firm to take those actions as given. The dominant firm inserts the fringe actions, as a function of time, into the consumer response functions it faces.

The state of the fringe, i.e. the size of its stock of repeat-purchasers, evolves according to three influences: the actions of the fringe, the state of the dominant firm[40], and the actions of the dominant firm. The presence of this third influence *does not* imply that

[40] The state of the dominant firm influences the state of the fringe as explained in section 2: given a exit rate from the dominant firm, the number of exits increases in proportion to the size of that clientele. Although a portion of the exits from the dominant firm joins the pool of "fast switchers", the rest joins the fringe, so the number of entries to the fringe is expected to increase when the number of exits from the dominant firm rises. Recalling that the number of exits from the dominant firm is proportional to its stock of repeat-purchasers, for a given marketing and pricing effort, the final result is that a higher stock of clientes at the dominant firm raises the number of entries to the fringe.

the first order conditions for the dominant firm's actions should include a new term to reflect it. That influence is captured by the dynamic constraint affecting each fringe firm, and this will influence the level of price and other action paths in the open loop Nash equilibrium. Those levels are taken into account by the dominant firm and by each fringe firm when deciding their own strategy paths.

Bringing all these adjustments together, the dominant firm solves program (D):

$$(D) \qquad \underset{\{p(t);M(t);q(t)\}}{Max} \left\{ V^D = \int_{t=0}^{\infty} \left[p \cdot A - F - c_a \cdot A - c_m \cdot M - c_q \cdot q \cdot A \right] e^{-rt} dt \right\}$$

subject to:

$$1) \quad \frac{dA}{dt} = (1 - \frac{A}{A^U}) \cdot C\left(p, M, q, \overline{p}_{fr}(t), \overline{q}_{fr}(t), \sum_{j \neq D} M_j^{fr}(t), A_{fr} \right) - A \cdot f\left(p, q, \overline{p}_{fr}(t), \sum_{j \neq D} M_j^{fr}(t), \overline{q}_{fr}(t) \right)$$

$$2) \qquad A^U = A^U \left(s_d p + (1 - s_d) \overline{p}_{fr}(t); \;\; M_i + \sum_{j \neq i} M_j^{fr}(t); \;\; s_d q + (1 - s_d) \overline{q}_{fr}(t) \right)$$

$$3) \;\; A(t) \geq 0 \qquad 4) \;\; M(t) \geq 0 \qquad and \;\; 5) \quad A(0) = A_o$$

where: s_d = market share of the dominant firm, i.e. $s_d = A / (A + A_{fr})$.

Program (D) differs from the program for pure monopoly in three respects[41]: first, the fringe's action paths influence entries and exits faced by the dominant firm, because those paths are inserted into the consumer response functions faced by the dominant firm. The system dynamics (constraint 1) in program (D)) exhibits a definite change from that in program (C), namely that fringe actions over time are taken into account by the dominant firm. Thus price and marketing levels in the full commitment situation are heavily influenced by the presence of the fringe and its anticipated (Nash) response, as compared to pure monopoly. Second, the relative size of the dominant firm given by its market share (s_d), influences the size of the potential market, for any given strategy. Third, the dominant firm takes into account that as its market share (s_d) grows, the impact of changes its price and quality on the size of the potential market A^U grows (see constraint 2)).

The combination of program (D) for the dominant firm and a set of programs like (B) for each fringe firm forms a differential game. Dockner at al (2000) in Theorem 4.2 show that in differential games the Nash equilibrium can be obtained by applying the Maximum Principle to each program and by applying the adjoint equation to the maximized Hamiltonian. The adjoint equation is critical because it determines $\lambda(t)$, the "value of a client", which in turn is needed to obtain the rule that is analogous to (1).

[41] There is a third difference with the case of a fringe firm: the dominant firm takes into account the impact of its size on the success rate of its marketing contacts, just as pure monopolies do.

Appendix 3 shows that this derivative takes the same form as in program (C), except for an adjustment due to the fact that the market share of the firm depends of the clientele A, in the full commitment case. This result is entirely due to the fact in this case, the rivals' and this firm's actions are a function of time alone and not of the state of the firm.

As shown in Appendix 3, the first order conditions for a pure monopoly remain approximately valid for the dominant firm in the steady state, as shown by equation (12):

$$(12)\qquad \eta^{eff} = \left[(1-\alpha)\cdot(\eta^{C}+\eta^{f})+\alpha\cdot\eta^{AU,p}\right]\cdot\left(\frac{\phi}{r\cdot(1-\alpha)+\phi\cdot\left[1+\alpha\cdot s\cdot(1-s)\cdot K\right]}\right)$$

where $K = (\eta^{AU,p})\cdot(p-p_{-i})/p - (\eta^{AU,q})\cdot(q-q_{-i})/q$

Equation (12) shows the appearance of a that new term modifies the impact of the "contact rate" (C/A^{U}) on the depreciation rate ϕ. As the market share of the firm (s) moves to either one or zero, keeping participation α constant, the optimal pricing rule reverts to that for a pure monopoly. As the participation α of the firm in the potential market falls to zero, rule (12) reverts to rule (1) for a monopolistically competitive firm. The new term is also a function of the elasticity of the potential market to changes in average prices and average quality.

The main difference between the effective price elasticity for a pure monopoly (rule (1')) and the effective price elasticity in (12) for a dominant firm is factor K. This factor exists because the dominant firm takes into account that as its market share grows, the impact of changes in its price and quality on the size of the potential market A^{U} grow (see constraint 2)). But when the following "symmetry" conditions obtain in equilibrium, changes of the dominant firm's price and quality have zero impact on the size of the potential market:

$$(13)\qquad p = \overline{p}_{-i} \quad ; \quad q = \overline{q}_{-i}$$

It is clear that K = 0 when (13) obtains. But the dominant firm will choose how much to deviate from condition (13), so the adjustment required by (12) is justified.

Of course, even if K=0, rule (12) does not imply that the level of pric e and other actions in the open loop Nash equilibrium are the same as in pure monopoly, because the system dynamics in programs (D) and (B) are definitely different from that in program (C).

Given these results, it is easy to cover the oligopoly case briefly. Each oligopolist takes into account the plans of its rivals when deciding its own strategy. In the full-commitment situation, where each oligopolist can commit to an action plan over time, the oligopoly game can be described as follows: each oligopolis t solves a program like (D), after replacing the average of the fringe's actions for the average of its rivals' actions.

The Nash equilibrium is obtained by applying the maximum principle to a set of programs like (D), one for each oligopolist. No additio nal terms appear in the first order conditions, as compared to the dominant firm case. The implication is that the optimal pricing and marketing rules for a dominant firm, including equation (12), apply to each oligopolist as well.

Again, this result does not suggest that the price and marketing *levels* that result in an oligopoly game are the same as those that result in a dominant firm game, when both remain within the full commitment situation. This is proven by the fact that in the oligopoly game, the system dynamics of rivals are those in program (D), while in the dominant firm game, the system dynamics of fringe firms are those in program (B).

5.2 No commitment

The no-commitment situation is modeled in the literature with the Markov Perfect approach, which assumes that each firm implements its chosen strategy as a function of its *state* and not of time alone. The state of the game as seen by firm i is the number of repeat-purchasers at firm i (A_i), and also the number of repeat-purchasers at its rivals (A_{-i}). Recall that the firm's entries E depend positively on the size of the rivals' clientele (A_{-i}) because given an overall number of exits from the rivals, the number of entries into this firm should increase, for a given marketing and pricing effort.

Therefore, the actions chosen by each firm are a function of its own state (A_i, the stock of its repeat-purchasers) and of time, as can be verified in Figure 1. For rivals, we denote these functions as:

$$(14) \qquad p_{-i}^*(t) = Z(A_{-i}, A_i, t) \quad ; \quad \sum M_{-i}^*(t) = W(A_{-i}, A_i, t) \quad ; \quad q_{-i}^*(t) = X(A_{-i}, A_i, t)$$

The no-commitment situation makes it natural to require the equilibrium to be subgame-perfect. These are equilibria where the actions of all players after an unexpected deviation from the original equilibrium path occurs, are also Nash equilibria.

To obtain the Markov Perfect Nash equilibrium (MPNE) for the game played by a single dominant firm and a large set of small fringe firms with no commitment, we proceed in the following way: each of the fringe firms follows the rules described in sections 2 and 3 for the monopolistically competitive firm, except for the fact that we abandon the assumption that the dominant firm chooses a time-invariant path. Each fringe firm's problem is program (B) augmented by (14), where its "rivals" include the dominant firm and the rest of the fringe. Note that the fringe programs are different from those used to analyze the full-commitment situation, because equations (14) were not used there.

For the dominant firm, the functions (14) consider as rivals the fringe firms only, so $A_{-i} = A_{fr}$, and these functions are inserted into the consumer response functions in the dominant firm's problem. The dominant firm's problem takes the form of program (E):

$$(E) \qquad \underset{\{p(t);M(t);q(t)\}}{Max}\left\{V^D = \int_{t=0}^{\infty}\left[p\cdot A - F - c_a\cdot A - c_m\cdot M - c_q\cdot q\cdot A\right]\cdot e^{-rt}dt\right\}$$

subject to :

$$1)\ \frac{dA}{dt} = (1-\frac{A}{A^U})\cdot C(p, M, q, Z(A_{fr}, t), X(A_{fr}, t), W(A_{fr}, t), A_{fr})$$
$$- A\cdot f(p, q, Z(A_{fr}, t), W(A_{fr}, t), X(A_{fr}, t))$$

$$2)\ A^U = A^U\left(s_d\cdot p + (1-s_d)\cdot Z(A_{fr}, t);\ \ M + W(A_{fr}, t);\ \ s_d\cdot q + (1-s_d)\cdot X(A_{fr}, t)\right)$$

$$3)\ A(t)\geq 0 \qquad 4)\ M(t)\geq 0 \qquad and\ \ 5)\ \ A(0) = A_o$$

where: s_d = market share of the dominant firm, i.e. $s_d = A / (A + A_{fr})$.

Program (E) differs from program (D) in a single and important way: the fringe's actions are expressed as a function of each firm's state and time, not only of time.

To find the MPNE, one must start finding the set of Nash equilibria of the game. As reported before, Dockner at al (2000) in Theorem 4.2 show that in differential games the Nash equilibrium can be obtained by applying the Maximum Principle to each program and by applying the adjoint equation to the maximized Hamiltonian[42]. In a second stage, one must discard the Nash equilibria that are not subgame perfect[43]. Contrary to the full-commitment case, the maximized Hamiltonian does not coincide with the ordinary Hamiltonian, because the optimal action paths depend of the state of the game in a different way due to the addition of (14). This implies that the maximized Hamiltonian has two additional sets of terms, as explained in Appendix 3. These terms change the "value of a client". An expression for the value of a client cannot be found without knowing the solution of the game.

The intuition is that it becomes important to take into account that if the state of the firm changes, the rivals will change their actions and this will modify the value of a client. But without an expression for the value of a client, it is no longer possible to express the optimal pricing rule in a general form like (1) or (12).

The same happens in the oligopoly case in the no commitment situation. Program (E) applies now to each oligopolist, after replacing the average of the fringe's actions for the average of the rivals' actions. For the reasons just explained, the rules we identified in section 5.1 are not valid for oligopolists in the no commitment situation.

6. Final comments

This paper presents a model of monopolistic competition in markets with brand loyalty. Its main purpose is to provide guidance to firms that face a dynamic trade-off in

---

[42] See Dockner at al (2000), p. 97-98.

[43] One may also try to find immediately the subgame perfect Nash equilibrium, by using the Hamilton-Jacobi-Bellman equations (a set of coupled partial differential equations). However, that approach requires finding the solution of the game, while this paper seeks only to identify the first-order conditions.

pricing: if the firm reduces price it attracts new customers who will be valuable in the future, but it also forgoes the opportunity to extract profits now from current customers. In the continuous time model proposed, which is based on general consumer response functions which can be estimated econometrically, the optimal resolution of this trade-off occurs in terms of a new family of rules that is highly intuitive.

It is shown that the “effective” price elasticity depends on the discount rate used by the firm, on the rate of depreciation of the clientele through exit from the stock of repeat purchasers, and on the sum of the price elasticities of entries into the stock of repeat-purchasers and exits from the stock of repeat-purchasers. None of these factors enter the optimal pricing strategy for a firm facing a conventional demand function where consumers are not repeat-purchasers.

The optimal rules for investment in marketing and in quality of service are derived jointly with optimal prices, and they turn out to be intuitive modifications to the well-known Dorfman-Steiner rule. It is shown also how these rules should be adapted in dynamic environments such as a growing market or announced changes in regulations.

With modest modifications, these optimal pricing and marketing rules apply also to pure monopolies, cartels, dominant firms and oligopolies that can commit to strategy paths. For dominant firms and oligopolists that cannot commit to strategy paths, it is not possible to obtain a rule with the same simple form.

References

Ariztía, J. (1997) "Análisis sobre los traspasos en el sistema previsional", en J. Ariztía, editor, *AFP: Las Tres Letras que Revolucionan a América*, CIEDESS, Santiago, Chile, p. 311-332.

Beggs, A. and P. Klemperer (1992) “Multiperiod competition with switching costs”, *Econometrica*, 60, p. 651-666.

Berndt. E. (1991) "Causality and Simultaneity between Advertising and Sales", capítulo 8 en *The Practice of Econometrics: Classic and Contemporary,* p. 361-448; Addison-Wesley Publishing Co., Reading, Massachusetts.

Bernstein, S. and A. Micco (2001) “Price competition and turnover: the Chilean pension fund industry”, mimeo Boston University and Inter American Development Bank, dated May 31 and presented at the Latin American Economic Association meetings in Montevideo, Uruguay, October 18-20, 2001.

Chen, Y. (1997) “Paying customers to switch”, *Journal of Economics and Management Strategy*, 6, No. 4, p. 877-897.

Dockner, E., S. Jorgensen, N. Van Long and G. Sorger (2000) *Differential Games in Economics and Management Science*, Cambridge University Press.

Dorfman, R. and P. Steiner (1954) “Opimal advertising and optimal quality”, *American Economic Review* 44, p. 826-836.

Evans, G.C. (1924) “The dynamics of monoply”, *American Mathematical Monthly*, vol. 31 (2), February, p. 77-83.

Farrell, J. and C. Shapiro (1988) “Dynamic competition with switching costs”, *Rand Journal of Economics*, 19, p. 123-137.

Fershtman, C. and M. Kamien (1987) “Dynamic duopolistic competition with sticky prices”, *Econometrica* 55 (5), September, p. 1151-1164.

Friedman, B. (1983) *Oligopoly Theory*, Canbridge University Press.

Fudenberg, D. and J. Tirole (1984) “The fat cat effect, the puppy dog ploy and the lean and hungry look”, *American Economic Review Papers and Proceedings*, v. 74, p. 361-368.

Gould, J.P. (1970) "Diffussion processes and optimal advertising policy", en E.S. Phelps, edit or, *Microconomic Foundations of Employment and Inflation Theory*, Norton, New York.

Klemperer, P. (1987) “Markets with consumer switching costs”, *Quarterly Journal of Economics*, 102, p. 375-394.

Klemperer, P. (1995) “Competition when consumers have Switching Costs: an overview with applications to Industrial Organization, Macroeconomics and International Trade”, *Review of Economic Studies* v. 62, No. 4, p. 515-539.

Lerner, A. (1934) “The concept of monopoly and the measurement of monopoly power”, *Review of Economic Studies*, June.

Nerlove, M. and K. Arrow (1962) “Optimal advertising policy under dynamic conditions”, *Economica* 39, p. 129-142, May.

XXX To be completed. Search literature reference

Padilla, J. (1995) “Revisiting Dynamic Duopoly with Consumer Switching Costs”, *Journal of Economic Theory* v. 67 (2), December 1, pp. 520-530.

Rizzo, J. (1999) “Advertising and competition in the Ethical Pharmaceutical Industry: the case of Antihypertensive drugs”, *Journal of Law and Economics* 42 (1), Part 1, p. 89-116.

Roberts, M. and L. Samuelson (1988) “An empirical analysis of dynamic, nonprice competition in an oligopolistic industry”, *Rand Journal of Economics* 19, No. 2, Summer, p. 200-220.

Schmalensee, R. (1976) “A model of promotional competition in oligopoly”, *Review of Economic Studies*, vol. 43, No. 4 (October), p. 493-507.

Simaan, M. and T. Takayama (1978) “Game Theory applied to dynamic duopoly problems with production constraints”, *Automatica*, 14, p. 161-166.

Stahl, X. (1996) “Oligopolistic pricing with heterogeneous consumer search”, *International Journal of Industrial Organization*, v. 14 (2), p. 243-268.

Tu, Pierre (1984) *Introductory Optimization Dynamics*, Springer-Verlag, New York.

Wernerfelt, B. (1991) “Brand loyalty and market equilibrium”, *Marketing Science*, 10 (3), Summer, p. 229-245.

Appendix 1: Solution to the optimization program (B)

For a monopolistically competitive firm operating in a stable industry, rivals' actions as fixed over time. As those arguments of demand are fixed, we can ignore them. Assuming the second and third restrictions in program (B) are slack, the Hamiltonian is:

(A.1) $H(p, M, q, A, \lambda) = e^{-rt} \cdot \{ (p-c_a)\cdot A - F - c_m\cdot M - c_q\cdot q\cdot A + \lambda\cdot[E(p,M,q) - \phi(p,q)\cdot A] \}$

$\lambda(t)$ is interpreted as the value of a new client attracted or lost at date t. This is the effect of a unit change in the clientele in the present value of the firm as of time t. As of time 0, its value is $\lambda\cdot e^{-rt}$.

First Order Conditions

From the Maximum Principle:

(A.2 a) $\partial H/\partial A = -d(\lambda\cdot e^{-rt})/dt$	=>	$(p- c_a- c_q\cdot q)/\lambda - \phi + d\ln\lambda/dt = r$
(A.2 b) $\partial H/\partial p = 0$	=>	$A = \lambda\cdot(- \partial E/\partial p) + \lambda\cdot (\partial\phi/\partial p)\cdot A$
(A.2 c) $\partial H/\partial M = 0$	=>	$c_m = \lambda\cdot \partial E/\partial M$
(A.2 d) $\partial H/\partial q = 0$	=>	$c_q\cdot A = \lambda\cdot [(\partial E/\partial q) + (^-\partial\phi/\partial q)\cdot A]$
(A.2 e) $\partial H/\partial(\lambda\cdot e^{-rt}) = dA/dt$	=>	$dA/dt = E(p,M,q) - \phi(p,q)\cdot A$

The right hand side of (A.2 a) shows the total return over capital invested in clients as of time t, which should be equal to alternativ e return r along the optimal path. Equation (A.2b) recommends an increase in price p until the gain from a higher income from the current clientele is fully outweighed by the sum of two losses: the reduction in entries and the increase in exits. Equation (A.2c) recommends an increase in marketing messages that attract new clients as long as the value of the new clients attracted is above the marginal cost of the message. Equation (A.2d) recommends that the quality of service, or the level of maintenance marketing messages, be raised as along as the increase in marginal cost of production remains below the marginal benefit, which in turn is the sum of value of the clients attracted and the value of the clients that are not lost.

These first order conditions are valid for interior optima only. If, for example, $A\cdot c_q > \lambda\cdot[(\partial E/\partial q) + (^-\partial\phi/\partial q)\cdot A]$, then q* = lower bound on quality, i.e. the firm should stick to lowest possible level of service quality (possibly set by regulations) and avoid maintenance marketing.

Second order conditions

The second order condition for program (B) is that second variations of the controls p, M and m over H be a negative semidefinite matrix. At the very least, the diagonal elements must be negative:

(A.4.a)	$H_{pp} < 0$	=>	$\lambda\cdot[\partial^2E/\partial p^2 - A\cdot \partial^2\phi/\partial p^2] < 0$
(A.4.b)	$H_{MM} < 0$	=>	$\lambda\cdot \partial^2E/\partial M^2 < 0$
(A.4.c)	$H_{qq} < 0$	=>	$\lambda\cdot [(\partial^2E/\partial^2q) + (^-\partial^2\phi/\partial^2q)\cdot A] < 0$

to which must be added the other conditions on cross products of the cross derivatives (not shown).

Border (Transversality) Conditions

In program (B), the border conditions are the initial condition $A(0) = A_o$ and the terminal or transversality condition, which is (A.3.a). Equation (A.3b) arises from (A.3a) and the second restriction A(t) = 0.

(A.3.a) $\lim_{t\to\infty} \lambda^*(t)\cdot e^{-rt}\cdot A^*(t) = 0.$

(A.3.b) $\lim_{t\to\infty} \lambda^*(t)\cdot e^{-rt} \geq 0$ (see Tu, 1984, p. 134).

Condition (A.3.a) is met if and only if r >0 and the product λ*(∞)·A*(∞) is bounded. Condition (A.3.b) is met only if r >0 and λ*(∞) is bounded. Therefore, the border conditions require that the discount rate is positive. Considering the definition of the effective price elasticity in equation (5) of the main text, it is assured that this elasticity cannot be negative, for any exit rate. Another implication of the border conditions is that (A.3.a) and (A.3.b) together require that A*(∞) be bounded. This may be interpreted as a requirement that marketing messages and price reductions should not be too effective in attracting new clients.

Existence of at least one steady state

If program (B) is consistent with economic experience, the optimal path in the control variables should at least exhibit one steady state. To verify this we use a phase diagram in the main control variable - price p - and the state variable, the number of clients A. Depending on the signs of the slopes of curves dA/dt = 0 and dp/dt =0 when they cross in this diagram, a steady state may or may not exist. In particular, if curve dA/dt = 0 has a negative slope and curve dp/dt =0 has a positive slope for all parameter values when they cross, then a steady state exists.

Program (B) includes three control variables: p, M y q. To simplify this proof, we suppress controls M and q, assuming they take fixed values.

Equation (A.2e) indicates that: dA/dt = E(p) - ϕ(p)·A Thus, there exists a curve where dA/dt = 0 in (p,A) space. The slope of this curve is obtained by differentiating (A.2e) totally. The result is:

(A.7) $$\left.\frac{dp}{dA}\right|_{\frac{dA}{dt}=0} = \frac{-f}{-E_p + f_p A}$$

Equation (A.2.b) implies that the denominator is positive because A/λ> 0. As ϕ> 0, equation (A.7) proves that the slope of curve dA/dt = 0 is negative for all parameter values.

To find the curve dp/dt =0 we use the other first order conditions. First, we differentiate (A.2b) with respect to time. Then we replace the term in dλ/dt using equation (A.2.a). We also replace the terms in dA/dt using equation (A.2.e). Finally the terms in λ are replaced using equation (A.2.b). The result is:

(A.8) $$\frac{dp}{dt} = \left(\frac{E}{A} - f\right)\frac{(-E_p)}{\left(-E_{pp} + Af_{pp}\right)} + \left(\frac{(p - c_a - c_q \cdot q)}{A/(-E_p + Af_p)} - (r+f)\right)\left[\frac{-E_p + Af_p}{E_{pp} + A\cdot(-f_{pp})}\right]$$

This curve represents the evolution of price over time along optimal paths. The curve dp/dt =0 is obtained setting the left-hand side of (A.8) to zero. The resulting relationship between p and A is:

(A.9) $$p = (c_a + c_q \cdot q) + \frac{(r+f)\cdot A}{-E_p + Af_p} + \frac{(f\cdot A - E)(-E_p)}{(-E_p + Af_p)^2}$$ (curve dp/dt=0)

The expression in both denominators is positive by (A.2b). To find the slope of this relationship in (p,A) space, we differentiate (A.9) totally with respect to p and A. The result is a complex expression, but it simplifies radically at the point(s) where this relationship crosses the curve dA/dt = 0, because in those crossings dA/dt = E - ϕA = 0. Thus, in every crossing point between the curves dp/dt =0 and dA/dt =0, the former curve has the slope:

(A.9) $$\left.\frac{dp}{dA}\right|_{\frac{dp}{dt}=0} = \frac{-E_p}{A}\cdot\frac{(r + 2f / r + f)}{(-E_{pp} + f_{pp}A)} > 0 \qquad where \ \frac{dA}{dt} = 0 \ \ is \ crossed.$$

Equation (A.4a) requires the denominator to be positive. As $E_p < 0$ by assumption, this expression is positive for any parameter value. It is possible that the curve dp/dt =0 may have a negative slope in other regions of plane (p,A), but locally, where the two curves cross, its slope is always positive. We have not ruled out the possibility of multiple crossings, and thus of multiple steady states.

Summing up, the curves dA/dt =0 and dp/dt =0 have slopes with the signs shown in Figure 1, at least locally . The laws of movement in this figure imply that a steady state exists (at least one). The requirement for this result is the validity of equation (A.4a) - the Hamiltonian is concave in price.

In the more general case where the firm controls M and q in addition to p, much more elaborate calculations are needed to prove uniqueness using these simple method. Generally speaking, uniqueness appears to require that the cross derivative E_{pM} not to be too large and positive. This is a natural assumption to make when the second order conditions are met, because they limit the size of this cross derivative.

Appendix 2: First order conditions of to optimization program (C)

For a pure monopoly, and assuming that the second and third restrictions in program (C) are slack, the Hamiltonian is:

(B.1) $H(p,M,q,A,\lambda) = e^{-rt}\cdot\{ (p-c_a)\cdot A - F - c_m\cdot M - c_q\cdot q\cdot A + \lambda\cdot[C(p,M,q) - A\cdot\{\phi(p,q) + C(p,M,q)/A^U(p,M,q)\}]\}$

From the Maximum Principle, the first order conditions for an interior optimum are:

(B.2 a) $\partial H/\partial A = -d(\lambda\cdot e^{-rt})/dt$ => $(p - c_a - c_q\cdot q)/\lambda - (\phi + C/A^U) + d\ln\lambda/dt = r$

(B.2 b) $\partial H/\partial p = 0$ => $A = \lambda\cdot(-\partial C/\partial p) + \lambda\cdot[\partial\phi/\partial p + \partial(C/A^U)/\partial p]\cdot A$

(B.2 c) $\partial H/\partial M = 0$ => $c_m = \lambda\cdot\partial C/\partial M - \lambda\cdot A\cdot[\partial(C/A^U)/\partial M]$

(B.2 d) $\partial H/\partial q = 0$ => $c_q\cdot A = \lambda\cdot(\partial C/\partial q) + \lambda\cdot[(-\partial\phi/\partial q) - \partial(C/A^U)/\partial q]\cdot A$

(B.2 e) $\partial H/\partial(\lambda\cdot e^{-rt}) = dA/dt$ => $dA/dt = C(p,M,q) - A\cdot\{\phi(p,q) + C/A^U\}$

In a steady state, $d\ln\lambda/dt = dA/dt = 0$. From (B.2 a) we obtain the value of a client:

(B.3) $\lambda = (p - c_a - c_q\cdot q)/(r + \phi + C/A^U)$

Developing the derivative $\partial(C/A^U)/\partial p$ in (B.2 b), replacing the definitions of price elasticities given in the text, replacing the value of a client given by (B.3) and using the steady state condition in (B.2 e) plus the definition of participation $\alpha = A/A^U$ to find that $C/A^U = \phi\cdot\alpha/(1-\alpha)$, the result is that (B.2 b) can be presented as equation (1') in the text. The analogous procedure yields equations (') and (6') in the text.

Appendix 3: First order conditions of to optimization programs (D) and (E)

Dockner at al (2000) in Theorem 4.2, condition (ii), show that in differential games the Nash equilibrium the adjoint equation analogous to (B.2 a) determines $\lambda(t)$ as a function of the derivative of the maximized Hamiltonian with respect to the firm's state (which is A). This condition is:

(C.1a) $$\frac{dl}{dt} = r\cdot l - \frac{\partial}{\partial A}\left(H^*_{Z_{-i}}(A, A_{-i}, l, t)\right)$$, which simplifies in the steady state to

(C.1b) $$l = \frac{1}{r} \cdot \frac{\partial}{\partial A}\left(H^*_{Z_{-i}}(A, l)\right)$$

where Z_{-i} = (Z_{-i}, X_{-i}, W_{-i}) in the notation of equation (14) in the main text = the rivals' actions, expressed a a function of time, the state of the game (A,A_{-i}) and possibly λ.

Equation (C.1) determines λ(t), the "value of a client", which is needed to obtain the rule that is analogous to (1) in the main text.

The maximized Hamiltonian differs from the ordinary Hamiltonian in that the controls (p,M,q) are replaced by their optimal values [p*(A,A_{-i}), M*(A,A_{-i}), q*(A,A_{-i})]. Replacing the rival's actions as well, the maximized Hamiltonian is:

(C.2) $$H^*_{Z\text{-}i}(A, A_{-i},\lambda,t) = e^{-rt}\cdot\{ (p^*(A,A_{-i})-c_a)\cdot A - F - c_m\cdot M^*(A,A_{-i}) - c_q\cdot q^*(A,A_{-i})\cdot A$$
$$+ \lambda\cdot[C(p^*(A,A_{-i}),M^*(A,A_{-i}),q^*(A,A_{-i}), Z_{-i}, W_{-i}, X_{-i}, A_{-i})$$
$$- A\cdot\{\phi(p^*(A,A_{-i}),q^*(A,A_{-i}), Z_{-i}, W_{-i}, X_{-i})$$
$$+ C/A^U(s(A)\cdot p^*(A,A_{-i})+(1-s(A))\cdot Z_{-i}), p^*(A,A_{-i}),M^*(A,A_{-i})+ W_{-i}, s(A)\cdot q^*(A,A_{-i})+(1-s(A))\cdot X_{-i})\}]\}$$

where: s = s(A) = A / (A+ A_{-i}) = market share of the firm.

It can be seen in (C.2) that the terms in ∂H*/∂A may be separated in four sets. The first set brings together the terms that appear in the pure monopoly case (for s =1) when the strategies (p*, M*,q*) of the firm and the market share s are taken as given. This is the set of terms that appears in the ordinary Hamiltonian, such as (B.2 a). The second set arises from the fact that the market share s = s(A) depends of state A when A_{-i} is given, a situation that does not occur in pure monopoly. The third set arises from the dependence of the strategies (p*, M*,q*) on A, and it appears also in the pure monopoly case. The fourth set arises from the possible dependence of actions of rivals (Z,W,X) on the firm's state A.

In the full commitment case the third and fourth set of terms vanish, because due to full commitment, both the firm's and rivals' actions are a function of time alone. This means that the strategies (p*(t), M*(t),q*(t)) do not depend on the firm's state A, and that the functions (13) simplify to Z(t), W(t) and X(t) and do not depend either on the firm's state A. This situation means that the maximized Hamiltonian coincides with the ordinary Hamiltonian.

Thus, the only difference between the full commitment case and the pure monopoly case is given by the second set of factors, which reflects the fact that the market share s = s(A) depends of state A when A_{-i} is given. Thus, the "value of a client" in the no commitment case is slightly different, for this reason, from the pure monopoly case. This implies that the first order conditions for program (D) take a form that is similar to that for program (C), as shown by equation (12) in the main text.

In the no commitment case, on the other hand, the third and fourth set of terms do not vanish. Now, the firm's strategies are expressed as a function of its state and the rivals' actions depend as well of this firm's state through term A in the contact function for each rival (this term is symmetric to term A_{-i} in the contact function C(p,M,q, Z_{-i}, W_{-i}, X_{-i}, A_{-i}) for this firm). Thus, the "value of a client" is a more complex expression for the no commitment case. We conclude that the first order conditions for program (E) take a form that is quite distinct from those for program (D).

The first order conditions in the full commitment case

Consider program (D) for a large firm that is not a pure monopoly. Assuming that the third and fourth restrictions in program (D) are slack, the Hamiltonian is:

(C.3) $$H = e^{-rt}\cdot\{ (p-c_a)\cdot A - F - c_m\cdot M - c_q\cdot q\cdot A + \lambda\cdot[C(p,M,q) - A\cdot\{\phi(p,q) + C(p,M,q)/A^U(s(A),p,M,q) \}] \}$$

Note that the market share s(A) appears as an argument of A^U. Note also that the consumer response functions C, ϕ and A^U have different levels from the analogous functions for program (C), because they are influenced by rivals. From the Maximum Principle, the first order conditions for an interior optimum are the same as (B.2), except for the first one:

(C.5 a) $\partial H/\partial A = -d(\lambda \cdot e^{-rt})/dt \quad \Rightarrow \quad (p - c_a - c_q \cdot q)/\lambda - \phi - C/A^U(1 - \alpha \cdot [\partial A^U/\partial s] \cdot [\partial s/\partial A]) + d\ln\lambda/dt = r$

where $\partial s/\partial A = s \cdot (1-s)/A$; $\quad \partial A^U/\partial s = [\partial A^U/\partial q] \cdot (q - q_{-i}) + [\partial A^U/\partial p] \cdot (p - p_{-i})$

and $s = A/(A + A_{-i})$ = market share < participation = $\alpha = A/A^U$

Equation (C.6) shows that the firm with a market share below 100% exhibits a new term in the equation that sets the value of a client. Using the definitions of elasticities given in the text, α drops out and:

(C.6) $\lambda = (p - c_a - c_q \cdot q)/[\, r + \phi + C/A^U \cdot \{1 + s \cdot (1-s) \cdot K\}]$

where $K = (\eta^{AU,p})(p - p_{-i})/p - (\eta^{AU,q})(q - q_{-i})/q$

As the other first order conditions for an interior optimum are the same as (B.2), the procedure described in Appendix 2 leads to an analogue of equation (1) for the steady state, where the effective price elasticity is:

(C.7) $$h^{eff} = \left[(1-a) \cdot (h^{C,p} + h^{f,p}) + a \cdot h^{AU,p}\right] \cdot \left(\frac{f}{r \cdot (1-a) + f \cdot [1 + a \cdot s \cdot (1-s) \cdot K]} \right)$$

Analogous adjustments to the denominator in equa tions (5') and (6') yield the appropriate Dorfman-Steiner conditions for a large firm in the full commitment case.