BRAND LOYALTY AND MARKET EQUILIBRIUM

BIRGER WERNERFELT
Massachusetts Institute of Technology

Two concepts of brand loyalty are defined, "inertial" brand loyalty resulting from time lags in awareness, and "cost-based" brand loyalty resulting from intertemporal utility effects. Their market level implications are formally derived in a continuous time model. It is found that inertial brand loyalty leads to equilibria with price dispersion, while cost-based brand loyalty also may allow single price equilibria. In all cases, as brand loyalty vanishes, so does the difference between the average trading price and the price which obtains with no brand loyalty. Consistent with empirical results, the theory predicts that the relationship between market share and performance is positive in cross-sectional studies, but flat in time-series studies. The theory is also consistent with the view that market share is an asset in itself. After developing the theory, several strategic implications are drawn. In the end, some questions for further theoretical and empirical research are raised.

(Brand Loyalty; Game Theory; Marketing Strategy)

1. Introduction

Brand loyalty is a fundamental concept in strategic marketing. It is generally recognized as an asset (Aaker 1984, p. 140), the reason being that it increases pricing flexibility (Staudt, Taylor and Bowersox 1976, pp. 140–141). There may be some debate over the extent of the phenomenon, but the consensus is clearly that loyalty is a concept of major importance (Jacoby and Chestnut 1978).

Given this, it is not surprising that a lot of consumer behavior literature is concerned with the sources of loyalty and the mechanisms through which it comes about (Kuehn 1962 is an early such paper). Similarly, there is a market level literature on how a firm, primarily through advertising, can foster brand loyalty (see, e.g., Day 1984, p. 105). And yet, at the market level, the cornerstone in our knowledge is missing. Conceptually the simplest and most basic problem is: "How can a firm exploit brand loyalty once it has it?" This question is not answered in the literature. The literature on strategic pricing (e.g., Dolan and Jeuland 1981) goes part of the way, but is typically confined to a monopolistic setting. The meaning of brand loyalty for such firms is clearly very different than that for firms facing competitors. With rational or learning competitors only those strategies which form a market equilibrium can be sustained. Accordingly, it is crucial that a game theoretic perspective is adopted.

The purpose of this paper is to examine the market level implications to two types of brand loyalty. More specifically, one type of brand loyalty is due to time lags in awareness, and the other is due to costs of switching. Brand loyalty of the first type is called "inertial" and brand loyalty of the second type is called "cost-based".

229

0732-2399/91/1003/0229$01.25
Copyright © 1991, The Institute of Management Sciences/Operations Research Society of America
It is shown that markets with inertial brand loyalty have no equilibria without price dispersion. Further, we describe an equilibrium in which all brands have time invariant, but different, prices. In this equilibrium, firms with larger market shares charge lower prices, but earn higher profits. Firms with lower market shares charge higher prices and find it unattractive to try to gain share, because this would require them to charge lower prices for some period of time.

In markets with cost-based brand loyalty, there are again equilibria without price dispersion (although the game also has equilibria with price dispersion).

The most interesting results are those which characterize the equilibria with a distribution of prices and market shares. The properties of these equilibria are consistent with (i) the empirical literature on market share and performance, and (ii) the belief that market share is an asset in itself.

(i) The empirical literature documents a strong cross-sectional correlation between market share and performance (Gale 1972; Buzzell, Gale and Sultan 1975), along with a negligible time-series correlation (Jacobson and Aaker 1985; Rumelt and Wensley 1981; Montgomery and Wernerfelt 1991). While its ability to explain the empirical evidence is an attractive feature of the theory put forward in this paper, it is important to be aware that several other theories can explain the same “stylized facts”. The theory suggested here is demand-based and independent of product differentiation. An alternative demand-based theory, due to Prescott and Visscher (1977), is that firms dominate segments of different sizes. The argument in supply-based theories is that different firms have different cost structures, such that some find it profitable to produce more (Demsetz 1973; Jacobson 1988).

These theories all suggest positive cross-sectional and flat time-series relationships between market share and performance. The theories are not mutually inconsistent, so discriminating empirical tests are not necessarily in order. However, relative importance can be assessed from the share-margin correlation, which is positive in the presence of economies of scale, but negative if only demand-based forces operate. In the aggregate, the evidence clearly favors the supply-based theories. Several studies (e.g., Buzzell, Gale, and Sultan 1975) report a positive correlation between margin and share.

This should not lead to the absurd conclusion that brand loyalty does not exist. Brand loyalty has been observed in numerous micro level studies and some possible market level implications have been found in the PIMS data by Fornell, Robinson, and Wernerfelt (1985). Most likely, both supply and demand factors operate simultaneously, in varying degrees, in different industries. Hence, it seems fair to conjecture that supply-based effects on the average dominate demand-based effects, but that the latter effects are important in specific industries.

(ii) The last point is indirectly supported by the widely held view that market share is an asset in itself. The supply-based theories mentioned above either portray market share as a result of an asset (a favorable cost position) or as a creator of an asset (production experience). Only in the brand loyalty mechanism developed in the present paper is market share an asset in itself. Since shares have to sum to one, it immediately follows that there are first mover advantages from getting it early. Conversely, later entrants face an entry barrier because they lack market share.

In sum, a large number of standard phenomena from the field of marketing strategy can be described and explained in the model presented below.

The paper is organized as follows. I define the two types of brand loyalty in § 2 and formally derive their market level implications in § 4. Because the model is very complicated, § 3 is devoted to its introduction and § 5 contains a discussion of the robustness of the results. Some strategic implications are highlighted in § 6 and suggestions for further theoretical and empirical research are given in § 7. A short conclusion, § 8, ends the paper.
2. Definitions

While the concept of brand loyalty has been used extensively in the marketing literature, the field has not reached consensus on a single definition. In fact, Jacoby and Kyner (1973, p. 1) assert that, "there are at least 8 major approaches to operationally defining brand loyalty", but "there are no conceptual definitions of brand loyalty". Recently, Colombo and Morrison (1989, p. 90) wrote that "only the researcher's imagination limits the number of plausible definitions for [brand loyalty]".

The tension between behavioral and cognitive approaches has been a contributing factor to the above problem. Until Day (1969) and Jacoby (1971), the field looked at brand loyalty in terms of outcomes (repeat purchase behavior), rather than reasons. The advantage of the cognitive approach is that one can distinguish between different mechanisms leading to repeat purchase behavior. In particular, one can differentiate between timeless qualities of preferences and truly dynamic effects. This involves a conceptual distinction between brand purchase due to fit between personal tastes and brand attributes and brand purchase due to past purchase. Operationally, this corresponds to a distinction between the unconditional choice probability and the incremental choice probability obtained by conditioning on past choice.

Despite the seemingly natural distinction between static and dynamic measures, the field appears to have adopted the view that "the best measure in any case is . . . situation specific" (Colombo and Morrison 1989, p. 90). In fact, the segmentation literature routinely defines brand loyalty in terms of high unconditional purchase probabilities (Grover and Srinivasan 1989).

The purpose of the present paper is to look at the implications of brand loyalty for market equilibria. Since the effects of heterogeneous preferences is well known, I here focus on the dynamic effects only. So I define a consumer as brand loyal if his purchasing pattern depends positively on the last brand purchased. In terms of dynamic choice models, this is then a markov effect. A consumer is brand loyal if his probability of buying a particular brand at time $t$, conditional on identical purchase at time $t-1$, is larger than the corresponding unconditional probability (Frank 1962; Massy 1966; Kahn and Meyer 1989).

To characterize equilibria we need to know how this dependency is affected by marketing variables. As a first cut, I here distinguish between pure awareness effects and utility effects. Specifically, I define two types of brand loyalty.

"Inertial brand loyalty": when brand utilities have no intertemporal dependence but consumers may be slow to become aware of the most attractive values.

"Cost-Based Brand Loyalty": when brand utilities have positive intertemporal interdependence, such that the brand last purchased has an advantage.

It is not the practice of the behavioral literature to define concepts in terms of utilities. However, in his classification of habitual purchasing behavior, Assael (1987) draws a very similar distinction. Specifically, he presents “brand loyalty” as a commitment due to favorable attitudes learned from past purchases. In contrast, “inertia” is defined as lack of search due to low involvement. So consumers who are “brand loyal” in Assael’s terms may search actively, but require a substantial price differential before they switch brands (1987, p. 57). On the other hand, consumers who are “inert” will exhibit very low search activity, but could well change in response to very small price differences once they become aware.

In sum, there are many definitions of brand loyalty in the literature. Among many possible distinctions, the definitions can be classified as behavioral versus cognitive or static versus dynamic. My definitions are cognitive and dynamic. Within this class of definitions, my term, “inertial brand loyalty,” roughly corresponds to Assael’s (1987) concept of “inertia” in purchasing behavior for low involvement products. Similarly,
my term, "cost-based brand loyalty," roughly corresponds to his concept of "brand loyalty" as supported by learned positive attitudes. Because my definitions are tied to a mathematical model they are much more precise, but also less rich than many definitions in the literature. Let me therefore discuss the sources of inertial and cost-based brand loyalty.

Following Assael (1987), inertial brand loyalty can be thought of as a result of "low involvement". Low involvement manifests itself as an absence of active search for, and evaluation of, members of the product class. In the context of rational utility maximization, such behavior results from high search costs combined with a belief that the rewards of search, in terms of price and quality, are small. These conditions are probably most closely satisfied for low price, frequently purchased, mature consumer goods. In fact, Assael (1987, p. 82) claims that "most purchase decisions are low in consumer involvement". According to Laurent and Kapferer (1985), low involvement is more likely when products are low in perceived importance, risk, and symbolic or hedonic value.

Cost-based brand loyalty may have many causes. First, it can be the result of brand-specific user skills (Stigler and Becker 1977; Wernerfelt 1985). That is, consumers may learn-by-using their current brand. Such a phenomenon certainly seems plausible for computer software (e.g., Lotus vs. Excel), but also for other products and services one must expect a small effect of this type. Because this learning does not transfer perfectly to other brands, some small switching costs result.

Second, we can have purely informational effects. You know the quality of your current brand, so why take a risk on changing? (Schmalensee 1982). This effect may be quite small in the sense that a minor price differential will overcome it.

Third, there may be behavioral effects relating to the dynamics of utilities. Preferences may form around current brands, such that other brands will seem less attractive (Carpenter and Nakamoto 1988).

Fourth, compatibility problems may mean that brand switching will cause the consumer to (at least partly) lose other investments which have to be made in order to use the product. Examples include such factors as training costs, quality checks and complementary hardware. (Broadly construed, this category includes the user skill story above.)

The switching cost argument should apply best to less frequently purchased consumer goods and industrial products. However, as argued in the first three points above, some effects of this type may affect frequently purchased consumer goods as well.

It is important to mention that the following only pertains to brand loyalty resulting from natural properties of the product class. Endogenous brand loyalty, induced by, e.g., advertising or pricing schemes, is not covered. This lack of coverage is not motivated by a judgment about relative importance. I believe that firms in many cases can induce brand loyalty. However, an analysis of that case is different from, and more difficult than, the one presented here.

I will now look at a formal model of the market level implications of the two types of brand loyalty.

3. Introduction to the Model

Because the mathematical formalization is very complicated, it is useful to start with a less formal introduction to the model.

The art of model building consists of stripping a situation of all nonessential complexity in order to analyze the key driving components in the most transparent manner possible. In the present case, the essential problem facing firms is how to trade off the desire to gain new customers against the benefits from charging high prices to current customers. Since this is an inherently dynamic problem, some complexity is unavoidable. In order
to minimize this complexity, I will try to formulate a model in which the trade off is time invariant. This is accomplished by postulating constant turnover in the pool of customers such that uninformed new customers continue to replace well-informed old ones. So even though all customers continue to learn, the average level of information is constant.

I next make a number of assumptions whose function is to ensure the dynamic consistency of the model. In particular, I assume that imperfectly informed customers and firms make decisions based on (probabilistic) beliefs which are true in the aggregate. For example, while customers do not know which firms charge which prices, their beliefs about the overall price distribution are correct. In a time-invariant equilibrium these assumptions imply that the players are not systematically fooled. I feel that these are reasonable assumptions for studying the phenomenon of brand loyalty.

The final set of assumptions are chosen to enhance the “smoothness” and thus the analytical tractability of the model. For example, by assuming that there are infinitely many firms and customers, some aggregate distribution functions are rendered differentiable. Similarly, concavity assumptions ensure that first-order conditions are sufficient.

Several “realistic” complications such as advertising, heterogeneity on either side of the market, price discrimination etc., are not addressed explicitly. This is a deliberate attempt to exclude phenomena which are not necessary for the results. So when we go to apply the model, we know that the results do not depend on these complications. As such, my modelling philosophy is analogous to the decision calculus rule of modelling (only) phenomena important for the problem at hand (Little 1970).

Naturally, if the analysis is to be applied as a decision support tool, then we must investigate the impact of relaxing some assumptions. For example, the retail environment for packaged goods is generally such that both firms and customers are heterogeneous. Similarly, many markets feature extensive advertising and real markets will be grainy (finite firms, finite customers). I examine some generalizations in § 5 and Appendix 1 after presentation of the basic models.

4. Market Level Implications

a. Inertial Brand Loyalty

Let us think about an existing (as opposed to new) nondurable consumer product that is sold in a market with many consumers and many firms each of which sells one brand. To keep a certain level of ignorance in the market, assume that there is some turnover among the consumers due to changing needs and geographical mobility. For simplicity, we abstract from production costs and assume that the products are essentially identical, such that perfectly informed consumers would prefer the lowest priced brand.

Formally, we look at a continuum of consumers¹ who are born and die with intensity \( \lambda > 0 \), while a continuum of firms exists in perpetuity. The assumption of infinitely many consumers and firms greatly simplifies the analysis. One should think of these assumptions as convenient ways to approximate markets with many buyers and sellers. (See further in § 5.) I assume that there are far fewer firms than consumers in the sense that their measure, \( n \), is significantly less than the measure of consumers.² The utility experienced by a consumer (say, \( i \)) per unit of time is given by the function \( U(1 - p_i y_i) \), where \( p_i \) is the price charged by \( i \)'s current supplier at time \( t \), and \( y_i \) is the rate of consumption. \( U: R \times R_+ \rightarrow R \) is thrice continuously differentiable and has sufficient

¹ Normalized to unit measure.
² That is, arrivals and departures follow a Poisson process with parameter \( \lambda \).
³ So each firm can be thought of as selling to a continuum of consumers.
curvature to make the optimal consumption rates \( y^*(p) \) declining functions for which monopoly price \( p_m \) is finite.\(^4\) For notational simplicity, we will generally operate with \( u: R_+ \rightarrow R \), the indirect utility function of price.\(^5\) The above assumptions imply that \( u \) is twice continuously differentiable.

When they first enter the market, consumers are ignorant about prices of particular brands and each consumer therefore starts buying a randomly chosen brand. Brand switching becomes possible once the consumer becomes aware of the prices of other brands. We assume that this occurs with intensity \( \lambda > 0.\)\(^6\) To capture the essence of inertial brand loyalty, it is assumed that, once the consumer is aware, there is no resistance to switching to a lower priced brand. On the other hand, it may be reasonable to think of \( \lambda \) as reflecting “accidental” awareness, from, e.g., word-of-mouth, rather than deliberate search. The fact that \( \lambda \) is finite reflects inertial brand loyalty in the sense that a firm can charge a higher price for a while before its customers switch. This effect has been called “dynamic monopoly power” by Arrow (1959).

Let \( \omega \) be a realization of the stochastic processes described above.\(^7\) Aiming to maximize expected lifetime utility, consumers find switching strategies as (measurable) functions of time, the price of the brand they most recently became aware of \([p(t, \omega)]\), the price of their current brand \([p_i(t, \omega)]\) and the marginal distribution of brands over prices \([G_i]\).\(^8\) By restricting the domain of the strategies to current prices, we rule out the influence of past prices documented in recent work by Winer (1986) and Lattin and Bucklin (1989). Such phenomena would not necessarily destroy the equilibria studied here, but the analysis is difficult and awaits future research.

In the following, we will define a steady-state equilibrium, as one in which firms charge constant prices. In such equilibria, the consumer’s problem is relatively simple. Specifically, the optimal switching strategies are of the reservation price type and consumer \( i \) is looking for a real-valued function \( \delta^*_i(p^0, p_i, G, t, \omega) \) to maximize the expectation of expected lifetime utility, discounted at \( r > 0 \):

\[
\int_0^\infty e^{-(r+\gamma)t} u[p_i(t, \omega)] \, dt
\]

(1)

given

\[
dp_i(t, \omega) = dM_i[p_i(t, \omega)|\delta_i(\cdot)]
\]

(2)

and a set of initial conditions, where \( dM_i \) is a stochastic process defined to ensure that \( p_i \) at all times equals the value of the price of consumer \( i \)'s current brand.\(^9\)

It is easy to see (by contradiction) that \( \delta^*_i(\cdot) = p_i \). Because different consumers buy different brands who in turn may charge different prices, this induces a marginal distribution of reservation prices which we call \( K \).

Turning now to the firms, it is reasonable to assume that they cannot price discriminate among consumers. We further assume that they have good aggregate information such that they can set prices as (measurable) functions of the marginal distribution of other

\(^4\) The technical condition for this is \( 2y^{**} + py^{*'} < 0 \), where ‘ denotes the first derivative and ” denotes the second derivative.

\(^5\) This is obtained by substituting \( y^*(p) \) into \( U(\cdot) \), such that we get a function of \( p \) only.

\(^6\) So changes in awareness are modelled as following a Poisson process with intensity \( \lambda \), where an “arrival” implies that the consumer becomes aware of the price of a particular brand.

\(^7\) That is, \( \omega \) is a particular infinite history, \( t \in [0, \infty), \) of all consumers’ arrival and departure patterns (the \( \tau \) process) and all consumers’ awareness patterns (the \( \lambda \) process). So contingent on \( \omega \), everything is known at all times.

\(^8\) This implies that consumers understand the game well enough to deduce (or guess) the relative number of brands priced at each level. Such a “rational expectations” assumption is standard in game theory models.

\(^9\) This implies that \( dM_i = p_i^0 - p_i \) when consumer \( i \) switches and zero at all other times.
brands over prices \([G]\), the marginal distribution of reservation prices \([K]\), their own market shares, and time. Denoting the market share of firm \(j\) by \(b_j(t, \omega)\), we look for real-valued functions \(p^*_j(b, G, K, t, \omega)\) to maximize the expectation of discounted profits

\[
\int_0^\infty e^{-\gamma t} y^*(p_j) p_j b_j(t, \omega) dt
\]

subject to a dynamic constraint, which specifies how market shares change over time as a result of (a) consumers leaving and entering the market, (b) consumers switching to the brand, and (c) consumers switching from the brand to lower priced alternatives. The three terms on the right-hand side of (4) depict these three effects.

\[
db_j(\cdot)/dt = \tau \{1/n - b_j(\cdot)\} + \lambda(1 - K(p_j))(1/n) - \lambda b_j(\cdot)G(p_j).
\]

The problem description is completed by the initial conditions.

Turning now to equilibrium concepts, the following definition is natural:

**Definition 1.** A steady-state equilibrium is a pair consisting of a distribution of posted prices and a distribution of consumer states (the price they currently pay) such that:

(a) the distribution of consumer states is in a steady state,
(b) any posted price in the support of the firms' distribution is optimal given the two distributions.

Given this, we can show that a (relatively) well-behaved steady-state equilibrium exists.

**Proposition 1.** In any steady-state equilibrium, the distribution function over prices, \(G\), is strictly increasing and continuous over a noninjinitesimal domain.\(^{10,11}\)

**Proof.** See Appendix 2.

The key implication of Proposition 1 is that a single price equilibrium does not exist. To see why, imagine first that all but one firm charge the price \(p > 0\). The firm in question can then profit from charging a bit below \(p\). Conversely, if \(p\) is zero, a firm can profit from charging a bit above zero.

**Proposition 2.** There exists an asymmetric steady-state equilibrium.\(^{12}\)

**Proof.** See Appendix 2.

To investigate the effects of inertial brand loyalty, we ask what happens when the effect vanishes in the sense that consumers become aware of alternative brands infinitely often. That is, we look at \(\lambda \to \infty\). Fortunately, the model is well behaved. If \(\tilde{K}\) is a steady state:

**Corollary 1.** For all \(\epsilon_1, \epsilon_2 > 0\) there exists a \(\lambda\) such that for all \(\lambda > \tilde{\lambda}\);

The steady-state distribution of prices satisfy \(\tilde{K}(\epsilon_i) > 1 - \epsilon_2\).

**Proof.** See Appendix 2.

This means that markets with small amounts of inertial brand loyalty (high values of \(\lambda\)) look almost like markets with no brand loyalty. So even though the price distribution remains nondegenerate, eventually almost all trading takes place at prices very close to marginal costs (here zero), the price when there is no brand loyalty. It is further true that firms make positive profits in this equilibrium (because of the discounting), and that these profits behave as expected when competitive pressures, the information level in the market, and initial market shares are varied.

\(^{10}\) Formally, \(G\) has connected support and no mass points.

\(^{11}\) For a finite number of firms, the result is no longer true because \(G\) and \(K\) then have mass points. (See, e.g., Wernerfelt 1984 and Appendix 1.) However, if one thinks of the information of the players as being a little bit "noisy", the subjective distributions \(G\) and \(K\) may once again become well behaved. See further in § 5.

\(^{12}\) Propositions 1 and 2 are proved by Wernerfelt (1988b) in general equilibrium versions.
COROLLARY 2. Profits in asymmetric steady-state equilibria are higher for slower awareness processes, faster turnover, and higher market share.

PROOF. See Appendix 2.

b. Cost-Based Brand Loyalty

To capture the essence of one type of cost-based brand loyalty, we now allow consumers to accumulate seller specific user skills over their tenure with a particular seller. (To avoid complications arising from endogenous switching costs, we assume that this process is automatic.) So the indirect, instantaneous utility function of consumer \( i \) now has the form \( u(p_{it}, z_{it}) \), where \( z_{it} \) is the age of current trading relationship. As before, we assume that \( u(\cdot) \) is twice continuously differentiable, increasing, and concave in \( z \geq 0 \), and yields a finite \( p_m \) for all \( z \).

We make analog assumptions on the information structure of this game, except that we allow the switching strategies to depend on the age of the trading relationships. In steady-state equilibrium, therefore, consumer \( i \) is looking for a switching strategy as a (measurable) function of \( p_0, p_i, G, z_i \) and time. The objective is to maximize

\[
\int_0^\infty e^{-(r+\gamma)t} u[p_i(t, \omega), z_i(t, \omega)]dt
\]

given

\[
dz_i(t, \omega) = dt + dN_i[z_i(t, \omega)|\delta_i(\cdot)],
\]

\[
dp_i(t, \omega) = dM[p_i(t, \omega)|\delta_i(\cdot)],
\]

and a set of initial conditions, where \( dN_i \) is a stochastic process defined to ensure that \( z_{it} \) jumps back to zero if the consumer switches. To capture the difference from the low involvement case of inertial brand loyalty, one should here think of \( \lambda \) as a result of active search.

Since this is formulated as a so-called piecewise deterministic Markov process (Davis 1984), it follows from Vermes (1985, Theorem 1) that a solution to (5)-(7) exists. The optimal \( \delta^*(\cdot) \) is of the reservation price type. Further \( \delta^*(p, \cdot) \) declines continuously from \( p_i \) as \( z_i \) grows above zero.

We assume that firms know the joint distribution of consumers over prices and reservation prices \( K(p, \delta) \). In this case the firm is looking for a measurable real-valued function \( p_f(b, G, K, t, \omega) \) to maximize

\[
\int_0^\infty e^{-\gamma t} y(p_f)b(t, \omega)dt
\]

13 The only significant assumptions concern the decreasing returns to scale in \( z \) and the differentiability at \( z = 0 \).

14 So \( dN_i = -z_i \) when the consumer switches, and zero otherwise.

15 It is possible to generalize the model further by making the search intensity a decision variable. Suppose there is a suitable utility cost of searching faster than \( \lambda^0 > 0 \), and consumers seek measurable functions \( \lambda^*(p_0, p_i, G, z_i, t, \omega) \). In this case, Jacod and Protter (1982) and Protter (1983), after I asked them for help, proved the existence of a solution to (7) in which the intensity now is state dependent.

16 Vermes operates with a finite-dimensional state space, so I am cutting a corner here. One should think of a large but finite population of firms and consumers.

17 To see that \( \delta^* \) is of the reservation price type, assume that \( \delta \) specifies switching for \( p^0 < p^1 \) or \( p^2 < p^0 < p^3 \) where \( p^1 < p^2 < p^3 \). Since the consumer has no impact on the market, \( \delta \) is obviously dominated by either "[\( p^0 < p^1 \)]" or "[\( p^0 < p^2 \)]".

18 Since the consumer's position immediately after switching is independent of his position immediately before, the desire to switch is lower if the present position is stronger. A higher \( z_i \) makes the present position stronger.
subject to initial conditions and the market share dynamics (analogous to (4)),
\[
db_j(\cdot)/dt = \tau[1/n - b_j(\cdot)] + \lambda \int_p^\infty [1 - k_p(x, p)] dx(1/n) - \lambda b_j(\cdot) \int_0^p [1 - k_\delta(p, x)] G(x) dx
\]
where \(k_p\) and \(k_\delta\) are the marginal densities of \(K\) on \(p\) and \(\delta\), respectively.

The asymmetric equilibrium described in § 4.a continues to exist in this model. However, we can now also get a much simpler, symmetric equilibrium. Let me start by defining the relevant equilibrium concept.

**DEFINITION 2.** A single-price steady-state equilibrium is a price \(p\) and a distribution \(\delta\) of consumer states (the length of staying at their current firm) such that:

(a) the consumers' distribution is in a steady state, and

(b) no firm will do better by altering its posted price.

Given this, we can show that a single price equilibrium exists,

**PROPOSITION 3.** A steady-state single-price equilibrium exists.\(^{19,20}\)

**PROOF.** See Appendix 2.

The reason that we get existence in this case is that the discontinuities from the earlier model disappear once a distribution of \(z\)'s is added to the picture.

The equilibrium price is defined by the first-order condition
\[
py' + y = py(2\lambda/(r + \tau))k
\]
where \(k\) is \(\partial K(p, \delta)/\partial \delta\big|_{\delta = p}\). So as in other switching cost models only the marginal consumers matter. We further find that the price behaves as expected when patience, the level of information in market, the competitive pressures, and market growth change.

**COROLLARY 3.** The symmetric equilibrium price and profits are higher for slower search, faster turnover, higher discount rates, and lower exponential market growth.

**PROOF.** See Appendix 2.

Also for this type of brand loyalty, the equilibrium approaches that with no brand loyalty as the magnitude of brand loyalty vanishes:

**COROLLARY 4.** The symmetric equilibrium price goes to zero as \(\lambda \to \infty\) or \(\partial u/\partial z \to 0\).

**PROOF.** See Appendix 2.

Note further that (10) gives the monopoly price as brand loyalty becomes very large (that is as \(\lambda \to 0\) or \(\partial u/\partial z \to \infty\)).

In Appendix 3, we look at an example which may illustrate the results better than the general derivations above.

\(c.\) **Comparison of the Models**

It is useful to consider the mechanisms underlying the differences between the equilibria of the two models. Specifically, why can a single-price equilibrium be sustained in the

\(^{19}\) It is important that \(u(\cdot, \cdot)\) be smooth at \(z = 0\). If there is a fixed component to switching costs, the discontinuities reappear. Following Shilonly (1977), one could conjecture that this can give mixed strategy equilibria, but it is unclear how one defines such strategies in differential games. Of course, a small amount of noise may smooth the aggregate discontinuities and eliminate this problem. See further in § 5.

\(^{20}\) In addition to this and the asymmetric equilibrium, I suspect that periodic (Gilbert 1977) or chaotic equilibria may exist as well.
case of cost-based brand loyalty only? The reason is that cost-based brand loyalty induces heterogeneous tastes, while inertial brand loyalty only induces heterogeneous information. The assumption that customers know the aggregate price distribution means that heterogeneous information disappears in single-price equilibria. So in order to make the profit function differentiable (check the incentives to undercut), and thus sustain an equilibrium, the single-price scenario requires suitably heterogeneous tastes. In the case of inertial brand loyalty, tastes are homogeneous, such that the price dispersion is needed for the profit functions to be well behaved.

5. Robustness of the Results

Let us examine some critical assumptions of the model.

a. Infinitely many firms and consumers. In a model with finitely many firms, the profit function of an individual firm is ill behaved. As the price drops below the price of a competitor, there is a discrete change in the rate of customer inflow. Formally, the right-hand side of (4) is not differentiable and the best response functions are discontinuous. In static games these properties typically lead to mixed strategy equilibria. However, this concept is not even defined in differential games and the analysis is very difficult. Since all real markets have finitely many firms this seems to be a rather damaging problem.

In a real market with many firms, no individual firm knows the exact price of all other firms. Rather, it has some probabilistic beliefs about these prices and this probability distribution is presumably rather smooth. Given this, the expected result of a unilateral price change should again be smooth and we are back to best response functions which look essentially as those with infinitely many firms. That is, if we assume that firms are perfectly informed, fragmented markets are modelled better by the model with infinitely many firms than by the model with finitely many firms.

As can be seen from Appendix 1, if one is willing to proceed without an existence proof, the results are not very different for cost-based brand loyalty (compare equations (10) and (14) in Appendix 2).

b. Switching costs grow continuously from zero. If there is a fixed component to switching costs, say $1, a firm contemplating a unilateral price change from the symmetric equilibrium of Proposition 3 would be able to increase price by a full dollar before any sales were lost. Given this, other firms would increase their prices, etc. etc. etc. However, at some point prices would be sufficiently high that a firm could benefit from cutting price by a bit more than one dollar. So existence of simple equilibria may well be a serious problem. An equilibrium probably exists, but it is likely to be highly complex.

There may be fixed switching costs for individual consumers, but in reality products and tastes will often be a bit differentiated, such that any price change will have action implications for some consumers. Given this, continuity is restored.

c. Constant price equilibria. As inspection of the proof will show, it is very difficult to establish existence of the asymmetric equilibrium with constant prices (Proposition 2). I have not been able to show existence of equilibria with time varying prices. However, I suspect that they exist.

d. No advertising. The intensity of the awareness process, $\lambda$, was an exogenous parameter in the model. It seems natural to think of individual firms as advertising to expand brand-specific awareness. That is, if firm $j$ incurs advertising costs at the rate $a_{jt}$, then consumers would become aware of its brand at the rate $\lambda(a_{jt})$.

For well-behaved $\lambda(\cdot)$, I see no reason that this extension would change the qualitative nature of the equilibria analyzed here. In the case of cost-based brand loyalty, one should still find single-price equilibria in which now all firms advertise at equal rates. Concerning

---

BRAND LOYALTY AND MARKET EQUILIBRIUM

239

inertial brand loyalty, Mortensen (1985) has analyzed a model quite similar to the one proposed. While he does not show that asymmetric equilibria with advertising exist, he shows that no other equilibria can exist (in his model).

e. One brand per firm. As long as firms produce finitely many brands, interactions between them will vanish in a model with infinitely many brands. In the finite case, the analysis would be much more complicated, although I suspect that the qualitative results would remain unchanged.

f. Customers become aware of one brand at a time. For a given \( \lambda \), if customers become aware of more than one brand at a time, there will be more competition and thus more downward pressure on prices. In fact, if two brands are always coupled they will have to charge the same price. However, as long as brands are grouped (systematically or randomly) in lots of finite size, the qualitative nature of equilibria for infinitely many brands should be the same as that described in § 4.

6. Strategic Implications

Assuming now that brand loyalty is of some (perhaps even significant) practical importance, what are the strategic implications suggested by the models? The comparative static results are summarized in Corollaries 1, 2, 3 and 4. The main direct result of the models, stated in Corollaries 1 and 4, is:

1. More brand loyalty permits higher equilibrium prices.\(^{22}\) Because of the reduced intertemporal price sensitivity, firms with positive market share will be able to price above costs, even if there are infinitely many of them. This is not a surprising result: both casual empiricism and marketing sense supports a positive association between brand loyalty and price levels. Indeed, brand loyalty has much the effect of product differentiation, it makes the brands less close substitutes.

It is presumably this effect which has led marketing scholars to look at brand loyalty as an asset (Aaker 1984, p. 140), which one should invest to created (Day 1984, p. 105).

Other results, stated in Corollaries 2 and 3, are contingent on the presence of brand loyalty in the market.

2. Lower firm-average levels of market communication permits higher equilibrium prices. If the average consumer becomes aware of competing brands on a less frequent basis, the current supplier will have more of a monopoly position and can change higher prices. Of course, individual firms have incentives to increase their own communication with potential buyers. (So we have a prisoner's dilemma type problem.) Similarly, all consumers have incentives to search more, and all consumers benefit from the search efforts of each of them.

On a cross-sectional basis, the implication is that loyal consumers are more valuable if they are harder to communicate with and less likely to search for alternatives. An example of the former could be buyers in less developed countries, and an example of the latter could be people with a very high value of time.

3. Higher market growth leads to lower equilibrium prices. Growth means that a greater fraction of buyers are unattached and thus more price-sensitive. On the other hand these buyers will, by assumption, turn brand loyal later. Once growth levels off, this means that the higher prices can be charged from a greater buyer base. So the net long-term effect of growth is attractive, rather than unattractive, from a profit perspective.

4. Lower discount rates lead to lower equilibrium prices. When discount rates are low, firms "invest" more aggressively in future brand loyal customers. Since the market price

\(^{22}\) Note that this is a market effect. The average price level will be higher in markets with more brand loyalty. Within a given market, prices in our asymmetric equilibrium are such that firms with larger share (who have more customers loyal to them) will price lower.
level is determined by competition for the marginal (unattached) consumers, this will drive prices down.

It is important to recall that the model is analyzed in terms of real prices. So only real interest rates matter. However, even these have moved substantially over the last ten years. It is therefore not unreasonable to expect a macroeconomic effect on price levels in industries with brand loyalty.

7. Suggested Further Research

A. Theoretical

On the theoretical side, the model suggests several extensions with strong managerial appeal. It would be interesting to develop richer models in which one could evaluate and sharpen the following conjectures.

1. Brand loyalty gives rise to first mover advantages. If firms enter sequentially, the first firm will get customers who are not already loyal to another brand. The second firm will get some from the first and some unattached, the third firm will be slightly worse off, and so on, until the equilibrium number of firms have entered. This mechanism is, of course, the logic behind the well-known “penetration” pricing strategy (Wernerfelt 1986; Dolan and Jeuland 1981), according to which firms should price lower when market growth (new customers) is higher.

In a formal model of this, I conjecture that sequential entry, ceteris paribus, will result in an eventual ranking of market shares and profits which rather closely mirrors the order of entry. Such a result would conform nicely to the empirical findings of Urban, Carter, Gaskin and Mucha (1986). Similarly, as the times between entries go to zero, such that eventually all firms enter simultaneously, I would expect that the “price of market share” will be bid up until all long-run profits disappear (Posner 1975).

This analysis could perhaps be undertaken in the context of a monopolist facing a single entrant. The inter-entry time could be measured by the fraction of all potential buyers who patronize the monopolist at the time of entry.

2. Brand loyalty makes market share an entry barrier. Consider again the model sketched in point 1 above. The entrant would initially have zero market share and thus zero profits. Only through prolonged underpricing would it gain share and subsequent profits. So the entry decision depends on a tradeoff between these “penetration” costs and later profits. This issue is partially illuminated in Wernerfelt (1988b) where I look at a general equilibrium model with habitual brand loyalty and free entry. The no-entry equilibrium condition (Wernerfelt 1988b, equation (7)) contains just the above factors. However, since the above-mentioned model posits a continuum of firms (just as the model in the present paper) it does not allow one to vary market share in a meaningful way.

Marketing folklore certainly supports this conjecture. For example, the argument could be applied to the computer software industry.

3. Brand loyalty enhances the advantages of product differentiation. In the formal model of cost-based brand loyalty, I assumed that user skills were completely brand specific. In a richer model where there are degrees of partial specificity, we can look at product differentiation as affecting the degree of specificity. To the extent that such differentiation makes it more difficult to transfer, e.g., user skills and quality judgments to other brands, the magnitude of brand loyalty will go up as brands become more differentiated. I conjecture that this will enhance profitability.

Again here the computer software industry could be an example (e.g., IBM presumably thought in this way when the PS/2 was introduced).

4. Brand loyalty invites price discrimination. In the present paper, I assumed that price discrimination was impossible. As an extreme case, consider the possibility of perfect
price discrimination in the context of cost-based brand loyalty. That is, assume that firms can identify the switching costs of each consumer and offer individual prices with no fear of arbitrage. In such a model, firms will want to offer lower prices to new customers, since these are more price sensitive.

Also this conjecture seems consistent with casual observation. As an example, many newspapers, magazines, and homeowners insurance companies offer cut-rate deals to new subscribers.

B. Empirical

On the empirical side any tests of the model presupposes that one can:

5. Characterize those industries where brand loyalty is strongest. At the microlevel, it should be possible to answer this question by checking for first-order effects in individual brand switching matrices for groups of consumers. With only macrolevel data, identification is more difficult, primarily because of confounding effects of scale economies.

Ex ante, it is not clear which industries are most strongly characterized by brand loyalty. Most folklore about brand loyalty is derived from consumer markets. I suspect, but do not know, that the phenomenon is equally important in industrial markets and, especially, in service industries. Wherever small heterogeneities set brands apart, there should be a basis for these effects. However, without research, this is all guesswork.

Assuming that one has developed a measure of the extent of brand loyalty in different markets, (e.g. an averaged difference between conditional and unconditional choice probabilities) the strategic implications highlighted above translate into research hypotheses.

Similarly, the conjectures 1–4 above could be tested:

6. Direct tests of the theory. Specifically, across markets, the extent of brand loyalty should correlate with (1) average market profitability, (2) some measure of first mover advantages, (3) concentration ratios, (4) some measure of product differentiation, and (5) some measure of the extent of price discrimination. The main difficulty in carrying out this research is that it involves merging of micro- and market-level data. However, it does not appear impossible, and these issues are certainly important for the marketing community.

8. Conclusion

In the context of fragmented industries, I have characterized how market equilibria change in response to two different types of brand loyalty. For inertial brand loyalty, symmetric equilibria cannot exist, and for both types of brand loyalty, we can have asymmetric equilibria. The main contribution of the paper is to begin a rigorous analysis of the competitive implications of brand loyalty. While the analysis is complicated, it yields intuitively reasonable predictions concerning a range of very important issues. In particular, it predicts a positive cross-sectional and a flat time-series relationship between market share and performance.

From a more technical perspective, it is interesting to think about the reasons for price dispersion at work in this paper. Models which yield equilibria with price dispersion generally rely on heterogeneity in either preferences or information. In some models, this heterogeneity supports specialist (or segmentation) strategies (Salop and Stiglitz 1977) and in other models the profit function is affected such that mixed strategies are necessary (Varian 1980). In the present model, the heterogeneity is produced endogenously as the stochastic awareness pattern unfolds. So the content of the heterogeneity is determined simultaneously with the equilibrium. In models based on different search costs, two buyers still differ after identical search outcomes. This is not true in the present model: only the price dispersion itself generates the segments which support it as an equilibrium.
outcome. In my view, this simultaneity is very attractive, since it allows us to generate an endogenous basis for price dispersion.\footnote{This paper was received December 1, 1989 and has been with the author 4 months for 2 revisions. This paper was processed by Marcel Corstjens.}

Acknowledgements. Comments from Anne Coughlan, Joe Farrell, David Sappington, Naufel Vilcassim, an anonymous referee, and seminar participants at the American Economic Association Meetings, Massachusetts Institute of Technology, Stanford University, The University of California at Berkeley, UCLA, and the University of Pennsylvania are gratefully acknowledged. Responsibility for any deficiencies rests with the author.

Appendix 1: A Finite Number of Firms

With a finite number of firms the anonymous formulation of the game is no longer satisfactory (see Reinganum and Stokey 1985). The appropriate strategy space is for firms to seek measurable functions of the vector of market shares, $G, K, i$, and $\omega$. It does seem possible to prove existence of equilibrium in this model, especially in the case with smooth consumer learning. I have, however, not been able to do so.\footnote{In Wernerfelt (1988a), I have a verification result for the even more difficult case where consumers are finite also. The problem is that I have been unable to generate a set of candidate strategies.} Instead, I have the following:

**Proposition 4.** If steady-state single-price closed loop equilibria exist, at least one is stable.\footnote{Wernerfelt (1984) contains an analogous argument and demonstrates existence in the open loop case.}

**Proof.** See Appendix 2.

**Corollary 5.** Prices in stable closed loop equilibria are higher than those in open loop equilibria.\footnote{Inspection of (14) in Appendix 2 reveals that the open and closed loop equilibria converge as $n$ goes to infinity.}

**Proof.** See Appendix 2.

Intuitively, a price cut will decrease competitors' market shares and this will in turn lead them to cut prices, leaving overall incentives lower (much like conjectural variations). The opposite result pertains for quantity games. There, volume expansion will lead to contraction by competitors, thus making it more attractive (see Fershtman and Kamien 1982).

Appendix 2: Proofs

**Proof of Proposition 1.** (By contradiction) Since only the ordering of prices matters, firms would move up in any gaps in $G$, so unconnected support is not compatible with equilibrium. Further, suppose there is a mass point above zero, say at $p^0$. In this case, there exists a $p' \in (0, p^0)$ such that $p^*(\cdot) = p'$ yields higher net present value than $p^*(\cdot) = p^0$. Finally, if there is a mass point at zero, then $p^*(\cdot) = p^*$, $p^* \in R_+$, yields positive profits, while $p^*(\cdot) = 0$ yields zero net present value. Q.E.D.

**Proposition 2** is proved via a fixed point theorem in function space. While this is much more complex than the familiar Kakutani-type fixed point theorem, it is analogous.

**Proof of Proposition 2.** By the fixed point theorem of Fan (1952) and Glicksberg (1952). The object is to find a $K$ such that steady-state prices are identical to optimal prices for all market shares. Define $S$ as the space of integrable functions from $[0, p^0]$ to $R$, and let $S_C$ be the subset of $S$ which is continuous. Now take any member of $S$, say $k_0$, and define $K^0(p) = \int_0^{p_0} k_0(x) dx$. If we insert into (4), using $k(\cdot) = g(\cdot)nb(\cdot)$, we get the steady-state condition

$$\frac{1}{n} - b(\cdot) + \left(1 - K(\cdot) - b(\cdot) \int k(\lambda)(1/b(x)dx) \right)(1/n) = 0.$$  \hspace{1cm} (11)

This identifies a unique, monotonic $b(p|k^0) \in S_C$. The inverse of this function, $\hat{b}(b|k^0)$, is continuous in $k^0$.

Given $k \in S_C$, (3)–(4) is a control problem and standard results (e.g. Fleming and Rishel 1975, Chapter 3) tell us that a solution $p^*(b|k)$ exists and is continuous in $k$. At this point

$$\hat{p}(b) = p^*(b|k)$$  \hspace{1cm} (12)

is an ordinary differential equation in $k$ with solution $k^* \in S$. So we can define $C$ as the correspondence from $b(p)$ to $k^*$. Any fixed point of $C$–$b(p|k)$ thus identifies a $k$ for which the optimal actions of each player (a) keep the state of that player constant and (b) preserve the aggregate state distribution.

From the second-order conditions and the monotonicity of $b_p$ we have that $C$ is upper semicontinuous. It is straightforward to verify nonemptiness and that $C$ is closed and convex valued. To check that $C$–$b(p|k)$ maps
S into $S$, start in $S$ and apply $\hat{b}$ to get into $S_b$ from which $C$ maps into $S$ again. Since $S$ is convex, we are done. Q.E.D.

**Proof of Corollary 1.** From (11) we get

$$1 - K(p) = \tau n(\hat{b}(p) - 1/n)/\lambda + \hat{b}(p)G(p).$$

Since $\hat{b}(p) \rightarrow 0$ as $\lambda \rightarrow \infty$, we are done. Q.E.D.

**Proof of Corollary 2.** Obvious from the implicit function theorem. Q.E.D.

**Proof of Proposition 3.** By construction. Assume that all firms have market shares $1/n$ and charge $p^*$ $> 0$ indefinitely. In that case, (8)–(9) is a control problem. Because of the assumptions on $u(\cdot, z)$, we avoid the discontinuities from the proof of Proposition 1 and, instead, have sufficient curvature to satisfy the second-order conditions for existence. Since all firms face the same problem, a symmetric equilibrium exists. Q.E.D.

**Proof of Corollary 3.** Implicit differentiation of (10), thinking of $r$ as containing negative growth. Q.E.D.

**Proof of Corollary 4.** Obvious from (10). Q.E.D.

**Proof of Proposition 4.** If a closed loop equilibrium exists, the market shares form a dynamical system. From the first-order conditions we see that the vector field points in on the boundary of its support (one prices lower if one has no market share). So we can apply the Poincaré-Hopf theorem (Guillemin and Pollack 1974, p. 134) to show that the dynamical system has at least one stable point. Q.E.D.

**Proof of Corollary 5.** By direct calculation, the closed loop analogue of (10) is

$$p^* \frac{\partial y}{\partial p} + y(p^*) = \left( p^* y(p^*) + \frac{p^*}{n} \left( \frac{2\lambda^0 k(p^*)}{r + \tau} \right) \right) \left( \frac{2\lambda^0 k(p^*)}{r + \tau} \right),$$

where $\partial y/\partial b = p'$ has been added in several places. Since this is positive in a stable equilibrium, the result follows. Q.E.D.

**Appendix 3: Example**

Suppose $u(p, z) = u(p e^{-\alpha z})$, $\alpha > 0$. In this case $\hat{b}(p, z) = pe^{-\alpha z}$ such that consumers are myopic in "effective prices". Now define $k(\cdot)$ as the density of such prices. If $g = G'$, we can write the dynamics of $k(\cdot)$ as

$$k(x) = \tau [g(x) - k(x)] + \lambda^0 \int_0^x g(s) ds = [1 - \tau - \lambda^0 G(x)] \alpha x dk/dx.$$

In the equilibrium described above, the steady state equations for $k$ are

$$0 = \tau (1 - k(p)) + (1 - \tau)\alpha p dk/dx|_{p^*},$$

$$0 = -\tau k(x) + (1 - \tau)\alpha x dk/dx|_{x < p},$$

in which $k(\cdot)$ has to satisfy

$$\int_0^{p^*} k(x) dx = 1 \quad \text{and} \quad \lim_{x \rightarrow p^*} k(x) = k(p).$$

This gives $k(p) = 1 - 1/[\tau/(1 - \alpha) + 1 + \alpha^{-1}]$ such that (14) reduces to

$$\frac{1}{\tau/[1 - (1 - \alpha)] + 1} \left( p \frac{\partial y}{\partial p} + y \right) = y \frac{2\lambda^0}{r + \tau}$$

by implicit differentiation, this gives:

**Finding 1.** Faster learning gives higher prices.

**References**


