SPATIAL BRANDING AND THE THEORY OF RETAIL CHAINS

by

Bretton J. Isaacs

M.A., Simon Fraser University, 1989

B.A., University of Calgary, 1987

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department

of

Economics

SIMON FRASER UNIVERSITY

May 2000

0-612-62620-2

Abstract

This thesis presents a theory of retail chains. A retail chain is defined as a set of spatially distributed stores selling a common product and sharing a common brandname. The theory shows that economies of scale in production and transportation costs are necessary, but not sufficient, as an explanation of such chains, for these two factors alone cannot explain spatial branding—the sharing of a common brandname across a set of spatially distributed stores.

The theory explains spatial branding by treating geographic space differently from other product-attribute spaces and, in particular, by doing away with the traditional assumption that the consumer has a fixed most-preferred geographic location. This assumption contradicts the fact that consumers often travel for reasons exogenous to their demand for many goods. When they travel, they take their demand for such goods with them, finding spatial distribution convenient and spatial branding informative.

When a consumer tries a new brand she learns a little bit about its attributes. Such information is important; not for the purpose of risk reduction, for the consumer is risk neutral, but rather for the purpose of improving future decisions. However, even perfect information has zero value if it never gets used, and the consumer's travel pattern puts a constraint on her ability to use such information. Retail chains restore the consumer's incentive to try the brand by providing a large chain with stores located along primary travel routes. The better a chain's coverage of the consumer's travel pattern, the greater her incentive to try the brand for the greater the likelihood that she will find the information contained in that purchase useful in future purchase decisions.

It is shown that the value of information contained in a purchase from a brand is *convex* in the size of the brand's retail chain. In discrete geographic space, this convexity implies a minimum-informative scale for a retail chain: consumers will not try the brand until its chain is of minimum-informative size. In continuous geographic space, this convexity results in the probability of first purchase from the brand also being convex. As such, the local-market share enjoyed by each store in the brand's chain is increasing in the size of its chain. Further, the speed with which a brand's local market share converges to its long-run value is shown to be increasing in the size of the brand's chain. This result is of particular importance for it is entirely independent of the forward-looking nature of consumers—i.e. it holds even if consumers behave myopically. A direct implication of this result is that not all brands will want to be spatially branded, some will prefer to remain independent single-store operations. Finally, the thesis concludes by demonstrating the implications these results have for the theory of franchising.

Dedication

To Shelene and Edna,

and

to my mother and father

who taught me,

not with their words, but with their actions,

everyday.

Acknowledgments

Sincere thanks to Professors Eaton and Jones for their guidance and encouragement. A special thank-you is also owed to Professors Dow and Church, and to my friends Erik, Bobby, Zu, Kevin, and the Squamish Chum.

Contents

List of Tables

List of Figures

Chapter 1

Spatial Distribution and Retail Chains

1.1 Introduction to the Theory of Retail Chains

Retail chains abound. From Starbuck's, to McDonald's, to Holiday Inns, one is likely not too far from where you are. "Chains are one of the dominant forms of organization of our times, with a single chain often having hundreds, even thousands, of units operating under a common trademark in diverse locations. Yet despite their importance we know very little about how chains work."(Bradach 1998, p. 1)

Why do retail chains exist? Clearly, before we can explain the success of retail chains, we must explain their existence. The two are obviously connected: existence is a necessary condition for success. However, despite the amazing success of retail chains, explanations for their existence are far from complete. This thesis seeks an explanation for the existence of retail chains.

Retail chains exist in a variety of forms and sell a diverse set of products and services. As such, it would be surprising if any one theory could explain their existence in all of their forms, and no such claim is made here. Nonetheless, the theory put forward herein is relevant to all chains, albeit some more than others. Exactly which chains are best suited to the proffered explanation will become apparent as the theory is developed.

1.1.1 Retail Chains: A Definition

Just as existence is a necessary condition for success, definition is a necessary condition for explanation. Before something can be explained, it must first be defined. What attributes do retail chains possess, and what services do they provide, which other firms do not? What distinguishes retail chains? A retail chain is here formally defined as follows:[1]

Definition 1.1 ***A Retail Chain.*** *A retail chain is a set of spatially distributed stores selling a common product and sharing a common brandname.*

It might be added that retail chains typically locate along highways and other major thoroughfares, in airports and bus terminals, and many other locations where a mobile population of consumers can be found. However, the more an object is defined, the less there is to explain. So rather than considering locational-choice a defining attribute of retail chains, the present work will consider this an object of explanation.

The theory of retail chains is not only important in its own right, but also for the theory of franchising. Almost invariably, theoretical work on retail chains is centered around the franchise relationship. This is understandable, for franchising has both empirical and theoretical import. Empirically, franchised establishments account for

[1]Some authors (e.g. Gosh and McLafferty (1987)) prefer the term "outlet-network" to "chain". The latter will be used here because it is briefer and because the concept it denotes in common usage does not differ significantly from that of theory.

over forty percent of retail sales in North America.(Bradach 1998, p. 1) Theoretically, "[t]he franchise relation raises fundamental questions concerning the nature of the firm and the extent of its integration."(Caves and Murphy II 1976, p. 572)[2]

Yet franchising and retail chains should not be confused. Franchising is an organizational structure used by some retail chains.[3] Thus explanations for the existence and form of the franchise relationship are not explanations for the existence and form of retail chains. Many retail chains are not franchised at all, yet the two terms are often used interchangeably, as if they represented identical concepts. Obviously, any theory of franchising must be consistent with the theory of retail chains. Thus, by contributing to the theory of retail chains, it is hoped that this work can also contribute to the theory of franchising. A start along these lines is made in Chapter 7.

1.2 Explaining Spatial Distribution

1.2.1 Economies of Scale in Production

A common explanation for the existence of retail chains is that such chains exist to realize economies of scale in production.[4] Such explanations are to be expected, for retail chains are firms: they produce for resale. As such, economies of scale are necessary for their existence. In the absence of scale economies, small scale production

[2] The economic and business literature contains many explanations for both the existence and form of the franchise relationship. See Dnes (1996) for a review of the academic literature on franchising. For accounts of the historical developments of chain stores and franchising in the United States see Hollander and Omura (1989) and Dicke (1992), respectively.

[3] The franchise contract is esentially a fixed-term licencing agreement that entitles a franchisee to use the business format and other intellectual capital of the franchisor to distribute products and/or services in a manner specified in the contract. Franchising and its relationship to the present work is discussed in Chapter 7.

[4] "Economies of scale in production" is a phrase used to describe a situation where the cost per unit falls as the number of units produced per period increases.

would be efficient. Each consumer could produce the product efficiently for herself, so firms, including retail chains, simply need not exist.

There is little doubt that economies of scale are important to the success of retail chains.[5] Yet, these economies are necessary for the existence of any firm, and we should not expect retail chains to be any different. Many of these economies could also be had by a purchasing association. As such, they do not constitute a distinguishing attribute of retail chains. More to the point, while necessary, scale economies are simply not sufficient as an explanation for existence of retail chains, for scale economies alone cannot explain why a retail chain is comprised of a set of spatially distributed stores, or why such stores share a common brandname. Economies of scale alone cannot explain the existence of retail chains.

1.2.2 Transportation Costs

Perhaps the presence of consumer transportation costs can further an explanation of retail chains. In the absence of such costs, there would be little reason for the existence of more than one store, or for those stores to be spatially distributed. Consumers could simply travel to one store, which presumably is producing efficiently in the presence of scale economies. Thus, like economies of scale, transportation costs must necessarily

[5] Yet some have questioned their importance. For example,

> "The McDonald's worker takes as long to grill a patty as the lunch counter's short-order cook requires. The chain employee probably does achieve higher output because of steadier flow of customers. However, the short-order cook's output can be attributed to his labor along with the efforts of perhaps three or four other workers in the business. At McDonald's there are hundreds of people working behind the scenes. ... McDonald's has a personnel staff where assistant vice presidents ponder the problems of overseeing an international workforce. There are armies of inspectors and staff architects to plan buildings. Whatever efficiencies may be achieved in the McDonald's kitchen may be partially negated by the inefficiencies involved in running a large chain. (Luxenburg 1985, p. 97)

be present for retail chains to exist.

Clearly transportation costs must play an important role in any theory of retail chains. These costs enter the theory of retail chains developed in this thesis by defining the set of goods for which spatial distribution is of particular importance:

Definition 1.2 *A* ***convenience good*** *is a product for which transportation costs, and other inconvenience costs, are a significant fraction of the gross social surplus derived from the good.*

Spatial distribution exists to reduce the transportation costs of acquiring the product. So long as the costs of an additional store are less than the transportation-cost savings enjoyed by consumers, there are gains from trading such costs. If transportation costs were not significant, there would be little reason to have more than one store, or for those stores to be spatially distributed. As such, retail chains are of greatest import in the provision of relatively inexpensive goods—i.e. convenience goods. Only when transportation costs constitute a significant fraction of the gross social surplus derived from the product does the convenience provided by a spatially distributed set of stores have value.[6] Retail chains excel in the provision of convenience goods by minimizing the transportation costs and other inconvenience costs of the consumer.

Convenience goods are typically small-ticket items within the general class of consumption goods. Like most useful concepts, however, the concept of a convenience good cannot be clearly delineated. Even expensive consumer durables need to be somewhat convenient: no one wants to drive out of town, even to buy a new car. The relative insignificance of transportation costs in the full cost of acquiring such durables is, however, reinforced by the relatively greater significance of scale economies at the

[6] Gross social surplus per unit is the value the consumer places on the good (gross of all acquisition costs) less the total unit cost of production. For convenience goods, transportation costs per unit distance are significant, so such costs will quickly dissipate this gross surplus. As a result, the consumer will not travel far to obtain such goods.

level of the store. Since durable goods are expensive, distributors face large fixed costs of holding inventory. These fixed costs result in scale-economies at the level of the store and, therefore, the need to serve a greater market.[7] As such, the role of spatial distribution in the provision and acquisition of consumer durables, while significant, is less important than in the provision and acquisition of convenience goods.[8]

1.2.3 The Hotelling Model

The explanatory role of economies of scale and transportation costs in the spatial distribution of firms has been elegantly formalized for years using the Hotelling model of geographic differentiation. The model is not only elegant, but extremely powerful. Using just these two factors, the model can easily explain the existence of a set of spatially distributed stores offering a common product. The explanatory power of the model is further increased with the introduction of location-specific and product-specific capital. Together, such factors can go a long way in explaining not only the geographic density of such stores, but the profitability of various business strategies

[7]Almost invariably, it is the convenient servicing of consumer durables which ultimately justifies the use of spatial distribution in these cases. Department-store chains are similar to durable-good chains. Spatial distribution plays less of a role in both cases due to economies of scale at the level of the store and willingness of the consumer to travel farther to purchase the goods sold by such stores.

Economies of scale at the store level should be carefully distinguished from those at the brand, or chain, level. The latter economies are associated with the production of factors common to all stores in the chain. Since stores in the chain are offering a common product, or business format, it is often possible to centrally produce (or purchase) some of these items and then distribute them to member stores. Obviously, it is these economies which would tend to increase the importance of chains, and not scale economies at the level of the store. Other things equal, scale economies at the level of the individual store would imply fewer stores per square unit-distance, and therefore a smaller role for chains. It can be expected, therefore, that the relative importance of the chain in the distribution of the product should increase with the importance of scale economies at chain level relative to that at the store level. See Section 2.1 for more on consumer durables.

[8]Department store chains are very similar to capital good chains in that the consumer is typically making substantial expenditures on a number of products and, therefore, is usually willing to travel farther to get to such a store (formally, the consumer surplus is greater in such purchases). As such, the role of spatial distribution, while significant, is less important than in the provision of convenience goods.

such as early entry, excess capacity, entry deterrence, horizontal integration, collusion, etc.[9]

Clearly retail chains provide convenience within the bounds imposed by efficient production. But they also provide much more. Not all that fundamentally defines a retail chain can be explained on the basis of economies of scale and transportation costs. The Hotelling model can explain the existence of a set of spatially distributed stores offering a common product. Yet, in such a model, a consumer typically buys from only one store. Thus why would such stores share a common brand name? The same transportation costs which are necessary for a set of spatially distributed stores results in the consumer purchasing her product from only one store. Thus the role played by brandname sharing in such a model is limited at best. This argument is summarized as follows:[10]

Proposition 1.1 *Economies of scale in production and transportation costs are necessary, but not sufficient, as an explanation for the existence of retail chains. Economies of scale are necessary for the existence of any firm, and transportation costs are necessary for a firm to have spatially distributed stores. Yet while these two factors can explain the existence of a set of spatially distributed stores offering a common product, they cannot explain why such stores share a common brandname.*

---

[9] A vast literature covers these arguments. Nonetheless, on the role of economies of scale and transportation costs in the spatial distribution of stores see especially Hotelling (1929), Kaldor (1935), Eaton and Lipsey (1977) and Lancaster (1979). Current theory often views geographic space and other preference spaces symmetrically, treating spatial distribution as just another type of product differentiation. The present thesis argues against this approach. A reader unfamilar with the economic theory of product differentiation can find a good explication in Eaton and Lipsey (1989).

[10] Additional support for this proposition can be found in Section 3.1.2. The proposition formally requires the consumer to have a fixed most-preferred geographic location as traditionally assumed in the Hotelling framework.

1.3 Possible Explanations for Brandname-Sharing

In this section, two common arguments for brandname-sharing are examined: *(i)* economies of scale in advertising, and *(ii)* risk aversion of the consumer.

1.3.1 Economies of Scale in Advertising

It might be said that the argument of the previous section has gone too far. That, in fact, economies of scale and transportation costs are sufficient as an explanation of the existence of retail chains. That, by just extending the concept of scale economies to include economies of scale in advertising, the sharing of a brandname by a set of spatially distributed stores can be explained. By sharing and advertising a common brandname, the presumably fixed costs of advertising can be spread over a greater number of outlets.

While this is true, it simply shifts the burden of explanation. Now the question is: Why advertise? Explanations for advertising are typically grounded in the consumer having some sort of imperfect information. For example, the consumer may know her own preferences, but not the attributes of the product. Advertising then either directly informs the consumer of such attributes, or provides an indirect signal of their level. Alternatively, even if the attributes of the brand are known, the consumer may not know what it is like to consume such a bundle of attributes. Advertising then gives the consumer a feel for such consumption by shaping the image of the product (real or not) in the consumer's mind.[11]

These arguments provide good explanations for advertising by many firms. However, being constructed to explain advertising by any firm whatsoever, these arguments do not take advantage of the particulars of the market environment which retail chains have evolved to serve. In particular, economies of scale in advertising suffers

[11] A more detailed discussion of these arguments can be found in the Appendix to this chapter.

as an explanation for brandname-sharing in retail chains, for such an explanation is neither sufficiently general, nor sufficiently specific. The explanation is not sufficiently general since it does not allow for the existence of independent single-store brands. An explanation of branding and retail chains should be general enough to allow for the existence of such brands, but an explanation built around scale economies does not: in the presence of scale economies, all brands should have large chains. Further, the explanation is not sufficiently specific, for it is not formally connected to consumer mobility. Advertising influences the choice of any and all consumers, independent of their degree of mobility. Being disconnected from consumer mobility, such arguments cannot explain the observed locational choices typically made by retail chains. Retail chains tend to locate their stores where consumers are highly mobile, such as bus stations, airports, and along highways. But economies of scale in advertising exist regardless of where such stores are located.

1.3.2 Risk Aversion and the Argument of Familiarity

Another common explanation for retail chains and brandname-sharing is the Argument of Familiarity: the consumer chooses the familiar brand over one which is unfamiliar. For example, in their fundamental work on franchising, Oxenfeldt and Thompson (1968, p. 4) state:

> No single factor will account for the startling expansion of franchising during the last decade. It appears that much of modern franchising is linked to the development of the automobile which creates highly mobile customers who seek familiar and reliable services related to food, lodging, auto repair and travel.

Why does the consumer value familiarity? Obviously risk aversion could explain the choice of a familiar brand over one which is unfamiliar. However, risk aversion is both empirically insignificant and theoretically unnecessary. Risk aversion is empirically insignificant since retail chains typically sell inexpensive goods. As such, risk aversion is unlikely to be an important factor in the purchase decisions of consumers. The lack of a warranty on the attributes of these goods testifies to the lack of a significant risk premium. Theoretically, the assumption of risk aversion is unnecessary, for so long as the value of the familiar brand is greater than the expected value of the unfamiliar brand, the former will still be chosen. In fact, this argument was made by Akerlof (1970) in the conclusion of his seminal, and now classic, article on adverse selection:

> Chains—such as hotel chains or restaurant chains—are similar to brand names. ... These restaurants, at least in the United States, most often appear on interurban highways. The consumers are seldom local. The reason is that these well-known chains offer a better hamburger than the *average* local restaurant; at the same time, the local customer, who knows his area, can usually choose a place he prefers. (p. 500, italics in original)

Clearly risk aversion is not a good explanation for the existence of retail chains.[12] However, with or without the introduction of risk aversion, the Argument of Familiarity cannot serve as an explanation for the existence of such chains, for this argument is incomplete.

By *assuming* the consumer is more familiar with one brand than with the other, the Argument of Familiarity puts the cart before the horse. Until it can be explained how the consumer came to be more familiar with one brand, this argument simply begs the question. To explain the choice of one brand over another on the basis of its familiarity borders on circularity. A complete theory of branding must explain how

[12]Explanations of advertising are also often grounded in risk aversion, and therefore suffer for the same reason.

a brand comes to be adopted by the consumer in the first place. You simply cannot explain the popularity of a brand by arguing that it is popular, or the existence of a large retail chain by arguing that it is large. What must be explained is how the brand came to be familiar and how the chain came to be large.

1.4 Appendix: Economies of Scale in Advertising

Economies of scale in advertising provides a possible explanation for the sharing of a brandname across spatially distributed stores. The strength of the explanation of brandname sharing is, however, determined by the strength of the explanation for advertising. Why advertise? Advertising is an activity which should produce value for the consumer if there are to be gains from trade. Typically, explanations of advertising are built around a presumption that the consumer lacks perfect information about either *(i)* the existence or attributes of the brand, or *(ii)* her preferences over such attributes.

Even if a consumer has perfect information about the attributes of a brand, she may still not be sure whether this bundle of attributes would be to her liking. If consumers are unsure about their preferences, there may be room to mould those preferences through advertising. This approach to branding builds on work done in theoretical psychology and sociology, and is typical of theory found in business journals.[13] Advertising is viewed as shaping the image of the brand in the consumer's mind. Whether this image is purely imaginary, or founded in real attributes of the brand, such advertising serves to distinguish the brand from other brands in the same product group, perhaps reducing price competition in the process.

While it would appear that much of modern advertising attempts to shape the preferences of consumers towards the brand, this approach contrasts significantly from that typically taken by economists. In theoretical economics, the consumer is usually assumed to be sovereign: she knows what she wants, she just might not know what is available. Put more formally, consumers have well-defined preferences over

[13] For an excellent introduction to this approach on branding, see Keller (1998).

product-attributes; but lacking knowledge of these attributes, they find advertising either directly informative, or indirectly informative as a signal of such attributes. For example, in their important work on the theory of franchising, Mathewson and Winter (1985, p. 504) offer the following explanation:

> The principal ingredient in most franchise contracts is the franchisee's right to use a national brand name in exchange for a share of profits to the franchisor. A significant increase in the use of franchise systems occurred in the mid-1950s with an apparent increase in the efficiency of national brand names. This we attribute to three factors. First, the development of television meant that there was a more efficient nationwide information technology, reducing the cost of establishing national brand names. Second, an increase in travel by consumers meant that consumers were more often shopping in unfamiliar geographic areas, in which national but not local brand names would serve as signals of quality, enhancing the value of a national brand name. Finally, a continuing increase in the real income of consumers led to a further increase in the opportunity costs of search in retail markets, again enhancing the information value of brand names.[14]

The focus of their paper is franchising, and the rest of their paper deals with that important topic. Nonetheless, what is being offered here is an explanation for the

[14] Similarly, in another important article on franchising, Caves and Murphy II (1976, p. 574) state:

> Some franchised goods and services are purchased by mobile customers in local markets where they do not regularly shop. The cost of search for them is very high relative to the expected benefit; the assurance provided by the franchise trademark of a minimum level of quality in an alien market becomes particularly valuable to the buyer and thus can yield a rent to the producer.

sharing of a brandname across spatially distributed stores. The explanation emphasizes the costs associated with the search by a mobile consumer who finds herself in an unfamiliar geographic location. The consumer knows her own preferences, but does not know the quality of each brand. National, but not local, advertising serves as a signal of quality, reducing the "search costs" of the consumer.

The **signalling argument** for the existence of retail chains is then that advertising acts as a signal of quality to consumers prior to purchase; that prior knowledge of quality has value to consumers; that the firm can capture the custom of consumers by expending resources on advertising; and, finally, that the cost of this signal can be lowered by spreading these advertising expenditures over a greater number of stores.

Signalling arguments have been used to explain the use of advertising and other market mechanisms in situations of both adverse selection and moral hazard. For example, in cases of *adverse selection*, quality is fixed, perhaps through past investment decisions, and the consumer must determine which firms produce high quality and which produce low. In such cases,

> In order for advertising to be an effective signal, high-quality firms must be able to recover advertising costs while low-quality firms cannot. Consumers must also be well enough informed about costs to realize that advertising is profitable for high-quality but not for low-quality firms. (Kihlstrom and Riordan 1984, p. 429)

However, technology may be such that the firm can vary quality from time to time, if it so chooses. If such is the case, the consumer then faces a *moral hazard* problem with respect to each firm's choice of quality. Here too, signalling models have been applied. For example, Klein and Leffler (1981) have argued that high quality may be signaled through the use of a price-premium which rewards the firm for maintaining a high-quality product. In such a model,

> Consumers are assumed to behave as if they know the cost functions of firms, and, given the prices charged, they can put themselves in the firm's position and calculate whether the benefit of producing high-quality items and maintaining a good reputation is greater than the cost involved. They are thus able to infer indirectly the quality of goods a profit-maximizing firm produces, even though they cannot directly observe it. (Allen 1984, 311)

Such models have greatly improved our understanding of quality uncertainty and the market mechanisms used to help reduce its severity. Nonetheless, both models suffer from a similar criticism: they simply shift the burden of knowledge possessed by the consumer. Rather than knowing something about an observable product, the consumer is assumed to know something about the unobservable cost structure of the firm. No doubt it is possible for a consumer to lack knowledge of the product's quality, yet possess knowledge of the product's cost. While possible, however, this is improbable.

Even more important, while a consumer may know how a firm's cost structure should vary with quality, there are many other attributes of a product which a consumer values and which have no clear relationship to costs. Should a sweeter product cost more, or less, than one that is less sweet? Clearly there is much more to a product than simple quality. Further, even if it is possible to advertise every conceivable attribute of a product, we are still left with the possibility that the consumer herself may not know whether she would like such a combination of attributes—at least until she tries the product. Fortunately for the consumer of inexpensive convenience goods, she can do just that: try it; so knowledge of the firm's cost structure is not necessary.

Perhaps advertising as a signal of quality is more applicable to the case of consumer durables, since the consumer cannot cheaply try such goods and many of their attributes (such as durability) will only become known after considerable use.[15] But here the role of the retail chain in the distribution of durable goods is more limited. As mentioned previously, retail chains, almost by definition, sell convenience goods: products for which transportation costs (and other inconvenience costs, such as waiting time, etc.) are a significant fraction of the gross social surplus derived from the good. If such costs were not an important factor in the purchase decision of the consumer, there would be little need for more than one store, or for those stores to be spatially distributed. Durable goods, on the other hand, are typically expensive and offer many periods of utility. Because of their expense, consumers search and research the characteristics of such goods before purchase, and advertising may (directly and/or indirectly) aid them in this task. But while search and research costs play a significant role in the purchase decision of a durable good, the cost of travelling to acquire the good typically does not. The transportation costs associated with the acquisition of such goods are typically small relative to the surplus derived, so the need for retail chains in the provision of such goods is more limited.[16]

[15]Nonetheless, even here it would seem that warranties might provide the best signal of all. Perhaps advertising warranties the warranty by signalling to the consumer that the firm expects to recoup their investment in advertising over many years of faithful service to the customer.

[16]Of course retail chains do sell durable goods, but the provision of such goods through the chain distribution system is typicially for the purpose of servicing such goods, usually under warranty.

Chapter 2

Temporal Branding and The Value of Information

2.1 Temporal and Spatial Branding

Professor Akerlof's essential insight is correct: "Chains ... are similar to brand names." At the heart of the success of retail chains lies the consumer's ability to make *two fundamental associations:* (*i*) associate a particular store with a particular chain, and (*ii*) associate a particular chain with a particular set of products and services. Retail chains facilitate the first association by having stores share a common brandname; they accomplish the second association by having these stores offer a common set of products and services.

Branding is a convention which warrants consistency. Under this convention, the consumer pools sample information obtained from separate purchases of the product made under the same brandname to form a single subjective estimate of the value of the next purchase made under that brand. Purchases of the product may be separated by both time and space. As such, branding can be naturally decomposed into its spatial and temporal components. **Spatial branding** warrants consistency at a point in time across stores sharing a common brandname. In contrast, **temporal branding** warrants consistency of a particular store over the time for which the brandname is used.

2.1.1 Money as an Analog

The relationship between spatial and temporal functions of branding is much like that which exists between the medium-of-exchange and the store-of-value functions of money. Money and branding are both conventions which, to the extent they are followed, serve to reduce the costs of exchange by facilitating a double coincidence of wants.

Money is anything which is held because it is generally and readily accepted in exchange for goods and services. This convention warrants the belief of the consumer that, prior to negotiating the exchange, *she has what the seller wants.* However, money just facilitates one side of the double coincidence of wants; advertising and branding facilitate the other. In particular, advertising typically identifies the merchant as selling products within a particular product group, whereas branding allows the consumer to distinguish products within the group. Once a brand has been purchased, the costs of subsequent exchanges are reduced. Branding is a convention which warrants the belief of the consumer that, prior to negotiating subsequent exchanges, the *seller has what she wants.* These conventions reduce the costs of exchange, and the value of each depends on the extent to which the convention is followed.

These conventions are supported by a myriad of market mechanisms. For example, the use of central bank notes as a generally accepted medium of exchange is supported by government fiat through the concept of "legal tender." The central bank further supports the use of its own notes, and the chequable and debitable deposits issued by private sector banks, by voluntarily controlling inflation. As well, private sector banks maintain their own credit worthiness by voluntarily holding sterile cash reserves, so the deposits they create through their extension of loans will also find general acceptability.

Branding, too, is supported by law. In fact, until the beginning of this century, trademark licensing (such as that contained within a franchise contract) was not allowed by American courts as it was contrary to the prevailing "source theory" of trademark use.[1] According to this theory, the function of a trademark was to indicate the source of a product. Thus permitting the use of a mark by more than one person violated that function. However, with the increasing use of franchising in the development of spatial distribution systems, the courts gradually relaxed their stance on trademark use, moving away from the "source theory" towards a "guarantee theory" of trademark use. Under this new theory, codified in Section 5 of the 1946 Lanham Trademark Act, a trademark could be used by "related companies", where a "related company" was defined to be one that is controlled by the registrant of the trademark as to the nature and quality of the goods or services with which the trademark is used. Thus consistency is a necessary requirement for the valid use of a trademark. This requirement of consistency is recognized by retail chains, and reflected in the strict restrictions imposed on franchisees in their franchise contracts.

Legal requirements notwithstanding, in the end, money and branding are simply conventions. Both conventions struggle to maintain a consistent standard while adapting to meet the changing requirements demanded by the markets within which they function. These conventions are valued, for they increase the gains from trade. As such, market participants expect these conventions to be followed, and are often

[1] See Thompson (1971, p. 13).

surprised when they are not. Nonetheless, in particular economies experiencing periods of hyperinflation or bank runs, the monetary convention may break down. So too with branding. A particular market may experience hard times. If such a market is susceptible to moral hazard by producers, the currency of branding may be expected to carry little store of value, and the convention breaks down. But such cases are the exception, not the rule.

2.2 The Value of Information: A Model of Temporal Branding

Branding creates value for the consumer in a variety of ways. The model of this section demonstrates that branding gives value to the information the consumer obtains from trying the product. Since the consumer is risk-neutral, this information has no value in risk reduction. This is important since retail chains are of particular significance in the distribution of convenience goods: inexpensive goods for which there is little, if any, risk premium.[2] The information derives its value by improving the consumer's future purchase decisions. Obviously future purchases take place in both time and space. Each of these dimensions creates a role for branding. This chapter builds a simple model of temporal branding which will be expanded upon in the analysis of spatial branding contained in the chapters to follow.

[2] As applied within this thesis the concept of "risk premium" is here defined to be the reduction in expected surplus the consumer is willing to accept in order to eliminate all variability in the outcome of the purchase. The argument made here is that because convenience goods are inexpensive they represent a small gamble to overall consumer welfare and, therefore, would be taken on actuarially fair odds. Put somewhat more formally, the utility function of the consumer is approximately linear for small changes.

2.2.1 Brand Selection as a Bandit Problem

Consider a risk-neutral consumer faced with a choice between two brands of unknown value (i.e. consumer surplus).[3] After she purchases a brand, she learns a little bit about its value which can be used to improve her future purchase decisions. The consumer would like a sequence of brand selections which maximize the expected present value of utility over her life.

Formally, this problem can be modelled as a "two-armed bandit". Bandit problems are a family of problems in the theory of sequential allocations of experiments.[4] The family gets its name from the problem-situation of a gambler facing two one-armed bandits of the mechanical kind. Each machine has a different but unknown probability of paying off when a coin is deposited and the arm is pulled. The gambler must sequentially decide into which machine she should put her next coin given her previous play experience. Over time, as she plays each machine, she obtains a pool of sample information which can be used to estimate the frequency of each machine's payoff. Finding an optimal policy, or sequence of play, can become quite complex.[5] Nonetheless, it is possible to impose simplifying restrictions and yet maintain the

[3] A brand's "value" is defined as the *net* increase in welfare (or utility) the consumer expects to derive from the purchase and consumption of a brand in a single period. The terms "value", "utility" and "surplus" are used interchangably and in combination as convenient.

[4] See, for example, DeGroot (1970).

[5] In discussing the contribution of an important paper by Professor J. C. Gittins (Gittins 1979), the following comment was made by Professor P. Whittle:

> [T]he problem is a classic one; it was formulated during the war, and efforts to solve it so sapped the energies and minds of Allied analysts that the suggestion was made that the problem be dropped over Germany, as the ultimate instrument of intellectual sabotage. In the event, it seems to have landed on Cardiff Arms Park. And there is justice now, for if a Welsh Rugby pack scrumming down is not a multi-armed bandit, then what is? (p. 165)

It is not surprising, given the problem's complexity, that few applications of the bandit model exist in the economic literature. However, see Rothschild (1974) for an application to pricing decisions under imperfect information about demand.

fundamental implications for consumer-choice theory embodied in such a policy.

Here the consumer's choice between brands is modelled as a choice between fixed but unknown distributions over consumer surplus.[6] Each brand is formally represented by a distribution of possible surplus-values; each purchase is an observation from the unknown distribution which represents that brand. For simplicity, brands are assumed to be characterized by distributions which are fixed (i.e. time-invariant or stationary), have a known variance, and are independent of each other. Since the consumer is risk neutral, her welfare in any period then depends only on the fixed but unknown mean of the distribution. As such, the problem can be simplified considerably by assuming that the variance of the distribution is equal to zero. The consumer will then obtain perfect information about the mean value of a brand after just one purchase.[7]

2.2.2 A Simple Numerical Example

The problem may be made more concrete through use of a simple numerical example. In particular, assume the mean value of Brand a is either 8 or 0 with equal probability. Similarly, the mean value for Brand b is either 6 or 2, again with equal probability.

[6] Note that the uncertainty associated with the value of a brand could be derived from the consumer's uncertainty concerning either the characteristics of the brand or her own preferences over such characteristics.

[7] The assumption that perfect information is obtained from just one purchase reduces the bandit problem to its simplest possible form, ruling out many possible realizations in the sequence of selections produced by an optimal policy. But, again, the essentials insights for consumer choice embodied in such a policy are preserved.

The work of this thesis has been expanded using computer simulation capable of handling a true bandit model in which (Bayesian) learning is more gradual. While space precludes full discussion of this work, it is briefly reported on in the Appendix to Chapter 3.

Following Hart (1942) and Jones and Ostroy (1984), the term **risk** will be used when all parameters of the distribution of interest are known, reserving the term **uncertainty** for situations when at least one of the parameters is unknown. Under risk-neutrality, we can replace distributions with their expected values in situations of pure risk, but not in cases of uncertainty. See also the discussion of Section 2.4, below.

It is well-known that despite the fact the prior expected value of each brand is 4, a risk-neutral consumer would not be indifferent between these brands when faced with this choice for the first time.

To see that this is the case, suppose the consumer has a two-period time horizon with a zero discount rate. A **purchase policy** π is a sequence of brand selections, one for each period. That is, $\pi = (\pi_1, \pi_2, ...)$ where π_t denotes the selection in period t under policy π. Such selections may be based on the beliefs of the individual in that period, as well as other parameters of the problem. Some purchase policies are forward looking, taking into account not only current utility but also the effect of the decision on the present value of future utility. A policy which does not take such future-value effects into account will be called myopic. Thus a **myopic policy** is one which selects the brand having the highest current-period expected-value. The consumer wants to choose an **optimal purchase policy**, π^*: one which maximizes the present value of her current and future utility. Consistent with the terminology of stochastic decision theory, a brand will be called **optimal** if it is the first selection under an optimal policy. Thus, an optimal policy may then be thought of as a sequence of optimal selections: $\pi^* = (\pi_1^*, \pi_2^*, ...)$. Let V^* denote the expected present value of such a policy.

Formally, finding an optimal purchase policy is a matter of folding back the decision tree (also known as backward induction or the principle of optimality, among other names). Thus, consider the consumer's last-period problem first. Obviously she wants to choose optimally in the last period. Since nothing could possibly be gained by foregoing some expected return in the last period, it is optimal for her to follow a myopic policy in that period. She will thus choose the brand with the highest expected return in the last period, given her beliefs at that time. We need now only consider policies which choose either Brand a or Brand b in current period and optimally thereafter (i.e. myopically in period 2). Obviously there are two such policies:

- **Policy** A: Choose Brand a in period 1, and optimally thereafter. Let V_a^* denote the expected present-value of this policy. If Brand a is chosen in period 1 and found to have value of 0, the consumer would switch to Brand b in the second period since its expected payoff is $4 > 0$. However, if Brand a is found to have value of 8, then it is optimal to choose Brand a in the second period as well. Each of these possibilities has prior probability of 1/2. Thus the expected present value of this policy is

$$V_a^* = 4 + \frac{1}{2}(8) + \frac{1}{2}(4) = 10 \tag{2.1}$$

- **Policy** B: Choose Brand b in period 1, and optimally thereafter. Let V_b^* denote the expected present-value of this policy. If Brand b is chosen in period 1 and found to have value of 2, the consumer would switch to Brand a in the second period since its expected payoff is $4 > 2$. However, if Brand b is found to have value of 6, then it is optimal to choose Brand b again in the second period. Each of these possibilities has prior probability of 1/2. Thus the expected present value of this policy is

$$V_b^* = 4 + \frac{1}{2}(6) + \frac{1}{2}(4) = 9 \tag{2.2}$$

Since $V_a^* > V_b^*$, Brand a is optimal. It follows that a risk-neutral consumer is not indifferent between the two brands, despite the fact that they have the same expected value. Quite naturally, the consumer values information about the brand for which she is more uncertain. Such information has value because it improves her future purchase decisions. This information value plays an important role in the consumer's initial choice between brands and has important implications for both the theory of consumer choice and the theory of retail chains.

2.3 Generalizing the Model

In the previous section it was shown that consumer's brand-selection problem is truly dynamic in that her decision in the second period potentially depends on what she learns in the first period. As such, the periods are "linked" and we should not expect a myopic purchase policy which simply selects the brand with the highest expected return for the current period to be optimal.

2.3.1 The Role of the Discount Rate

The consumer's decision problem will now be allowed to have an infinite horizon with future purchases discounted at rate $i > 0$. Although the assumption of an infinite horizon is quite unrealistic, the alternative of a known arbitrary time of death (i.e. a finite horizon) is both equally unrealistic and too grim to contemplate. Thus, it is assumed that the consumer acts as if she should could live forever, but with some small probability of death any period. Discounting future purchases at rate $i > 0$ captures not only this small probability of death, but also plain impatience. A positive discount rate means the future is somewhat less important to the consumer than the present.

It is important to note that all models of this thesis are formally unchanged if one allows the consumer's purchase decisions to be separated by a given number of periods, or even by an uncertain number of periods with a known distribution. These extensions will simply determine the value of the discount rate by affecting the expected frequency of purchase. In fact, the primary role of the discount rate herein is to capture the frequency with which the consumer expects to purchase any of the brands from the product group. In particular, the smaller the expected frequency of purchase, the greater the discount rate.

2.3.2 Representing Consumer Beliefs

It will prove useful to have some notation for the beliefs of the consumer. Thus, let x_a and x_b denote, respectively, the distributions characterizing the beliefs of the consumer about Brand a and Brand b. Further, let $\bar{x}_a$ and $\bar{x}_b$ denote the respective means of these distributions. When necessary, prior beliefs for Brand a will be denoted by x'_a and posterior beliefs by x''_a.

The simple numerical example above illustrated the optimality of selecting the brand with the more uncertain value. This was Brand a. Even though the expected current-value to be derived from each brand was the same, the consumer stood to learn more by trying Brand a first. Since this result is only dependent on the relative uncertainty of the brands, the model can be further simplified by letting Brand b have a known expected value of μ_b. This leaves the consumer uncertain only about the expected value of Brand a.

In particular, let the consumer's prior uncertainty about the expected value of Brand a be characterized by a distribution with its mass distributed evenly over just two points, $\mu_a + \sigma_a$ and $\mu_a - \sigma_a$. Let $1/2_{\mu_a \pm \sigma_a}$ denote these prior beliefs. Thus, $x'_a = 1/2_{\mu_a \pm \sigma_a}$. These are the beliefs of an **inexperienced consumer**—one who has not bought Brand a previously. It is easily verified that the consumer's prior distribution over the expected value of Brand a itself has mean μ_a and variance σ_a^2.

After the consumer has tried Brand a, she receives perfect information about its expected value. Thus her posterior distribution will have all of its mass over just one of the two points which support her prior distribution. Let $1_{\mu_a+\sigma_a}$ and $1_{\mu_a-\sigma_a}$ denote these two equally-likely posterior distributions. Thus $x''_a \in \left\{1_{\mu_a+\sigma_a}, 1_{\mu_a-\sigma_a}\right\}$. These are the possible beliefs of an **experienced consumer**—one who has bought Brand a previously and knows its expected value. Such valuation may be purely subjective; consumers need not agree about the value of Brand a's attributes. Finally, since the consumer has no uncertainty about the expected value of Brand b, let 1_{μ_b} denote her prior beliefs for this brand. The consumer's beliefs about Brand b are, therefore, static and suppressed in the functions to follow.

2.3.3 Finding An Optimal Policy

For simplicity assume $\mu_b > 0$ so the consumer never entirely reserves purchase. Let $\Delta \equiv \mu_b - \mu_a$ denote the expected opportunity cost of selecting Brand a. To make the problem interesting, the following restriction is imposed:[8]

Parametric Restriction 2.1 *It is assumed that neither brand is dominant. Algebraically,* $\sigma_a > \Delta > 0$.

Since Parametric Restriction 2.1 implies $\mu_a - \sigma_a < \mu_b < \mu_a + \sigma_a$, it follows that an optimal policy π^* is a member of a family of policies which satisfy the following condition:[9]

$$\pi^*\left(x_a''\right) = \begin{cases} a & \text{if} \quad x_a'' = 1_{\mu_a+\sigma_a} \\ b & \text{if} \quad x_a'' = 1_{\mu_a-\sigma_a} \end{cases} \tag{2.3}$$

That is, an optimal policy will require an experienced consumer to select Brand a if it is of high utility-value, but Brand b otherwise. We now turn to finding the remaining portion of such a policy: that which applies to a consumer who has no experience with Brand a (i.e. that for $x_a' = 1/2_{\mu_a \pm \sigma_a}$).

[8] This restriction rules out prior dominance of either brand. In particular, if $\Delta \leq 0$ then $\mu_b \leq \mu_a$ so there would be no opportunity cost to selecting Brand a and it would always be optimal. Further, if $\sigma_a \leq \Delta$ then $\mu_a + \sigma_a \leq \mu_b$ so there would be nothing to gain by selecting Brand a and it would never be optimal.

[9] Note that this family of policies is stationary (i.e. not an explicit function of time) and thus the time subscript has been dropped.

To this end, let $V^*(x_a)$ denote expected present-value of the optimal policy. Such a policy will use the information contained in the consumer's beliefs, x_a. As before, in finding an optimal policy, attention can be restricted to policies which choose either Brand a or Brand b in the current period, and optimally thereafter. Let $V_a^*(x_a)$ and $V_b^*(x_a)$ denote, respectively, the expected present-values of these policies. It will be useful to express the values of these policies somewhat more explicitly than in the previous section.

To this end, let $u_k(x_a)$ denote the expected current utility derived from the selection of Brand $k \in \{a, b\}$ under beliefs x_a. Then we can write

$$V_k^*(x_a') = u_k(x_a') + E_{x_a''|k,x_a'} V^*(x_a'') \tag{2.4}$$

where $E_{x_a''|k,x_a'}$ denotes the expectation over posterior beliefs x_a'', conditioned on prior beliefs x_a' and the selection of Brand k in the current period. Note that only if $k = a$ will the consumer's beliefs change (i.e. $x_a'' = x_a'$ if $k = b$). Now the values of Policies A and B can be found as before:

- **Policy** A : Choose Brand a in the current period, choosing optimally thereafter. Thus we have

$$\begin{aligned} V_a^*\left(1/2_{\mu_a \pm \sigma_a}\right) &= u_a\left(1/2_{\mu_a \pm \sigma_a}\right) + E_{x_a''|a,1/2_{\mu_a \pm \sigma_a}} V^*(x_a'') \\ &= \mu_a + \frac{1}{2}V^*\left(1_{\mu_a+\sigma_a}\right) + \frac{1}{2}V^*\left(1_{\mu_a-\sigma_a}\right) \end{aligned} \tag{2.5}$$

If, after trying Brand a, the consumer finds it to be of high value (i.e. $x_a'' = \mathbf{1}_{\mu_a+\sigma_a}$), then it is optimal to choose Brand a thereafter. So $V^*\left(1_{\mu_a+\sigma_a}\right) = (\mu_a + \sigma_a)/i$. However, if the consumer finds Brand a to be of low value (i.e. $x_a'' = \mathbf{1}_{\mu_a-\sigma_a}$) then it is optimal to choose Brand b thereafter. So $V^*\left(1_{\mu_a-\sigma_a}\right) =$

μ_b/i. Thus the value of this policy is

$$V_a^*\left(1/2_{\mu_a\pm\sigma_a}\right) = \mu_a + \frac{\mu_a + \sigma_a + \mu_b}{2i} \tag{2.6}$$

- **Policy** B : Choose Brand b in the current period, choosing optimally thereafter. Thus we have

$$V_b^*\left(1/2_{\mu_a\pm\sigma_a}\right) = u_b\left(1/2_{\mu_a\pm\sigma_a}\right) + E_{x_a''|b,1/2_{\mu_a\pm\sigma_a}} V^*\left(x_a''\right) \tag{2.7}$$

Under this policy, the beliefs of the consumer do not change, i.e. $E_{x_a''|b,1/2_{\mu_a\pm\sigma_a}} V^*\left(x_a''\right) = V^*\left(1/2_{\mu_a\pm\sigma_a}\right)$. It follows that if it is optimal to choose Brand b in the current period, it is optimal thereafter since the future looks identical to the present.[10] So, under this hypothetical, $V^*\left(1/2_{\mu_a\pm\sigma_a}\right) = \mu_b/i$. Thus the value of this policy is

$$V_b^*\left(1/2_{\mu_a\pm\sigma_a}\right) = \mu_b + \frac{\mu_b}{i} \tag{2.8}$$

[10] In solving for the value of Policy B, the argument was used that, since the beliefs of the consumer do not change when Brand b is selected in the current period, the future looks identical to the present. As such, if Brand b is optimal in the current period, it is optimal thereafter. In their work on the general theory of bandits, Berry and Fristedt (1985, Theorem 5.2.2, p. 92) have shown that this argument is available only if the discount sequence satisfies a particular regularity condition. Roughly, this condition requires that the value placed on future benefits (i.e. discount factors) not rise too much over time. If, at a later date, the consumer places a higher value on future benefits than she does today, it may pay for her to acquire information about the uncertain brand later on, even though it does not pay for her to do so presently. Fortunately, the geometric discount sequence (i.e. constant discount rate) typically used in economic models (and used herein) satisfies the regularity condition. In the Hotelling model of Chapter 4, however, this argument will not be available to us despite the use of a geometric discount sequence. Nonetheless, an optimal policy is found.

Comparing the values of these two policies from Equations 2.6 and 2.8, it follows that Brand a is optimal if

$$\begin{aligned} V_a^*\left(1/2_{\mu_a \pm \sigma_a}\right) &\geq V_b^*\left(1/2_{\mu_a \pm \sigma_a}\right) \\ \mu_a + \frac{\mu_a + \sigma_a + \mu_b}{2i} &\geq \mu_b + \frac{\mu_b}{i} \\ \frac{\sigma_a - \Delta}{2i} &\geq \Delta \end{aligned} \tag{2.9}$$

2.3.4 Interpreting the Optimal Policy

Inequality 2.9 implicitly defines the optimal policy for the inexperienced consumer. Explicitly stated, this policy requires

$$\pi^*\left(1/2_{\mu_a \pm \sigma_a}\right) = \begin{cases} a & \text{if} \quad \sigma_a \geq (1+2i)\,\Delta \\ b & \text{if} \quad \sigma_a < (1+2i)\,\Delta \end{cases} \tag{2.10}$$

Does such a policy make sense? If the consumer was to buy Brand a in period 1 she would be giving up an expected gain of Δ in the current period. The inequality in the optimal policy of Equation 2.10 says that no matter how large is Δ, there always exists a value for σ_a which will induce the consumer to give up this current-period gain for information about Brand a that could significantly improve (by an amount $(\sigma_a - \Delta)/i$) her future purchase decisions. As Professor Gittins explains in his discussion of Professor Bather's important paper on the general theory of bandits (Bather 1981), the trade-off is between the exploitation of current information and the exploration for new information:

> This strategy, in which we simply use the information that is already available, is what I would like to put under the *exploitation* heading. *Exploration,* on the other hand, is the consideration which this does not take into account: that is, the need to gather information for subsequent use,

which may conflict with the desire to achieve the best possible immediate return. (p. 283, italics in original)

The optimal purchase policy is perhaps best understood in terms of the information value associated with the purchase of Brand a. To see this, let V^m denote the present value of a myopic policy, π^m. Under the assumption $\Delta > 0$, such a policy would select Brand b exclusively and earn a present value of $\mu_b + \mu_b/i$. Thus the expected net present value of information is

$$V^* \left(1/2_{\mu_a \pm \sigma_a}\right) - V^m \left(1/2_{\mu_a \pm \sigma_a}\right) = \begin{cases} \frac{\sigma_a - \Delta}{2i} - \Delta & \text{if} \quad \sigma_a \geq (1+2i)\,\Delta \\ 0 & \text{if} \quad \sigma_a < (1+2i)\,\Delta \end{cases} \tag{2.11}$$

The first term, $(\sigma_a - \Delta)/2i$, is the expected value of a perpetual option on Brand a. Should Brand a be of high value (i.e. "in the money"), it will have a net pay off of $(\sigma_a - \Delta)/i$. This occurs with probability $1/2$. The second term is the forgone current-period expected-value Δ: the premium paid for this option. Obviously, only if the expected value of the option on Brand a exceeds its premium will the consumer try Brand a for the first time. These features of the consumer's choice problem are summarized in the following proposition:

Proposition 2.1 *The consumer's incentive to experiment with an unknown brand (Brand a) is (i) increasing in the consumer's uncertainty (σ_a) about the value of the brand, (ii) decreasing in the opportunity cost (Δ) of trying the brand, and (iii) increasing in the frequency of purchase from the product group (i.e. decreasing in the discount rate i).*

2.4 Implications for Brand Value: The Importance of Consistency

What determines the willingness of the consumer to experiment? Obviously the greater the value of the known brand, the greater the opportunity cost of the experiment. But costs must be weighed against benefits; or to paraphrase Professor Gittins: the exploitation of what is known must be weighed against exploration of what is not. Other things equal, the value of experimentation with a brand is greater, the greater the consumer's uncertainty about the brand. Here, it is important to distinguish the concepts of risk and uncertainty. Conceptually, the prior randomness the consumer perceives in the value of a purchase from a brand can be separated into two parts:

1. **Risk:** Irreducible randomness inherent in the brand, perhaps derived from variations in the value of the inputs or the production process itself.
2. **Uncertainty**: Reducible randomness simply due to the consumer's lack of familiarity with the brand.

It might be expected that in the absence of risk aversion, consistency would have little value. Yet despite the absence of risk aversion, consistency still matters. The consumer formally views a brand as a *consistent* distribution of possible surplus-values—i.e. one that is stationary, or constant, across both time and, as will be shown in the next chapter, space. Consistency entices the consumer to experiment, much like an honest dealer entices a gambler to bet. Experimenting with a branded product is therefore much like sequentially betting on flips of a marked coin: regardless of whether the coin is fair or biased, the marking allows the coin to be identified as one which has been flipped before, and to use the outcome of previous flips to improve bets laid on future flips. In fact, a biased coin is to be preferred in a sequential game. A fair coin represents a situation of pure risk: there is nothing to be learned by flipping a fair coin which could improve future betting decisions. But a biased coin provides an opportunity to learn about the bias with each flip. As the degree of bias increases, the value of experimentation also rises, regardless of on which side the bias lies. In the limit, a coin which is perfectly biased represents a situation of pure uncertainty and no risk; such a coin need only be flipped once to determine on which side the bias lies. As argued in the introduction to this chapter, branding is a convention which increases the value of experimentation, effectively by allowing consumers to identify the coin they are betting on each period.[11]

2.4.1 Experience Goods

The concept of branding as formalized in the present chapter is closely related to the concept of experience goods, first introduced by Nelson (1970). **Experience goods** are products having attributes about which the consumer is uncertain, but the value of which she can come to learn through purchase and consumption of the product. Effectively, the present chapter has argued that a consumer will take into account the expected value of "experience information" obtained through the purchase of an **experienced brand**—i.e. a branded experience good. This "experience value" is

[11]The relationship between risk and uncertainty is much like that between "cost" and "sunk cost." Like cost, risk is a static concept: to understand these concepts a one-period model will do. In contrast, sunk cost and uncertainty are dynamic concepts: they can only be understood in a multiperiod setting.

A study of consistency and its relationship to brand value would be important in itself. I have made some strides in this regard, building on the work of Professor Jones, in particular (Jones and Ostroy 1984) and (Jones 1995). However the interest of the present work lies elsewhere. In this thesis all consumers start off uncertain about the value of a brand, whatever the determinants of that uncertainty may be. The interest of the present work lies not in the determinants of brand-value generally, but in the determinants of brand-value within a spatial setting. That is, we seek the factors which are specifically relevant to the determination of brand-value for retail chains. Given a level of uncertainty, what are the factors which specifically affect the value of a retail chain's brandname?

distinct from its complement, "search value", which is the value of information gained through means other than purchase and consumption.

Almost all brands are experience brands, but not all brands have high experience-value. Branded consumer durables, for example, are replaced infrequently so information acquired about the brand through the purchase and consumption of the good will not be used in a future purchase decision for some time. By that time the good may have undergone significant changes in its attributes. In fact, consumers expect such changes as part of product development. Further, the durability of these goods gives rise to their expense. As such, a significant risk premium is associated with their purchase, as evidenced by the warranties typically offered with such purchases.[12] The expense of durable goods and the infrequency with which they are purchased means it is unlikely that consumers will buy such goods for the purpose of improving their future purchase decisions.

Consumer durables have low experience value, but high search value. Convenience goods, in contrast, are experience goods with high experience value. Because they are nondurable, they are purchased frequently. As a result, information acquired through purchase will find quick use in many future purchase decisions—even if that information dictates that the brand in question not be purchased. Further, because they are inexpensive, there is little, if any, risk premium associated with their purchase. The consumer comes to learn the value of such goods, not by searching, but simply by trying.

My senior supervisor puts the argument of this section quite nicely: "For durable goods, "experience" is not an attractive way to learn because the value of learning is not realised, if at all, for a very long time, and because there may be a lack of consistency—the good may very well change, so today's experience is no guide to

[12]This is not to say that warranties are always evidence of risk aversion: a warranty could, for example, arise out of a signalling equilibrium even if consumers are risk neutral.

"tomorrows" purchase. For non-durable, inexpensive goods, "search" is not attractive because "experience" is cheap."[13] Within the context of this thesis, the distinction between convenience goods and consumer durables is captured by the rate of discount. As the time between purchases increases, so too does the rate of discount. As is evident from the model above, an increase in the discount rate reduces the value of information.[14]

[13]Professor B. Curtis Eaton, comments on first draft, January, 2000.

[14]Consumer durables are often produced at one or just a few production facilities. The centralized control over production means the manufacturer can provide detailed product information and warrant the accuracy of this information, as well as the attributes of the product in question. Warranties and central production facilities means the job of maintaining consistency across outlets in a chain that sells durable goods is essentially reduced to the job of maintaining a consistent level of customer service. Branding is still extremely important to such chains, but the relative importance of spatial branding is reduced. In contrast, chains which distribute products largely produced at the local outlet rely on their brandname to assure spatial consistency to the customer. Restaurant chains, hotel chains, and almost all chains for which service is the primary product being purchased, would fall into this category. This argument explains, in part, the classification of franchised chains into "business-format franchising" and "product franchising." The classification is important, and empirical work in franchising should typically not group these classes together.

Chapter 3

The Theory of Spatial Branding and Retail Chains in Discrete Geographic Space: The Concept of Minimum-Informative Scale

3.1 A Simple Model of Retail Chains

In the Introduction it was argued that economies of scale and transportation costs are necessary conditions for the existence of retail chains. Economies of scale are necessary for the existence of any firm, and if transportation costs did not exist, consumers could simply travel to one store, so a set of spatially distributed stores would not be necessary. Yet while economies of scale and transportation costs can explain spatial distribution, they appear less able to explain spatial branding.

One might expect that since the relative importance of spatial distribution increases with the relative importance of transportation costs, so too does the relative importance of spatial branding. However, the relationship between transportation costs and spatial branding is not as simple as it first appears. In a Hotelling model, for example, the same transportation costs necessary for the existence of a set of spatial distributed stores results in the consumer purchasing her product from only one store. Thus, why would such stores share a common brand name?

Consider, for example, the analogy with money. If consumers buy from only one store, money would only have to function as a store of value; there would be little need for a *generally* accepted medium of exchange. And so too with branding. If consumers buy the product from only one store, branding would only have to function temporally, not spatially. As such, there would be little need for stores to share a common brandname.

The model of the last chapter formalized the concept of temporal branding—a key component of all branding. Fundamental to the concept of temporal branding is consistency of the product over time. Such consistency allows the consumer to use information acquired through a current purchase in her future purchase decisions. One might expect that since temporal branding simply requires temporal consistency, the only additional requirement for spatial branding is spatial consistency. However, this too is not the case. While consistency is necessary, the model of this chapter shows that the value of a spatial brand is determined by more than mere consistency.

3.1.1 The Importance of Space

We are interested in determining whether there exist factors other than consistency which might affect the value of information conveyed by branding. As such, perfect consistency is assumed, but now both temporally and spatially. It is shown that geographic space, consumer mobility, and the size of a brand's retail chain can combine to affect the value of a spatial brand. In particular, in a spatial setting, limitations on the use of information may be imposed by the consumer's travel behavior.

The argument is very simple. The value of information obviously depends on the extent to which it can be employed, for even perfect information has zero value if it never gets used. Since consumers are not willing to travel far to obtain convenience goods, their travel pattern imposes limitations on the ability to use information acquired through purchase. Retail chains attempt to restore the value of experimenting with a brand by providing a high level of convenience. Convenience not only determines the availability of the brand in the current period (or more generally, the current cost of acquisition), but also the likelihood of the brand entering into future purchase decisions. Thus, a forward-looking consumer will take availability of the brand into account when faced with the opportunity to purchase the brand for the first time.

It is shown that the greater the size of a brand's retail chain—or, more precisely, the better it covers the consumer's travel pattern—the more opportunity the consumer has to use any information acquired about the brand. As the chain grows, the willingness of the consumer to try the brand for the first time increases. Formally stated, the probability of first-purchase is convex in the size of the retail chain. As such, there exist increasing returns to chain size.

In the discrete geographic space of the present model, the increasing returns take a rather severe form: a "minimum-informative scale" for the chain below which the value of information will not cover the consumer's opportunity cost of trying the brand. Thus, until the brand's chain has reached this minimum-informative size, no sales will be made.

The model also shows that being first into a market has value to a brand. The cost of trying a new brand is the value of the alternative forgone. Being first in a market means being there before consumers have alternatives. Without alternatives, there is little opportunity cost to trying the brand. Consumers are more willing to experiment with a new brand, if they do not have to forego a known brand. In fact, it may be possible for an incumbent brand to create loyal customers by "preempting their beliefs." In particular, consumers may believe the likelihood of improving their future purchase decisions through experimentation with a new brand is small, if the incumbent brand is already offering them a product of sufficiently high value. As a result, consumers may appear to purchase the brand habitually, unwilling to experiment with a new brand even though it may offer them a product of higher value.

Obviously, the identification and understanding of the factors which influence the value of consumer experimentation are extremely important; for before a brand can have a loyal customer, it must have a customer and all purchases start with the first. Economies of scale are limited by the extent of the market (Smith 1976), and the extent of the market for a retail chain is determined by the extent to which the chain can get inexperienced consumers to try the product. As will be shown, it is the sharing of a common brandname across spatially distributed stores which creates the market by inducing mobile consumers to try the brand for the first time. Until the market is created, few sales can be made, and the scale-economies which are commonly attributed to the success of retail chains cannot be had. The seminal contribution of scale is, therefore, on the demand side: through the sharing of a common brandname new customers are attracted, the market is extended, and the limit on cost-side economies is thereby lifted.

3.1.2 Geographic Space, Convenience, and Consumer Mobility

As noted previously, while the relationship between transportation costs and spatial distribution is quite simple, the relationship between transportation costs and spatial branding is not. For the same transportation costs which are necessary for the existence of a set of spatially-distributed stores, results in the consumer buying her product from only one store, thereby leaving little role for spatial branding.

The missing piece of this puzzle is quite simply this: Although consumers will not travel far to obtain convenience goods, they do, nonetheless, travel. They travel to go to work, to take vacations, to obtain capital goods; they travel for a variety of reasons, the majority of which are entirely independent of the demand for convenience goods. This observation is fundamental to the theory outlined herein; as such, it is formalized in the following axiom:

Axiom 3.1 ***Mobility and Demand.*** *For convenience goods, the consumer's travel pattern is primarily exogenous to (i.e. separable from) the demand for the good.*

Axiom 3.1 implies that, to understand the demand for convenience goods, we must downplay their importance in the consumer's travel behavior. In particular, the traditional view that a consumer leaves home to go get the good, and returns home after purchase, is a simplification which hides the importance of space in spatial branding. This traditional view is built around the assumption that consumers have a most-preferred location in geographic space—e.g. their home. And it is the distance of stores from this most-preferred location which, in part, determines from which store the consumer buys the product.

In making this assumption, the consumer's preferences over store location are being treated like her preferences over all other attributes of the product. However geographic space is not just another attribute-space for the product. Geographic location is an attribute common to *all* products. As such, it cannot be held fixed when analyzing the purchase decision for one product when the consumer obviously travels to purchase other products.

This is especially true for convenience goods. When consumers travel for reasons exogenous to the demand for the good in question, their most-preferred location varies, and they take their demand for the good with them. To give the most obvious example: when you go to work, or take the kids to hockey practice, or anywhere else, your demand for gasoline goes with you. People almost never leave home solely for the purpose of purchasing gasoline from the closest store—it is a convenience good, you purchase it when you need it and where it is convenient.

To assume consumers have a most-preferred location which is fixed is to treat geographic space like all other preference spaces. While many attribute spaces for a product may be amenable to partial-equilibrium analysis—being separable from other activities of the consumer—geographic space in the analysis of convenience goods is not. It is the exogeneity of the consumer's travel pattern which is the primary factor explaining the use of spatial branding in spatial distribution. When consumers travel for reasons exogenous to the demand for the good, they take their demand for the good with them, and find the information conveyed by spatial branding useful in their spatially-distributed purchase decisions.

The concept of consumer mobility outlined here, and formalized in this and the following chapter, leads to an expanded theory of consumer choice. Consumers no longer choose the single store which is closest to their fixed most-preferred location; rather, they choose the brand whose retail chain comes closest to covering their travel pattern. The comparative advantage of retail chains lies in serving the needs of mobile consumers. Because consumers travel for reasons exogenous to the demand for the good, they find spatial distribution convenient; and because they purchase from more than one store, they find spatial branding informative. One of the goals of this thesis is to demonstrate that the primary explanatory factor in the use of spatial branding and, therefore, the existence of retail chains, is the exogeneity of the consumer's travel pattern from the demand for the product.

3.1.3 Some Simplifying Assumptions

This chapter builds on Axiom 3.1, formally modelling the concept of a convenience good by taking the consumer's travel pattern to be entirely exogenous to the demand for the good.[1] This assumption is relaxed in the model of Chapter 4; but for now it is instructive, taking the concept of convenience to its limit and, thereby, providing bounds on the observable possibilities. For many consumers, however, this assumption may not lie too far from the truth:

> You will find many customers are lazy—they will not move an extra foot, they will not make an extra turn, they will not wait an extra minute in line, and they will not compromise. .... The understanding of convenience, particularly in this era, is the greatest opportunity for all business, not just retailers. (Salvaneschi 1996, p. 117)[2]

The model of this chapter adapts the notation and adopts the assumptions of the previous chapter, while incorporating the following additional structure:

Assumption 3.1 ***Consumer Mobility.*** *The consumer's travel pattern is characterized by a uniform distribution over a finite number, n, of locations—thus, each of the n locations has an equal probability of being visited.*

Assumption 3.2 ***The Distribution of Stores.*** *Brand b is assumed to serve all locations in the consumer's travel area, whereas Brand a serves only a nonproper*

[1] The assumption of a entirely exogenous travel pattern effectively defines the limiting concept of spatial branding. Much like the absence of product differentiation defines a perfectly competitive market structure, or the absence of sunk investment defines a perfectly contestable market structure, the absence of exogenous travel defines a perfectly temporally-branded market structure. As the exogeneity of consumer travel increases, so too does the role played by spatial branding.

[2] Salvaneschi oversaw the opening of 5,000 stores as the former Vice-President in charge of real estate administration for McDonald's; of 1,200 stores as the former Senior Vice-President responsible for new markets and locations for Kentucy Fried Chicken; and of 1,700 stores as the former President of Blockbuster Entertainment Corp.

subset of these. Thus, $1 \leq n_a \leq n_b = n$, *where* n_a *and* n_b *are the number of stores in each brand's retail chain.*

As in the model of Chapter 2, the consumer makes one purchase per period, but by now selecting one of the brands from those available at her current location before moving on to her next location.[3] Of interest is the relationship between the size of Brand a's retail chain, or, more specifically, the coverage the chain provides for the consumer's travel pattern, and the resulting information-value such coverage generates for the consumer. Given the consumer's travel pattern, the probability the consumer travels to a location served by Brand a is $c_a \equiv n_a/n$.[4] This probability is a measure of Brand a's coverage of the consumer's travel pattern and is assumed to be known by the consumer. Such knowledge is at least imparted to the consumer through her travels, but other possibilities clearly exist.[5]

3.1.4 Finding an Optimal Policy

The first modification imposed by geographic space and consumer mobility is with respect to the constraints on an optimal policy (formerly given by Equation 2.3). In

[3] While Brand b will be formally interpreted as an incumbent chain which the consumer has tried previously, there is, however, an alternative interpretation for Brand b which the reader may find more appealing. In particular, Professor Robert A. Jones has suggested Brand b might be thought of as the label the consumer gives to a set of independent single-store brands, one per location, which the consumer treats as a group offering a mean surplus value of μ_b. Such an interpretation can only be formally justified, however, if the consumer visits a very large number of locations so that sampling with replacement (i.e. forgetting about a particular independent) is a reasonable approximation.

[4] It follows that the probability she visits a location served only by Brand b is $1 - c_a$.

[5] For example, the consumer may become aware of stores at locations she has not yet visited through the newspaper or television advertising of the chain. However, it should be clear that the role of advertising in this regard is not formally required in this model and, therefore, will not be further examined here. In particular, it need not be assumed that the consumer has perfect information about the true coverage provided by each brand. The argument made herein is strengthened, for example, if the true coverage is taken as an upper bound on the consumer's belief about coverage. It is required, however, that the consumer's belief about coverage is positively dependent on the true coverage.

particular, the additional structure implies that an optimal policy will be a member of a family of policies which satisfy the condition:

$$\pi^* (x''_a, \lambda) = \begin{cases} a & \text{if} \quad (x''_a = 1_{\mu_a+\sigma_a}) \wedge (\lambda \in \Lambda_a) \\ b & \text{if} \quad (x''_a = 1_{\mu_a-\sigma_a}) \vee (\lambda \notin \Lambda_a) \end{cases} \tag{3.1}$$

where Λ_a will be used to denote the set of locations served by Brand a and where $\wedge$ and $\vee$ denote the logical operations "*and*" and "*or*", respectively. Equation 3.1 says that if the consumer knows Brand a to be of high utility-value and finds herself at a location served by Brand a, she will purchase Brand a, but Brand b otherwise. Our task is to find the remaining portion of the optimal policy: that which applies to an inexperienced consumer with prior beliefs $x'_a = 1/2_{\mu_a \pm \sigma_a}$ at a location $\lambda \in \Lambda_a$.

To this end, let $V^* (x_a, \lambda)$ denote expected present-value of an optimal policy for a consumer with beliefs x_a who finds herself at location λ. Almost surely the consumer will eventually find herself at a location served by Brand a, and there she will face her first brand-selection decision. Thus assume $\lambda \in \Lambda_a$. In finding an optimal policy we can, as before, restrict our attention to policies which choose either Brand a or Brand b in the current period and optimally thereafter. Let $V_a^* (x_a, \lambda)$ and $V_b^* (x_a, \lambda)$ denote, respectively, the expected present-values of these policies. These values may be expressed somewhat more explicitly. To this end, let $u_k (x_a, \lambda)$ denote the expected current utility derived from the selection of Brand $k \in \{a, b\}$ under beliefs x_a at location λ. Then we can write,

$$V_k^* (x'_a, \lambda) = u_k (x'_a, \lambda) + E_{x''_a|k,x'_a} E_\lambda V^* (x''_a, \lambda) \tag{3.2}$$

where E_λ denotes the expectation of λ over Λ and $E_{x''_a|k,x'_a}$ is as defined previously in Equation 2.4. As in the last chapter, the values of these policies can be found as follows:

- **Policy** A : Choose Brand a in the current period, and optimally thereafter. Thus,

$$\begin{aligned} V_a^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) &= u_a\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) + E_{x_a''|a,1/2_{\mu_a \pm \sigma_a}} E_\lambda V^*\left(x_a'', \lambda\right) \\ &= \mu_a + \frac{1}{2} E_\lambda V^*\left(1_{\mu_a+\sigma_a}, \lambda\right) + \frac{1}{2} E_\lambda V^*\left(1_{\mu_a-\sigma_a}, \lambda\right) \end{aligned} \tag{3.3}$$

If, after trying Brand a, the consumer finds it to be of high value, it is optimal to choose it whenever it is available in future periods. So $E_\lambda V^*\left(1_{\mu_a+\sigma_a}, \lambda\right) = \left(c_a\left(\mu_a + \sigma_a\right) + (1 - c_a)\,\mu_b\right)/i$. However, if the consumer finds Brand a to be of low value, then it is optimal to choose Brand b thereafter. So $E_\lambda V^*\left(1_{\mu_a-\sigma_a}, \lambda\right) = \mu_b/i$. Thus the value of this policy is

$$V_a^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) = \mu_a + \frac{c_a\left(\mu_a + \sigma_a\right) + (2 - c_a)\,\mu_b}{2i} \tag{3.4}$$

- **Policy** B : Choose Brand b in the current period, and optimally thereafter. Thus,

$$V_b^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) = u_b\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) + E_\lambda E_{x_a''|b,1/2_{\mu_a \pm \sigma_a}} V^*\left(x_a'', \lambda\right) \tag{3.5}$$

Under this policy, the beliefs of the consumer do not change, so $E_{x_a''|b,1/2_{\mu_a \pm \sigma_a}} V^*\left(x_a'', \lambda\right) = V^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right)$. Further, it follows that if it is optimal to choose Brand b in the current period, it is also optimal thereafter since the future looks identical to the present. Thus, under this hypothetical, $E_\lambda V^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) = \mu_b/i$, so the value of this policy is

$$V_b^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) = \mu_b + \frac{\mu_b}{i} \tag{3.6}$$

It follows that Brand a is optimal if

$$\begin{aligned} V_a^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) &\geq V_b^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) \\ \mu_a + \frac{c_a\left(\mu_a + \sigma_a\right) + \left(2 - c_a\right)\mu_b}{2i} &\geq \mu_b + \frac{\mu_b}{i} \\ c_a &\geq \frac{2i\Delta}{\sigma_a - \Delta} \equiv c_a^{\min} \end{aligned} \tag{3.7}$$

where the reader will recall that $\sigma_a > \Delta > 0$ (Parametric Restriction 2.1). The optimal policy thus dictates the following behavior for the inexperienced consumer:[6]

$$\pi^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) = \begin{cases} a & \text{if} \quad \left(c_a \geq c_a^{\min}\right) \wedge \left(\lambda \in \Lambda_a\right) \\ b & \text{if} \quad \left(c_a < c_a^{\min}\right) \vee \left(\lambda \notin \Lambda_a\right) \end{cases} \tag{3.8}$$

This policy is best understood in terms of the information value contained in a purchase of Brand a. Under the assumption of $\Delta > 0$, the expected net present-value of information contained in such a purchase is

$$V^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) - V^m\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) = \begin{cases} \frac{\sigma_a - \Delta}{2i} c_a - \Delta & \text{if} \quad c_a \geq c_a^{\min} \\ 0 & \text{if} \quad c_a < c_a^{\min} \end{cases} \tag{3.9}$$

[6] The optimal purchase policy, stated in its entirety, could be expressed as follows:

$$\pi_t^*\left(x_a, \lambda_t\right) = \begin{cases} a & \text{if} \quad S_t\left(x_a, c_a^{\min}\right) \wedge \left(\lambda_t \in \Lambda_a(t)\right) \\ b & \text{if} \quad \text{otherwise} \end{cases}$$

where λ_t is the location of the consumer at time t, $\Lambda_a(t)$ is the set of locations served by Brand a at time t, and $S_t\left(x_a, c_a^{\min}\right)$ is the following (Boolean) statement:

$$S_t\left(x_a, c_a^{\min}\right) \equiv \left(\left(x_a = 1/2_{\mu_a \pm \sigma_a}\right) \wedge \left(c_a(t) \geq c_a^{\min}\right)\right) \vee \left(x_a = 1_{\mu_a + \sigma_a}\right)$$

where $c_a(t)$ is the coverage of the consumers travel pattern provided by Brand a in period t and where $c_a^{\min} \equiv 2i\Delta / \left(\sigma_a - \Delta\right)$. This statement is true if either *(i)* the consumer is inexperienced and Brand a chain is of minimum informative-scale, or *(ii)* the consumer is experienced and Brand a is of high utility-value. Implicity in this statement is the assumption that Brand a cannot commit to opening stores in the future. The consumer therefore takes the coverage of Brand a in future periods to be that provided by the set of stores currently in operation.

which is obviously convex in c_a.[7] This result is summarized in the following proposition:

Proposition 3.1 ***Convexity of the Value of Information.*** *The value of information contained in a purchase of an experience brand, such as Brand a, is increasing and convex in the size of its retail chain.*

The return a consumer derives from experimentation with an experience brand is increasing in the size of the brand's chain. As argued in the last chapter, the consumer effectively holds a perpetual option on the purchase of Brand a, but she does not know if the option is "in the money" (i.e. if Brand a is of high-value). To find out, she must pay the current-period premium $\Delta \equiv \mu_b - \mu_a$. However, the introduction of geographic space and consumer mobility has put restrictions on the consumer's ability to exercise her option and, therefore, the value of this option. As the size of Brand a's chain increases, so too do the opportunities for the consumer to exercise the option. As a result, the expected value of the option increases with the size of Brand a's chain. Obviously, the expected value of the option on Brand a must exceed current-period premium of Δ (the expected cost of exercising the option) to warrant the consumer trying the brand for the first time.

The convexity of the value of information in c_a means there exists increasing returns to chain size. In the discrete-space model of the present chapter, this result takes the rather severe form of a "minimum informative-scale" for Brand a's retail chain. Before a consumer will try Brand a, the chain must reach the minimum scale necessary to make the information obtained through its purchase sufficiently valuable to forego the expected current-period return of Δ.[8] The consumer-behavior described

[7] The reader is invited to compare expression 4.8 with its sister, expression 2.11, on page 31.

[8] The idea that there are economies of scale in the use of information was perhaps first noted by Arrow (1974) (although not in this context). Later, Radner and Stiglitz (1984) formalized the concept and Wilson (1975) has used it to argue that a competitive equilibrium could not exist in its

by the optimal policy is summarized in the following proposition:[9]

Proposition 3.2 ***Minimum Informative-Scale for a Retail Chain.*** *Within the context of the present model, a consumer will only experiment with an experience brand (i.e. try it for the first time) if the brand's retail chain covers a sufficient portion of her travel area. In particular, there exists a minimum informative-scale for the retail chain which is increasing in Δ and i, and decreasing in σ_a.*

3.1.5 Loyal Customers (or Habitual Purchases) and Market Preemption

Sometimes consumers appear to purchase the same brand habitually. From the point of view of the brand, such customers are merely loyal. These customers will not try new brands that come on the market, even though such brands may be of higher utility-value. This behavior can be generated within the context of the present model. In particular, from Equation 3.7 there exists a utility-value for Brand b (the incumbent brand), say μ_b^L, which will induce loyalty in a customer such that she will not try Brand a (the new brand) regardless of the size of its chain. To see this, let $\Delta^L \equiv \mu_b^L - \mu_a$ denote the minimum differential which will induce customer loyalty. Obviously Δ^L is defined by the condition that $c_a^{\min} = 1$, so even full coverage by Brand a is not able to induce the customers of Brand b to experiment with Brand a. Thus,

$$\frac{2i\Delta^L}{\sigma_a - \Delta^L} = 1 \tag{3.10}$$

presence.

Risk aversion only strengthens this result, increasing the size of the minimum informative-scale, $c_a^{\min}$. In particular, let R be the risk premium on Brand A and let $\Delta' \equiv \Delta + R$. Substituting Δ' in place of Δ in the definition of $c_a^{\min}$ contained in Equation 3.7 simply increases its value.

[9]Note that $\partial c_a^{\min}/\partial\Delta$ and $\partial c_a^{\min}/\partial i$ are both positive, while $\partial c_a^{\min}/\partial\sigma_a$ is negative.

so

$$\mu_b^L = \mu_a + \frac{\sigma_a}{1+2i} \tag{3.11}$$

Note that $\mu_b^L < \mu_a + \sigma_a$ for $i > 0$, so that even though Brand a could give the consumer higher utility-value, the consumer will not try it, even in the absence of risk aversion. Brand b has effectively preempted the beliefs of consumers: by providing consumers with a product of sufficiently high value, consumers do not believe there is much to be gained by experimenting with other brands, such as Brand a. The result is summarized by the following proposition:

Proposition 3.3 ***Brand Loyalty (Preemption of Consumer Beliefs).*** *A brand can obtain the loyalty of its customers by entering a market early and offering them a product of high utility-value. By doing so, such a brand can preempt the beliefs of consumers, making them unwilling to try the products of subsequent entrants even though such products may yield consumers a greater surplus.*

One might wonder how a brand of lower quality can survive competition from higher-quality brands. The answer is simple, if inexperienced consumers never try the brand, they will not know its quality is high. Experimentation with an unknown brand has value for future purchase decisions. Naturally, the greater the value of experimentation, the greater the value of the known brand required to induce loyalty. Thus, as the degree of prior uncertainty σ_a increases, or the rate of discount i falls, the value of the known brand μ_b which will induce loyalty rises.[10]

[10]Building on the work of Schmalensee (1982), Bagwell (1990) models the ability of a entrant with uncertain quality to penetrate a market served by an incumbent of known quality. While his model is nonspatial and concerned with price rather than chain size, similar results are obtained in that the incumbent may be able to deter entry of a brand offering greater consumer surplus.

The theory of spatial branding has rather mixed implications for entry deterrence. Entry of a new brand into a geographic area is governed by competition for inexperienced consumers. On the one hand, the theory predicts that consumers are likely to be more willing to experiment with new brands entering the market for branded convenience goods than perhaps they would be in other markets where information value is less important and risk aversion is more important. On the other hand, the minimum-informative scale (Proposition 3.2) and brand loyalty (Proposition 3.3) results derived from this theory make entry more difficult for a new brand, even in the absence of risk-aversion.

3.1.6 Robustness and Market Saturation

The model above has relied on the assumption of an incumbent brand offering a product of known value and serving all locations. This assumption results in the consumer's opportunity cost of acquiring information about Brand a being both positive and constant over locations, and therefore over time. If, however, the opportunity cost of trying Brand a varies from period to period—say because *(i)* some locations are served by another incumbent which the consumer values differently than Brand b or because *(ii)* the consumer's travel pattern results in her distance from Brand a varying from period to period (see, e.g., the Hotelling-model of Chapter 4)—the analysis is somewhat more complicated, but the central proposition of this chapter (Proposition 3.1) still obtains.

The same cannot generally be said, however, for Propositions 3.2 and 3.3. It should be noted that these propositions remain valid even if Brand b does not serve all locations, so long as such locations are served by other brands of equivalent or higher value. However, these propositions no longer hold if locations not served by Brand b are served by another brand which the consumer values less than or equal to

μ_a (or if they are not served by any brand at all). The reason is that such locations present the consumer with an opportunity to acquire information about Brand a at zero opportunity cost since she need not forego an alternative brand with value greater than μ_a. As a result, by establishing a store at such a location, Brand a will obtain positive sales without having to attain minimum informative scale.[11] Further, if Brand a is preferred to the incumbent brand, consumers which would otherwise have remained loyal and not tried the new brand, will buy it wherever it subsequently becomes available, including now locations which the incumbent does serve. As might be expected, an incumbent's chain is only as strong as its weakest link, and its weakest link is locations which it does not serve. Perhaps this explains why many chains are ubiquitous, saturating the market by establishing stores at locations which might otherwise be considered marginal.

[11]Note, however, that all first-time purchases will have to be made through these locations. Only after consumers have tried Brand a at locations where it has a monopoly will they (possibly) go on to purchase it at locations where it competes with Brand b. Further, there still exists a minimum-informative scale for first-time purchases at duopoly locations. What does it look like? First, recall that the cost of obtaining information about Brand a is the expected surplus the consumer foregoes by not purchasing Brand b (i.e. $\Delta \equiv \mu_b - \mu_a$). Thus, locations are served only by Brand a (or served by a brand which the consumer values less than μ_a) present the consumer with an opportunity to obtain information about Brand a at zero cost. Since perfect information is acquired after just one purchase, this opportunity to try Brand a at zero cost is foregone should she first try Brand a at a location served by both brands. Thus, the opportunity to try Brand a at zero cost must be valued in her decision-problem to try Brand a at locations served by both brands. When this is done, expression 3.7 becomes

$$c_a \geq \frac{2\Delta}{\sigma_a - \Delta}\left(i + c_a'\right) \equiv c_a^{\min}$$

where c_a is the proportion of locations served by both brands and c_a' is the proportion of locations served only by Brand a. Quite naturally, the increased opportunity cost of trying Brand a at duopoly locations results in an increase in the minimum informative scale necessary to induce the consumer to try Brand a at such locations.

3.2 Implications for Locational Choice

Why are retail chains located along major roads and highways, in bus terminals, shopping malls, and airports? This thesis argues that spatial distribution is of primary importance in the provision of convenience goods. The characteristics of this class of goods explains the structure of the market which retail chains have evolved to serve. Although consumers will not travel far to obtain convenience goods, they do, nonetheless, travel. So to make convenience goods convenient, producers of such goods must locate their stores along the primary travel routes of consumers. Since experience information has value to consumers, it can be expected that competition will exist among producers to create and provide such information. Retail chains create information-value by having a spatially distributed set of stores which share a common brandname, and by locating those stores in the locations of highest consumer mobility.

A direct implication of the theory of spatial branding is that independent single-store operators, who are not spatially branded, will have a very difficult time competing for the custom of mobile consumers. They simply do not provide the coverage necessary to induce mobile consumers to experiment with the product. However, not all consumers are highly mobile, some are relatively sedentary. The model of this chapter also allows the purchase decision of less mobile consumers to be analyzed as a special case.

In the context of the present model, a non-mobile consumer can be defined as one who has a fixed most-preferred location and, therefore, only visits stores within her community. As such, an independent single-store brand covers the travel pattern of a non-mobile consumer just as well as a brand with a very large number of stores. Formally, the purchase policy of a nonmobile consumer can be found from that for a mobile consumer by setting $c_i = 1$, for any brand i which serves the consumer's location. As a result, the information contained in a purchase from any store has the same opportunity of being used, whether that store be an independent single-store operation or a member of a chain which covers the country.

To avoid competition with independent single-store operators, chains typically locate their stores in areas of high mobility. The comparative advantage of a retail chain lies in the market for mobile consumers typically found at "mobile locations." Not all locations are created equal. The mobility of the population at some locations is greater than that at others. For example, locations in bus terminals, airports, and along highways can be expected to have a much more mobile population than do locations "off the beaten trail," such as those found in isolated small towns or on the back roads deep in the interior of a community.

To a consumer who never leaves her community, spatial branding has no value; all branding is temporal. Recall again the thoughts of Professor Akerlof (1970, p. 500) noted in the previous chapter:

> Chains—such as hotel chains or restaurant chains—are similar to brand names. ... These restaurants, at least in the United States, most often appear on interurban highways. The consumers are seldom local. The reason is that these well-known chains offer a better hamburger than the *average* local restaurant; at the same time, the local customer, who knows his area, can usually choose a place he prefers. [italics in original]

Akerlof is correct: chains are similar to brand names, but so are independent single-store operations: they are branded too, albeit purely temporally. Consumers are not born with a knowledge of local brands; nor are they endowed with such a knowledge when they move to a new community. As noted in Chapter 1, by assuming the consumer is more familiar with one brand than with the other, the argument of familiarity puts the cart before the horse. Until it can be explained how the consumer came to be more familiar with one brand, this argument simply begs the question. A complete theory of branding must explain how a brand comes to be adopted by the consumer in the first place.

The theory of spatial branding completes this argument, for it explains why the local customer knows his area. Knowledge of the local area will find many opportunities for future use by a local customer, but not by one who is not local. Thus, experimentation with local brands is an investment worth making for a local customer, but not for a transient customer who is unlikely to return, or returns with less frequency.

3.3 Appendix: Computer Simulation

Locations are differentiated, and geographic space matters, but not in the traditional sense. Consumers are not defined by a unique location in geographic space, but rather by a unique travel pattern over that space. Firms compete for consumers in the space of consumer mobility by placing their stores at locations which provide the greatest coverage of the consumer's travel pattern. As the present chapter has attempted to show, such coverage generates not only convenience-value for the consumer but also information-value, and the latter of these values, unlike the former, give rise to increasing returns to scale for the chain.

Many of the simplifying restrictions of the present model have been relaxed using computer simulation, allowing the results of this chapter and this thesis to be extended in a number of ways. The model used in computer simulation will be very briefly outlined here.

The model is a variation of a true two-arm bandit problem with each brand paying 0 or 1 with unknown probability p_i, $i = a, b$. Consumers revise their beliefs about each brand on the basis of the experience with the brand according to Bayes theorem. Since the sampling process is Bernoulli, a conjugate prior from the Beta family of distributions is used so that the posterior is also a member of the Beta family. The model is a variant of this standard formulation in the following respects:

1. The model is made very general by interpreting 0 and 1 not as payoffs, but rather product types; thereby allowing for the analysis of both vertical and horizontal differentiation. If all consumers prefer to receive a 0 (or a 1) then the model is one of pure vertical differentiation; whereas if some consumers prefer to receive a 1 and others prefer to receive a 0, the model is one of horizontal differentiation.

2. Consumers are distributed across travel patterns (mobility space). A travel pattern is described by a Markov "mobility matrix", with elements of the matrix giving the consumer's probability of moving from one location to another. Different consumers have different travel patterns.

3. Consumer mobility results in their purchase decisions being distributed over time and space. Brands have less than perfect coverage of the market, thus not all arms (brands) are available at each decision epoch. This is a fundamental departure from the standard bandit model.

3.3.1 Solving for an Optimal Policy

There are two basic approaches to solving for an optimal policy to a bandit problems:

> "The variety of policies that have been suggested for the multi-armed bandit reflect the diversity of criteria by which such policies can be judged. The simplest criteria place the problem in an optimization framework by postulating a prior distribution for the unknown parameters and maximizing the expected sum of rewards over a finite horizon, or the discounted sum of rewards over an infinite horizon. The backwards induction algorithm is available for the finite horizon formulation; the attraction of the discounted formulation is the stationarity of intertemporal comparisons implied by it. These criteria place zero or finite weight on the infinite future; in contrast, the asymptotic optimality requirement puts zero weight on any finite time interval. Asymptotic optimality is an appealing property, but as a criterion its exclusive concern with tail behavior seems to throw away more than just the bathwater."(Bather 1981, Dr. F. P. Kelly, p. 285 of the discussion of Professor Bather's paper)

The computer model uses a finite-horizon stochastic dynamic-program built around a modification of the Gittins Index Theorem (Gittins 1979) to solve for an optimal policy. Optimality is defined over a finite horizon with a positive discount rate. As such, the model does not throw out the baby. However, so as not to place zero weight on the indefinite future, policies which approximate optimality over periods longer than that used for the backwards induction algorithm were obtained through terminal-date approximations adapted from Berry and Fristedt (1985).

The Gittins Index Theorem allows the complex two-brand bandit problem to be separated into two one-brand bandit problems. Each of the one-brand problems is solved separately to obtain an index function for the brand in question. The index function gives the value of the brand for all possible consumer beliefs and locations—much like assigning a certainty equivalent value in a problem of pure risk, except that the index value includes the value of information. Under the Index Theorem it is then optimal to select whichever brand has the highest index value, given the consumer's beliefs (purchase experience) and location. A graph of an index function generated by the computer program is shown in Figure 3.1.

A purchase from a brand can result in "success" or "failure". The number of successes obtained in previous purchases from the brand is measured by the a-axis, the number of failures by the b-axis. The index function is increasing in successes but decreasing in failures, and includes a measure of the value of information to be obtained from another purchase. A population of consumers with various travel patterns (Markov mobility matrices) travel around a fixed number of locations, each independently selecting the brand at their current location which has the highest index value. Time paths of sales similar to those derived in Chapter 5 are easily generated.

Unfortunately, the Index Theorem of Gittin's only proves optimality of the index-policy in the standard bandit problem where all arms (brands) are available for selection at every decision epoch. However, the problem of spatial branding differs from

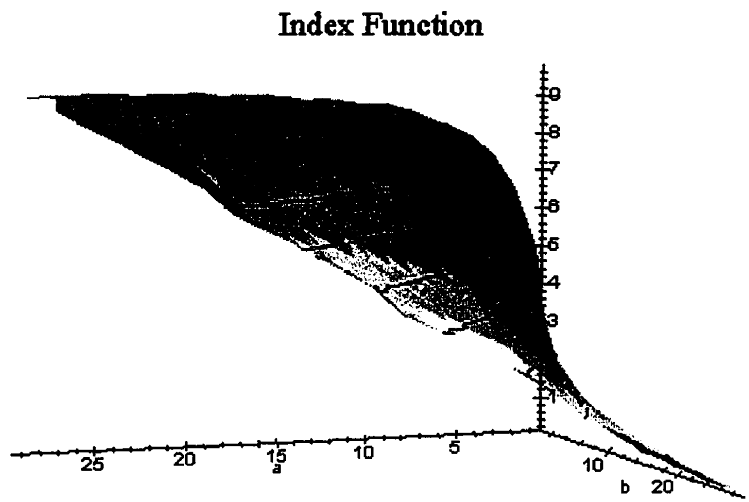


Figure 3.1: The Modified Gittins Index Function

the standard problem in that the availability of the brand depends on the size of the chain and the consumer's travel pattern. Because of this difference, optimality of an index-policy has not been generally proven for this modified problem (although the program has been validated by reproducing results published by both Gittin's and Berry-Fristedt for the standard bandit).

There is, however, a correspondence between mobility matrices and possible discount sequences. As such, it is conjectured that such a proof is possible for a restricted set of mobility matrices which satisfy the regularity condition of Berry and Fristedt noted earlier in footnote 10 on page 29. Unfortunately, a formal proof of a Modified Index Theorem for the mobile consumer will have to be postponed for future work—as will the model itself.

Chapter 4

The Theory of Spatial Branding and Retail Chains in Continuous Geographic Space: A Hotelling Model

4.1 Introduction to the Model

In the present chapter the model of the last chapter is further extended, in the Hotelling tradition, to a continuous and uniform geographic space represented by a circle of unit circumference. The expanded concept of geographic space is used to more fully develop the formal concepts of convenience and consumer mobility.

4.1.1 Extending the Concepts of Consumer Mobility, Convenience, and Information Value

Recall the argument of Section 3.1.2. Spatial distribution is of particular importance in the sale of convenience goods: inexpensive items for which transportation (and other inconvenience) costs are a significant fraction of the gross social surplus derived from the good. Even though consumers will not travel far to obtain such goods, they do nonetheless travel. And when they travel, they take their demand for convenience goods with them. Spatial branding is of particular importance because consumers are mobile, travelling for reasons exogenous to their demand for convenience goods (Axiom 3.1). As a result, these consumers find spatial distribution convenient, and spatial branding informative. The comparative advantage of retail chains lies in the provision of convenience goods to mobile consumers.

The model of Chapter 3 formalized Axiom 3.1 with the heuristic assumption that the consumer's travel pattern was *entirely* exogenous to such demand. This assumption usefully illustrated the concept of convenience by taking it to its limit. The model of the present chapter uses a more general formalization of geographic space to relax this assumption and, thereby, more fully capture Axiom 3.1. It does so by decomposing consumer mobility into primary and secondary travel patterns. Quite naturally, this expanded conceptualization of consumer mobility gives rise to an expanded conceptualization of convenience, which will be shown to now include both primary and secondary forms.

Expanding the concepts in this way yields a fuller understanding of retail chains and the information value they create for mobile consumers. However, these expanded concepts do present a problem for finding an optimal policy. In particular, the argument that the future looks like the present can no longer be used in the solution to the optimal policy (recall footnote 10). Nonetheless the optimal policy is found and shown to generate one of the primary results of this thesis: that the local market share of each store in a retail chain is convex in the size of the chain, i.e. there are increasing returns to scale.

4.1.2 Geographic Space and the Distribution of Stores

Consider a city represented by a circle of unit circumference. Throughout the circular city, stores from each of the two brands are assumed to be distributed, with locations determined exogenously. To simplify the model, a certain amount of symmetry must be imposed. In particular, assume Brand b has n_b stores which are uniformly distributed around the city. In the present model it will be useful to refer to the set of locations contained within the geographic area between any two stores from Brand b as a *community* and denote such a set by Λ. As such, the size (i.e. geographic length) of each community is identical and equal to $1/n_b$. As in the model of the previous chapter, Brand a will serve a subset of these communities. Thus, $1 \leq n_a \leq n_b$. If Brand a serves a community, its store is assumed to be positioned at the midpoint between the two stores from Brand b.

These assumptions imply two types of communities: those served by both brands, and those served only by Brand b. Further, the symmetry imposed on the space means that particular communities within each type are identical and need not be distinguished. Thus, let Λ_a denote a representative community served by Brand a and let Λ'_a denote one which is not. As well, it will be useful at times to let C denote the unit circle—a city comprised of the set of all communities. Thus, Λ_a and Λ'_a are representative elements of C. Finally, the consumer's location can now be defined by her distance λ from the midpoint of the community. As a result, in communities served by Brand a, λ is the distance of the consumer from Brand a. An example of a city with these geographic features is depicted in Figure 4.1.

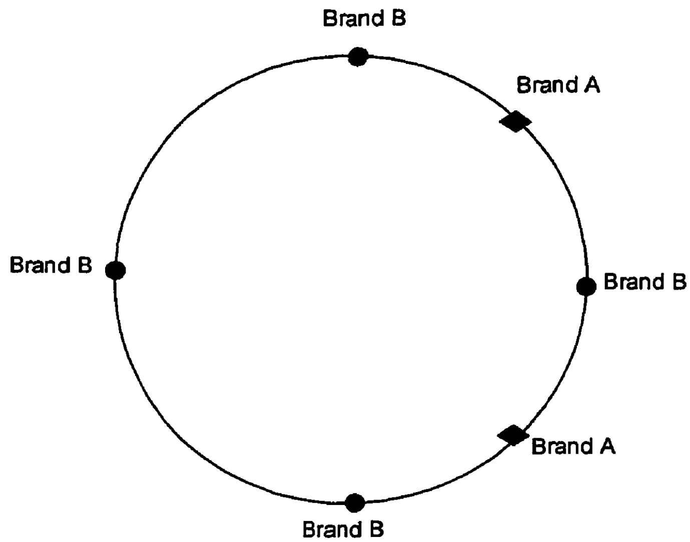


Figure 4.1: A Circular City with Four Communities, Two Served by Brand a

4.1.3 Consumer Travel Patterns

Consumers travel around the circumference of this circular city. However, the movement of consumers is somewhat different from that traditionally assumed. Consumers have no unique location and their travel pattern is less than fully dependent on the demand for the product. In particular, consumers are assumed to make two types of trips: primary and secondary.

Primary trips (formally identical to those modelled in the previous chapter) are entirely independent of the consumer's demand for the product, whereas **secondary trips** are solely for this purpose. For example, the primary purpose of a trip may be to get to work, or take the kids to school. While on a primary trip, the consumer might find it convenient to stop off and pick up the product. This is the secondary trip. Similarly, a travelling salesman may be in an out-of-town airport and desire a room to stay in for the night. Obviously the salesman's primary trip is for the purpose of business, but his secondary trip is to obtain accommodations. In both of these examples, the primary travel pattern is independent of the demand for the product, while the secondary travel pattern is fully dependent on this demand and, therefore, identical to that usually assumed in the traditional Hotelling model.[1] Formally, the consumer's travel pattern will be represented as follows:

Assumption 4.1 ***Consumer Mobility.*** *The consumer's primary travel pattern is assumed to be described by a uniform density over the unit circle. Thus, regardless of the consumer's current location, all points on the unit circle have an equal probability of being visited next period. After her primary trip takes her to a new community, she will make a secondary trip to one of the two stores closest to her within that community. The cost per unit distance of secondary travel is assumed constant and equal to* $t \geq 0$.

4.1.4 Consumer Utility

The consumer's beliefs about each brand are unchanged from the previous models. So long as consumer surplus from a brand is expected to be positive and greater than that which can be obtained from other brands, she travels to buy the brand and spends a fixed amount in its purchase. Thus, if, for example, through her primary travels, she found herself immediately next to a store from Brand a, a purchase from this brand would yield an expected surplus of $\bar{x}_a$, the mean of the distribution describing her beliefs about Brand a. Her expected consumer surplus declines from there as her distance from Brand a increases, until it reaches a value of zero, or until it falls below

[1] The reader might conceptualize primary travel as occuring across the interior of the circular city along highways (not shown), and secondary travel occuring on secondary roads along the circumference of the circle. Though perhaps useful, this metaphor will not be formally used here.

the surplus expected to be received from a purchase from Brand b, whichever comes first.

More formally, in communities served by Brand a, there will be some maximum distance that the consumer will travel to purchase Brand a. Let this critical distance be λ^k. Further, let the function $u_a(x_a, \lambda) = \bar{x}_a - t\lambda$ give the current expected-utility received by a consumer with beliefs x_a at location λ who chooses Brand a. Finally, let $u_b(\lambda) = \mu_b - t\left(\frac{1}{2n_b} - \lambda\right)$ denote a similar function for Brand b.

Given the linearity of travel costs, the use of piecewise functions can be minimized by imposing the following simplifying restrictions. These restrictions are imposed on parameters so that only stores closest to the consumer will be candidates for selection.

Parametric Restriction 4.1 *Communities served only by Brand b are assumed to be covered so consumer surplus is nonnegative at all locations within such a community. Algebraically, $\mu_b - t/2n_b \geq 0$.*

This assumption will simplify the statement of an optimal purchase policy for communities served only by Brand b. The following additional restrictions will simplify its statement for communities served by both brands.

Parametric Restriction 4.2 *A consumer standing in front of a store from Brand a is assumed to prefer Brand a even if it is of low utility-value. Algebraically, $\mu_a - \sigma_a > \mu_b - t/2n_b$. Similarly, a consumer standing in front of a store from Brand b is assumed to prefer Brand b even if Brand a is of high utility-value. Algebraically, $\mu_a + \sigma_a - t/2n_b < \mu_b$.*

Although price competition between brands is not being analyzed here, the above restrictions are of the "no mill-price undercutting" variety. Neither of these restrictions are formally required for the results, but the absence of such restrictions will complicate the analysis unnecessarily.

Given these assumptions, the current utility a consumer with beliefs x_a can expect to derive from travelling to a community served by Brand a is

$$\begin{aligned} v_{\Lambda_a}\left(x_a, \lambda^k\right) &= E_{\lambda \leq \lambda^k | \lambda \in \Lambda_a} u_a\left(x_a, \lambda\right) + E_{\lambda > \lambda^k | \lambda \in \Lambda_a} u_b\left(\lambda\right) \\ &= 2n_b \int_0^{\lambda^k} \left(\bar{x}_a - t\lambda\right) d\lambda + 2n_b \int_{\lambda^k}^{1/2n_b} \left[\mu_b - t\left(\frac{1}{2n_b} - \lambda\right)\right] d\lambda \\ &= 2n_b\lambda^k \left(\bar{x}_a - \frac{t}{2}\lambda^k\right) + \left(1 - 2n_b\lambda^k\right)\left[\mu_b - \frac{t}{2}\left(\frac{1}{2n_b} - \lambda^k\right)\right] \end{aligned} \tag{4.1}$$

where $E_{\lambda \leq \lambda^k | \lambda \in \Lambda_a}$ denotes the expectation over locations in a community served by Brand a and satisfying the restriction $\lambda \leq \lambda^k$. Note that $2n_b\lambda^k$ can be interpreted as either *(i)* the probability of a consumer purchasing from Brand a, given she travels to a community served by this brand, or *(ii)* the proportion of trips made to communities served by Brand a which result in purchases from this brand.

Since communities served only by Brand b are covered, when the consumer travels to such communities she will always purchase Brand b, choosing the store closest to her. Thus the current utility a consumer can expect to derive from travelling to such a community is

$$\begin{aligned} v_{\Lambda'_a} &= E_{\lambda \in \Lambda'_a} u_b\left(\lambda\right) \\ &= 2n_b \int_0^{1/2n_b} \left[\mu_b - t\left(\frac{1}{2n_b} - \lambda\right)\right] d\lambda \\ &= \mu_b - \frac{t}{4n_b} \end{aligned} \tag{4.2}$$

where $E_{\lambda \in \Lambda'_a}$ denotes the expectation over locations contained in a community not served by Brand a.

4.2 Narrowing the Problem

4.2.1 The Form of an Optimal Policy: Policy-Family $\mathcal{P}_0$

We seek an optimal policy governing the consumer's choice between brands when she travels to communities served by both brands. Given the previous assumptions, the analysis can be restricted to purchase policies which are members of the following family of policies:

Definition 4.1 ***Policy-Family*** *$\mathcal{P}_0$ is the set of all policies π^k satisfying the following condition:*

$$\pi^k(\lambda, x_a) = \begin{cases} a & \text{if} \quad (\lambda \leq \lambda^k(x_a)) \wedge (\lambda \in \Lambda_a) \\ b & \text{if} \quad \text{otherwise} \end{cases}$$

where $\lambda^k(x_a)$ is the maximum distance a consumer with beliefs x_a will travel to purchase from Brand a under Policy k. Implicit in this statement is the condition that, if the consumer visits a community not served by Brand a, she buys Brand b from the store closest to her location.

Such policies dictate the purchase Brand a when the consumer's primary travel pattern takes her to a community served by Brand a and the distance from Brand a is no greater than $\lambda^k(x_a)$. In what follows it will sometimes prove convenient to refer to Policy k by its critical value function $\lambda^k(x_a)$ rather than the entire function π^k. Similarly the following definition will also prove useful for future reference:

Definition 4.2 ***The Value of Policy* k**. *The current utility a consumer can expect to derive under Policy k when travelling to a community served by Brand a is $v^k_{\Lambda_a}(x_a) \equiv v_{\Lambda_a}(x_a, \lambda^k(x_a))$, found by evaluating Equation 4.1 at the critical value function $\lambda^k(x_a)$ which defines Policy k.*

4.2.2 Two Restrictions

In this section the problem of finding an optimal purchase policy will be narrowed by examining restrictions which such a policy must satisfy. First, an optimal policy must be optimal for all beliefs the consumer may hold (i.e. for both experienced and inexperienced consumers). Second, an optimal policy must perform at least as well as a myopic policy, which does not take into account future periods. Examining such restrictions will not only serve to break the problem of finding an optimal policy into parts, but also serve to improve our intuitive understanding of the policy once found.

Analysis is restricted in this section to the behavior of a consumer who visits a community served by both brands. The consumer must decide to which brand her secondary trip should be made. This is nothing more than the behavior traditionally assumed in the Hotelling framework, save for the modification that the consumer's most-preferred location (the starting point of the secondary trip) is stochastically determined (rather than fixed, as traditionally assumed). The stochasticity associated with her most-preferred location is derived from the stochasticity of her primary travel pattern. Once such behavior is better understood, Section 4.4 will extend the analysis to examine the behavior of a consumer whose primary trip-taking can take her to any community in the city, thus visiting communities served by Brand a only with a given frequency (strictly less than unity).

The Behavior of an Experienced Consumer: Policy-Family $\mathcal{P}_1$

This section examines restrictions imposed on an optimal policy by the condition that such a policy must be optimal for an experienced consumer—i.e. one who has previously purchased Brand a and knows its utility-value. Completion of this task will thus take us part way to finding an optimal policy. Finding an optimal policy will then amount to finding the additional restrictions imposed by the optimality condition for an inexperienced consumer. This latter task is reserved for Section 4.4.

To this end, consider the situation faced by an experienced consumer who travels to a community served by Brand a. Such a consumer will face one of two possible sets of valuation curves, each being comprised of the sole valuation-curve for Brand b and one of the two possible valuation-curves for Brand a. These curves are as depicted in Figure 4.2.

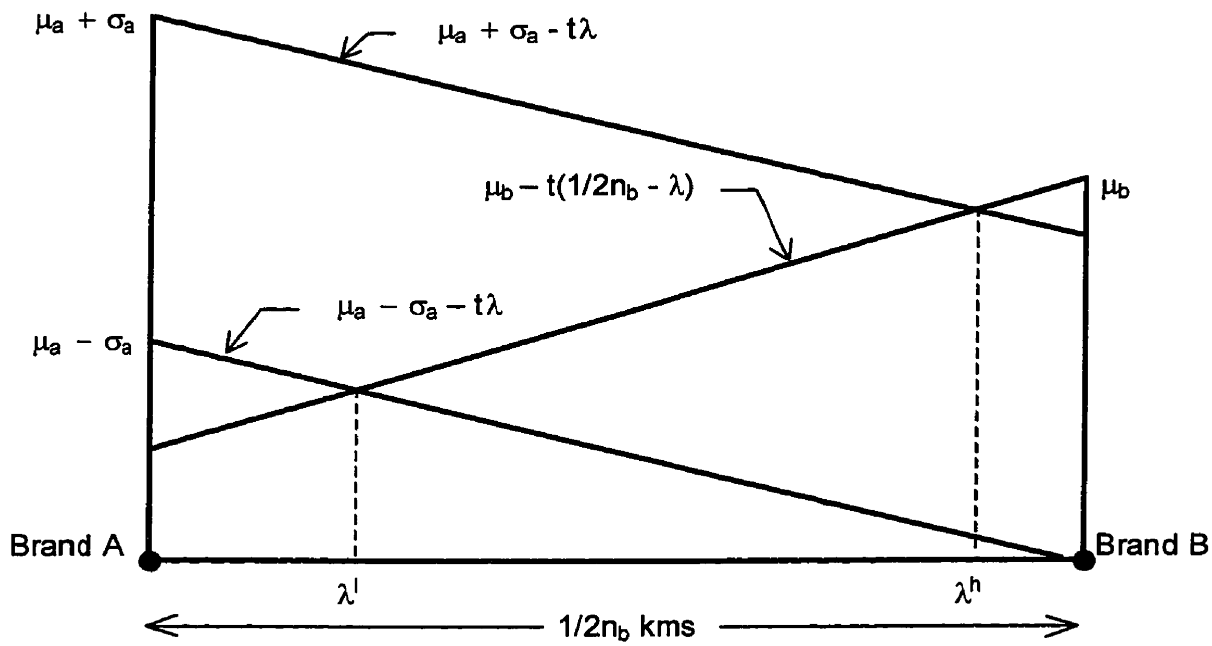


Figure 4.2: Optimal Choices for an Experienced Consumer in a Two-Brand Community

If $x''_a = 1_{\mu_a+\sigma_a}$ then an experienced consumer will buy Brand a for all locations $\lambda \leq \lambda^h$. However, if $x''_a = 1_{\mu_a-\sigma_a}$ then she will buy Brand a for all locations $\lambda \leq \lambda^l$. It follows, then, that we can further restrict our search to policies which take the following form:

Definition 4.3 ***Policy-Family*** $\mathcal{P}_1$ *is the set of all policies* $\pi^k \in \mathcal{P}_0$ *satisfying the following restriction:*

$$\lambda^k(x_a) = \begin{cases} \lambda^h & \text{if} \quad x_a = 1_{\mu_a + \sigma_a} \\ \lambda^k & \text{if} \quad x_a = 1/2_{\mu_a \pm \sigma_a} \\ \lambda^l & \text{if} \quad x_a = 1_{\mu_a - \sigma_a} \end{cases}$$

where the value λ^k *defines the* k^{th} *member of this family and where the values* λ^h *and* λ^l *are as diagrammatically defined above (and algebraically defined in Table 4.1 below). Obviously,* $\mathcal{P}_1 \subset \mathcal{P}_0$*.*

Note that members of this family of policies differ only in the critical location λ^k governing the choice of an inexperienced consumer. Further, all dictate the same optimal sub-policies (λ^h, λ^l) governing choices made by an experienced consumer. It should be clear, then, that in seeking a generally optimal policy $\lambda^*(x_a)$, we seek a policy which is a member of $\mathcal{P}_1$ and which is distinguished by its assignment of an optimal critical location λ^* governing the choice made by an inexperienced consumer.

At this point it will be useful to introduce some additional notation for the values received by a consumer under any policy $\pi^k \in \mathcal{P}_1$. To this end, recall from Definition 4.3 that $v^k_{\Lambda_a}(x_a) \equiv v_{\Lambda_a}(x_a, \lambda^k(x_a))$ is the current utility a consumer can expect to derive under any policy π^k when travelling to a community served by Brand a. Given Definition 4.3, it follows that the current utility a consumer will receive under policy $\pi^k \in \mathcal{P}_1$ is

$$v^k_{\Lambda_a}(x_a) = \begin{cases} v^l_{\Lambda_a} & \text{if} \quad x_a = 1_{\mu_a + \sigma_a} \\ v^k_{\Lambda_a} & \text{if} \quad x_a = 1/2_{\mu_a \pm \sigma_a} \\ v^h_{\Lambda_a} & \text{if} \quad x_a = 1_{\mu_a - \sigma_a} \end{cases} \tag{4.3}$$

where $v^l_{\Lambda_a} \equiv v_{\Lambda_a}\left(1_{\mu_a-\sigma_a}, \lambda^l\right)$, $v^k_{\Lambda_a} \equiv v_{\Lambda_a}\left(1/2_{\mu_a\pm\sigma_a}, \lambda^k\right)$, $v^h_{\Lambda_a} \equiv v_{\Lambda_a}\left(1_{\mu_a+\sigma_a}, \lambda^h\right)$.[2] In particular, note that while value $v^k_{\Lambda_a}$ is particular to policy k, the values $v^l_{\Lambda_a}$ and $v^h_{\Lambda_a}$ are common to all policies which are members of Family $\mathcal{P}_1$. These latter values are derived in the next section and presented in Table 4.1.

Policy-Family $\mathcal{P}_2$ and a Myopic Purchase Policy

This section examines the restrictions imposed by the myopic purchase policy. An optimal policy, which considers the value of information contained in a purchase from Brand a, must perform at least as well as a myopic purchase policy, which does not consider such value. Examining a myopic policy serves two purposes. First, the bound imposed by a myopic policy serves to not only narrow the problem of finding an optimal policy, but also aids our intuitive understanding of such a policy. Second, and more importantly, once an optimal policy is found, the role of chain size in creating information value for the consumer can be usefully analyzed by comparing and contrasting the optimal policy to the purchase behavior dictated by a myopic policy.

A **myopic purchase policy** is defined by the condition that the brand chosen is that offering the highest expected *current* utility, given the consumer's location and beliefs. Graphically, a myopic purchase policy looks as depicted in Figure 4.3.

[2] Since all policies under consideration differ only by the value governing the choice of an inexperienced consumer, we are denoting such policies according to the values so assigned. Nonetheless, whenever there is the possibility of confusion, the policy function $\lambda^k(x_a)$ and the policy value λ^k for $x_a = 1/2_{\mu_a\pm\sigma_a}$ will be carefully distinguished by the presence of the argument x_a on the former but not on the latter. Similar comments apply to the function $v^k_{\Lambda_a}(x_a)$ and its value $v^k_{\Lambda_a}$ for $x_a = 1/2_{\mu_a\pm\sigma_a}$. Note, as well, that such policies may also use other parameters of the decision problem (in addition to the consumer's beliefs), but these will be suppressed for now.

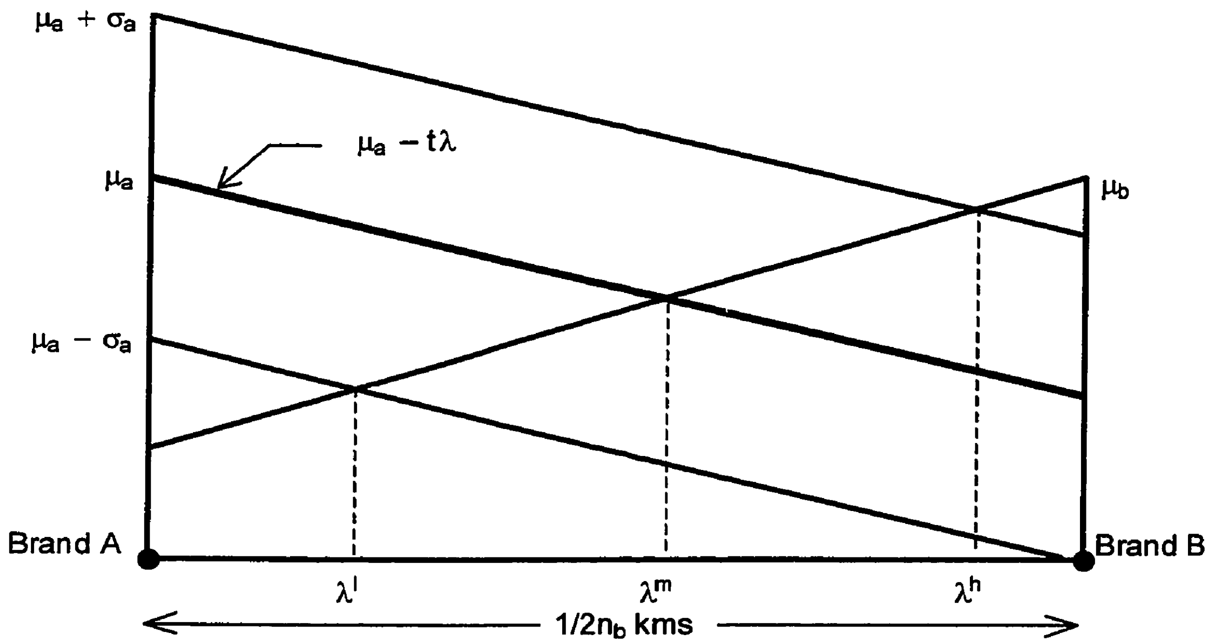


Figure 4.3: λ^m, A Myopic Purchase Policy

Formally, a myopic policy $\lambda^m(x_a)$ is defined by the condition:[3]

$$\begin{aligned} u_a(x_a, \lambda^m(x_a)) &\equiv u_b(\lambda^m(x_a)) \\ \bar{x}_a - t\lambda^m(x_a) &\equiv \mu_b - t\left(\frac{1}{2n_b} - \lambda^m(x_a)\right) \\ \lambda^m(x_a) &\equiv \frac{1}{2n_b}\left(\frac{1}{2} + \frac{\bar{x}_a - \mu_b}{\tau}\right) \end{aligned} \tag{4.4}$$

where $\tau \equiv t/n_b$ is the "normalized" **cost of secondary travel** for the consumer—a

[3]Note that this condition is equivalent to:

$$\frac{\partial}{\partial\lambda} v_{\Lambda_a}(x_a, \lambda^m(x_a)) \equiv 0$$

measure of the inconvenience of stopping off while on her primary trip.

Since $x_a \in \left\{1/2_{\mu_a \pm \sigma_a}, 1_{\mu_a+\sigma_a}, 1_{\mu_a-\sigma_a}\right\}$, the myopic policy can be stated more explicitly as follows:

$$\lambda^m(x_a) = \begin{cases} \lambda^h & \text{if} \quad x_a = 1_{\mu_a+\sigma_a} \\ \lambda^m & \text{if} \quad x_a = 1/2_{\mu_a \pm \sigma_a} \\ \lambda^l & \text{if} \quad x_a = 1_{\mu_a-\sigma_a} \end{cases} \tag{4.5}$$

where the definitions of Table 4.1 are employed (recall also that $\Delta \equiv \mu_b - \mu_a$).

Table 4.1: Some Definitions

$\lambda^l \equiv \alpha^l/2n_b$	$\lambda^m \equiv \alpha^m/2n_b$	$\lambda^h \equiv \alpha^h/2n_b$
$\alpha^l \equiv \alpha^m - \sigma_a/\tau$	$\alpha^m \equiv 1/2 - \Delta/\tau$	$\alpha^h \equiv \alpha^m + \sigma_a/\tau$

Parametric Interpretation 4.1 *Note that Parametric Restriction 4.2 implies* $0 \leq \alpha^l \leq \alpha^m \leq \alpha^h \leq 1$ *and therefore* $0 \leq \lambda^l \leq \lambda^m \leq \lambda^h \leq 1/2n_b$. *These λ's are the maximum secondary-travel distances for consumers under alternative assumptions about the experience of consumers and the underlying value of Brand a. Given the primary travel pattern of consumers is uniform, these maximum distances imply the α's may be interpreted as probabilities of purchase from Brand a. For example, an experienced consumer who has found Brand a to be of high-value will choose Brand a if her primary travel pattern takes her within a distance of λ^h of a store from Brand a's chain. Thus, given her primary travel pattern is uniform and takes her to a community served by Brand a, the probability that such a consumer buys from Brand a is α^h, where* $\alpha^h \equiv \lambda^h/(1/2n_b) = 2n_b\lambda^h$.

A myopic policy is a member of Policy-Family $\mathcal{P}_1$ and can be used to further the search for an optimal policy. What distinguishes a myopic policy within the family of policies $\mathcal{P}_1$ is that it would have an inexperienced consumer buy from Brand a for all locations $\lambda \leq \lambda^m$ and from Brand b for all locations $\lambda > \lambda^m$. It follows that, for $\sigma_a \geq 0$, an optimal policy λ^* will be a member of a family of policies $\mathcal{P}_2 \subset \mathcal{P}_1$ satisfying the restriction $\lambda^m \leq \lambda^* \leq \lambda^h$. This is true since an optimal policy takes into account the present value of information about Brand a acquired through its purchase. Since a myopic policy does not take into account the value of such information, the distance an inexperienced consumer will travel to purchase from Brand a under an optimal policy will be greater than that under a myopic policy. Thus we have:

Definition 4.4 ***Policy-Family*** *$\mathcal{P}_2$ is the set of all policies $\pi^k \in \mathcal{P}_1$ satisfying the condition that $\lambda^m \leq \lambda^k \leq \lambda^h$, where λ^k is the critical distance governing the behavior of an inexperienced consumer under policy π^k.*

Proposition 4.1 *An optimal policy π^* is a member of Policy-family $\mathcal{P}_2$. In particular, $\lambda^m \leq \lambda^* \leq \lambda^h$ for $\sigma_a^2 \geq 0$.*

The values for λ^l and λ^h in Table 4.1 imply values for $v^l_{\Lambda_a}$ and $v^h_{\Lambda_a}$ in Equation 4.3. These are the values received by an experienced consumer under any policy $\pi^k \in \mathcal{P}_1$. These values, along with the value $v^m_{\Lambda_a}$ received by an inexperienced consumer under a myopic policy are derived in the Appendix and presented below:

$$\begin{aligned} v^l_{\Lambda_a} &\equiv v_{\Lambda_a}\left(1_{\mu_a-\sigma_a}, \lambda^l\right) \\ &= \mu_b - \frac{\tau}{4} + \frac{\tau}{2}\left(\alpha^l\right)^2 \end{aligned} \tag{4.6a}$$

$$\begin{aligned} v^m_{\Lambda_a} &\equiv v_{\Lambda_a}\left(1/2_{\mu_a\pm\sigma_a}, \lambda^m\right) \\ &= \mu_b - \frac{\tau}{4} + \frac{\tau}{2}\left(\alpha^m\right)^2 \end{aligned} \tag{4.6b}$$

$$\begin{aligned} v^h_{\Lambda_a} &\equiv v_{\Lambda_a}\left(1_{\mu_a+\sigma_a}, \lambda^h\right) \\ &= \mu_b - \frac{\tau}{4} + \frac{\tau}{2}\left(\alpha^h\right)^2 \end{aligned} \tag{4.6c}$$

Note that the value $v^m_{\Lambda_a}$ is unique to a myopic policy, whereas the values $v^l_{\Lambda_a}$ and $v^h_{\Lambda_a}$ are common to all members of the policy family $\mathcal{P}_1$.

4.3 The Value of Information: A Restricted Analysis

This section examines the value of information (contained in a purchase from Brand a) a consumer can expect to derive within a single period given her primary travel pattern takes her to a community served by both brands. Thus, both the temporal and spatial nature of the consumers problem is here restricted. The analysis is performed both graphically and algebraically.

4.3.1 Finding the Value of Information Graphically

To better understand Proposition 4.1, and the value of information contained in a purchase from Brand a, consider Figure 4.4. This diagram makes it clear that a community served by both brands is comprised of three geographic areas.[4] At each end of the community is a segment where one brand is dominant. Information acquired about Brand a will be of no current value should the consumer's primary travel pattern take her to a location in either of these two areas. In contrast, such information does have value if her primary travel pattern takes her to a location in the middle region of the community. This is the decision zone, where she faces a nontrivial choice problem.

[4] Actually, this is only one half of the community. However, the other half (on the other side of Brand a) is analytically identical to this one, so we will speak as if this is the entire community.

The shaded triangles represent the potential gains from knowing the true value of Brand a and making the correct purchase decision on the basis of that information. The areas of these triangles are directly related to the value of information under a myopic purchase policy, though such a policy does not take into account this value.

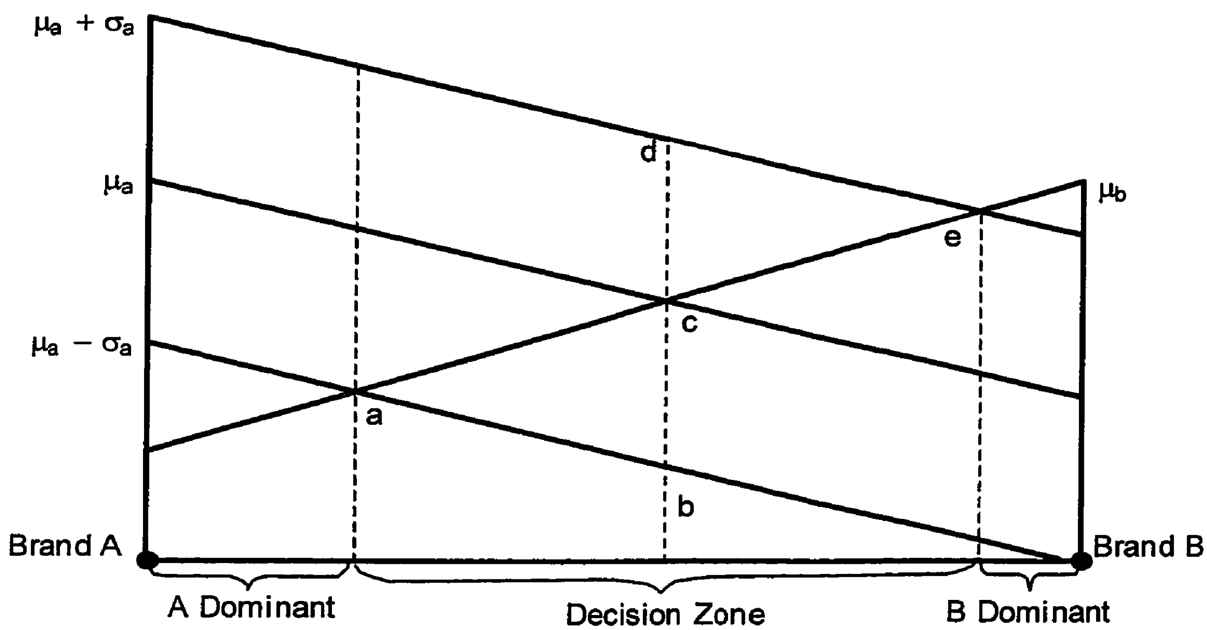


Figure 4.4: The Value of Information under a Myopic Policy

The height of the triangles represents the cost of making incorrect future purchase decisions about Brand a, whereas the length of the decision zone determines the likelihood that such decisions will have to be made. The two triangles are symmetric and have equal probability of "occurring", so we need only find the expected value of one to find the value of information. To this end, consider the left-hand triangle, abc. The difference in the heights of the curves bounding this triangle is $2t\lambda$ and the length of the triangle is $\lambda^m - \lambda^l = \sigma_a/2t$. Given the consumer's primary travel-pattern is uniform, all locations in this space of length $1/2n_b$ are equally likely. Thus, the expected value of this triangle, F, is

$$\begin{aligned} F &= 2n_b \int_0^{\sigma_a/2t} 2t\lambda d\lambda \\ &= \frac{\sigma_a^2}{2\tau} \end{aligned} \tag{4.7}$$

This result is discussed below in Section 4.3.3.

4.3.2 Finding the Value of Information Algebraically

The expected value of information under a myopic policy can also be found somewhat more formally from the following identity:

$$F = E_{x_a''|a,1/2_{\mu_a \pm \sigma_a}} v_{\Lambda_a}^m\left(x_a''\right) - v_{\Lambda_a}^m\left(E_{x_a''|a,1/2_{\mu_a \pm \sigma_a}} x_a''\right) \tag{4.8}$$

where $E_{x_a''|a,1/2_{\mu_a \pm \sigma_a}}$ is as defined previously in Equation 2.4. The first term on the right-hand side of Equation 4.8 is the expected value of being an experienced consumer. Since a myopic policy $\pi^m \in \mathcal{P}_1$ it follows, from Equations 4.3, 4.6a and 4.6c, that this value is

$$\begin{aligned} E_{x_a''|a,1/2_{\mu_a \pm \sigma_a}} v_{\Lambda_a}^m\left(x_a''\right) &= \frac{1}{2} v_{\Lambda_a}^l + \frac{1}{2} v_{\Lambda_a}^h \\ &= \mu_b - \frac{\tau}{4} + \frac{\tau}{2}\beta \end{aligned} \tag{4.9}$$

where $\beta \equiv (\alpha^m)^2 + (\sigma_a/\tau)^2$.[5] Since $E_{x''_a|a,1/2_{\mu_a \pm \sigma_a}} x''_a = 1/2_{\mu_a \pm \sigma_a}$, the second term on the right-hand side of 4.8 is value expected to be received by an inexperienced consumer under a myopic policy. From Equation 4.6b, this value is $v^m_{\Lambda_a} = \mu_b - \frac{\tau}{4} + \frac{\tau}{2}(\alpha^m)^2$. Thus, it follows, by substitution, that

$$\begin{aligned} F &= \left(\mu_b - \frac{\tau}{4} + \frac{\tau}{2}\beta\right) - \left(\mu_b - \frac{\tau}{4} + \frac{\tau}{2}(\alpha^m)^2\right) \\ &= \frac{\sigma_a^2}{2\tau} \end{aligned} \tag{4.10}$$

as before.

4.3.3 Analysis

F is the expected one-period value of having perfect information about Brand a when travelling to a community served by Brand a.[6] This value is increasing in the uncertainty of the consumer's beliefs about Brand a (as measured by the variance of her prior distribution, σ_a^2) and decreasing in the cost of secondary travel τ—the cost of stopping while on her primary trip.[7]

---

[5]Note that the restriction $\alpha^h \leq 1$ implies that $(\alpha^h)^2 = \alpha^2 + \sigma_a^2/\tau^2 + 2\alpha\sigma_a/\tau \leq 1$ and therefore $\alpha^2 + \sigma_a^2/\tau^2 = \beta \leq 1$. Note also that

$$\beta = \frac{(\alpha^l)^2 + (\alpha^h)^2}{2}$$

[6]This value may also be found by purely geometric methods. Again, consider the left-hand triangle, *abc*. Use point a as the top of the triangle and cb as its base. Drop a perpendicular down to its base so that two right-angle triangles are formed. The height of each these triangles is $\lambda^m - \lambda^l = \sigma_a/2t$ and their base is $\sigma_a/2$. Thus the combined area of both right-angle triangles is

$$Area\,(abc) = \frac{\sigma_a^2}{4t}$$

Since the probability (loosely speaking) the consumer's primary travel pattern takes her to any location within the length of this triangle is $2n_b$ the result follows.

[7]Formally, $\partial F/\partial\sigma_a = \sigma_a/\tau > 0$ and $\partial F/\partial\tau = -\sigma_a^2/2\tau^2$.

The effect of each of these two factors on the information value of Brand a is apparent from Figure 4.4. In particular, an increase in the consumer's uncertainty about Brand a has the effect of increasing both the height of the shaded triangles and the length of the decision zone. In contrast, as the inconvenience cost of the secondary trip falls, the valuation curves for each brand get flatter and the decision zone expands, again increasing the value of information about Brand a. The role of uncertainty has been previously discussed in Section 2.4, so it is the role of convenience which is of particular interest here. This role is summarized in the following proposition:

Proposition 4.2 *A reduction in the cost of secondary travel, τ, not only reduces the cost of the current purchase, but also increases the value of information contained in that purchase. Both of these effects increase the likelihood of the consumer purchasing the brand. The increase in the value of information is depicted graphically in Figure 4.4 by an increase in the length of the decision zone which accompanies a reduction in τ.*

The present model has expanded the concept of convenience from the formalization which was used in the model of Chapter 3. Convenience is provided to a brand's customers through more than just chain size. Many other factors associated with the attributes of the individual stores in the chain, rather than the chain itself are also important in the provision of convenience. Retail chains are often concerned with quick service; with the ease of access to and egress from premises; and with the general convenience of purchase. Proposition 4.2 states that these convenience services not only reduce the consumer's current costs of obtaining the brand, but also increase the value of information associated with its purchase, making them all the more important.[8]

[8] The central purpose of this paper is to present an argument for the existence of chains, not to analyze competition between them. In keeping with this purpose, the present model has restricted

Remark 4.1 *Chain size, c_a, determines the brand's coverage of the consumer's primary travel pattern. As such, it will sometimes be referred to as the brand's level of "primary convenience." In contrast, τ is the cost of secondary travel and, as such, will sometimes be referred to as determining the brand's level of "secondary convenience."*

4.3.4 Implications for an Optimal Purchase Policy

As noted earlier, an optimal policy will consider not only the expected current utility to be derived from the purchase of Brand a, but also the expected present value of information. Let F^* denote the expected present value of information obtained under an optimal policy.[9] An intuitive understanding of the role of information value in

---

the convenience costs (i.e. τ's) for each brand to be the same. This restriction has the effect of separating the effect of convenience on information-value from its effect on the current cost of purchase. Diagrammatically, the midpoint of the market is maintained while the length of the decision zone and the heights of the yellow triangles are allowed to expand and contract with changes in τ. Obviously if convenience costs were to be an object of competitive analyses, the model would need to be generalized to allow different brands to have different τ's. Similar comments apply to pricing and other variables typically analyzed in competitive choice.

At times the interpretation of the metric for τ has been rather loose. This is intentional, but it stretches the formal interpretation of τ and space of secondary travel. Formally, the space of secondary travel in the present model is a uniform geographic space. This is the simplest possible interpretation for there is the requirement that the "space of secondary travel" be physically linked to the "space of primary travel," otherwise the consumer could not move between these spaces. For example, highway travel must be physically linked to secondary-road travel. Obviously, the cost per unit distance of secondary travel can be affected by such thing as size of parking lot; the ease of access and egress from premisses; whether the store has a drive-thru window; whether the store is on the opposite side of the primary-travel road the consumer is on; whether there is a meridian dividing such the road, etc. Two brands may be directly across the intersection from each other, yet one may be much more accessible than the other, despite the fact that they are very close (on a distance metric). Thus, at least informally, the space of secondary travel being considered is far from uniform, and might be best thought of in terms of a "convenience metric", rather than the traditional distance metric. Note also that if τ is sufficiently large, the space of secondary travel might not be "covered", allowing for the possibility that the consumer stays on her primary route, not stopping to buy any brand.

An excellent guide to locating retail stores is provided by Salvaneschi (1996). In particular, he examines factors as the size and positioning of parking lots, the position of entrances and exits, where to locate stores at intersections, if there is a bend in the road, etc. All these factors can affect the inconvenience of a consumer stopping while on her primary trip.

[9]Note that F^* will differ somewhat from F (derived above) for reasons discussed below in Section 4.4—if it did not, the job of finding an optimal policy would be completed.

affecting the purchase of Brand a can be gained through Figure 4.5. As depicted, the value of information F^* increases the height of an inexperienced consumer's valuation-curve for Brand a making the purchase of this brand more likely under an optimal policy.

A myopic policy is defined by the condition that it maximize current-period utility-value. Since information has value only in future periods, a myopic policy does not expend resources on capturing such value. In particular, the current-period cost of acquiring information under any policy $\pi^k \in \mathcal{P}_2$ is $2t\left(\lambda^k - \lambda^m\right)$, which is the incremental travel cost beyond that borne by a myopic policy. This cost is depicted in Figure 4.5 for the optimal policy by the length of the dashed line-segment $gh = 2t\left(\lambda^* - \lambda^m\right)$. Since the cost of acquiring information is borne in the current period, a myopic policy minimizes it by setting $\lambda^k = \lambda^m$. Thus, even though information value exists under a myopic policy (since Brand a could be purchased under such a policy), such a policy does not take this value into account in setting λ^k for an inexperienced consumer. However, an optimal policy sets $\lambda^k = \lambda^*$ so that $gh = F^*$, where the marginal (current) cost of acquiring information is just equal to the present-value of the information's marginal future benefit.

The following proposition is stated here for heuristic purposes, next to its associated graph, although more formal algebraic justification for the proposition is reserved for Section 4.5.

Proposition 4.3 *Any factor which increases the convenience of Brand a also increases the value of information and, therefore, the critical distance λ^* which governs the purchases of an inexperienced consumer under an optimal policy. In the context of the present model there are two classes of such factors. First, any factor which reduces τ, the cost of secondary travel (i.e. the inconvenience of stopping while on a primary trip), such as quick service and the ease of access and egress from premises,*

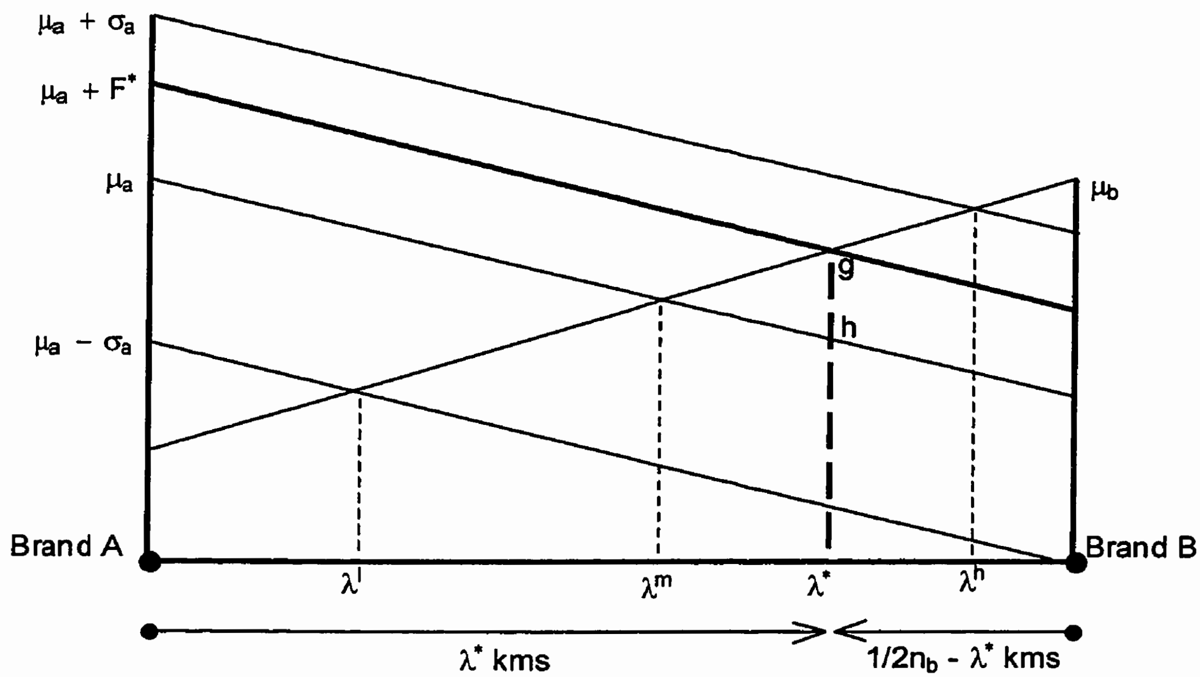


Figure 4.5: λ^*, An Optimal Policy for an Inexperienced Consumer

will function in this manner. Second, chain size also functions in this manner. The greater the size of Brand a's retail chain, the greater the coverage of the consumer's primary travel pattern, and the greater the value of any information acquired about Brand a.

The following point is of a more technical nature and of lesser importance to the understanding of an optimal policy. As such, it is made in the form of the following remark:

Remark 4.2 *An optimal policy not only affects the cost of acquiring information, but also its value. To see this note that the value of information depends on how it affects the consumer's decisions. As such, the value of information is not some* *universal constant, but rather depends on (i) how the information is used once acquired and (ii) the decisions which would be made prior to its acquisition. With respect to (i), all purchase policies (examined herein) use information in the same way: i.e. optimally. However, with respect to (ii), not all purchase policies make the same decisions prior to the acquisition of information. In particular, the myopic policy sets* $\lambda^k = \lambda^m$*, whereas the optimal policy sets* $\lambda^k = \lambda^*$*. This difference in the prior decisions made by these policies affects the value of information. Consumers who use different purchase policies value information differently. Graphically, the value of information under an optimal policy is related to the area of the lightly shaded triangles in Figure 4.6 (to avoid clutter, the valuation line* $\mu_a + F^*$ *which goes through point "g" has been removed). Note that the combined area of these triangles is greater*

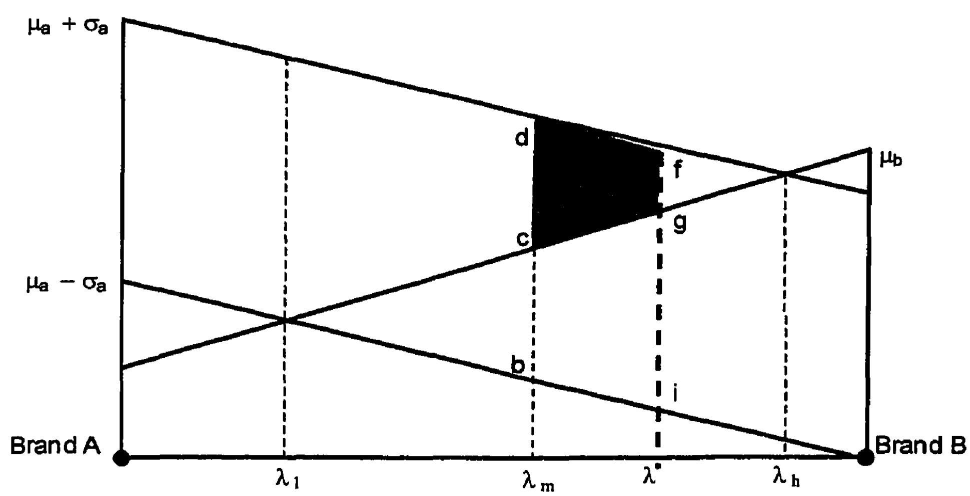


Figure 4.6: The Value of Information Under an Optimal Policy

than that under a myopic policy, so the value of information under an optimal policy is also greater. Intuitively, an optimal policy makes fewer mistakes if Brand a is of high-value (represented by the darkly-shaded area "cdfg"), but more mistakes when

Brand a is of low value (represented by the expanded lightly-shaded area "bcgi"). The mistakes made by an optimal policy prior to the acquisition of information are larger than those made by a myopic policy—this is simply the price paid by an optimal policy to acquire information which will improve future decisions.

4.4 Finding the Optimal Purchase Policy

We seek an optimal purchase policy which governs the choice of an inexperienced consumer in a community served by Brand a. Thus, suppose the consumer's primary travel pattern takes her to a location where both brands are available. Given her uniform travel pattern, she will, almost surely, find herself at such a location eventually.

4.4.1 A Problem Created by Expanding the Concept of Convenience

In solving for the optimal purchase policy, we have, up to now, been able to use the argument that if it is optimal to buy Brand b in the current period, then it is optimal to buy it in all future periods as well. If this argument could be used in the present model, the problem of finding an optimal policy could be solved by simply extending the analysis of the previous section. This argument was available for use in the models of the previous chapters since the decision-problem faced by the consumer in those models was time-invariant. Unfortunately, despite the fact that a geometric discount sequence is assumed,[10] the consumer's decision-problem in the present model no longer possess this time-invariance property.

The consumer's decision-problem is no longer time-invariant due to the continuous nature of the geographic space and the existence of secondary travel within that space.

[10] Recall the discussion in Remark 10.

In particular, the consumer's primary travel within this space results in her distance from Brand a and, therefore, her cost of purchasing Brand a, changing from period to period. In the model of the last chapter, however, the cost of acquiring Brand a was a constant equal to the foregone value of the known brand, Brand b. Given the size of Brand a's retail chain, the consumer either purchased Brand a in the current period or purchased Brand b then and forever after. In the present model, however, there is a random variable at work: For reasons exogenous to the demand for this product, the consumer may find herself in neighborhood of Brand a and, therefore, may find the purchase of Brand a, and the information acquired through its purchase, is relatively inexpensive to acquire. In particular, it is no longer the case in the present model that the cost of acquiring information about Brand a is a constant. Thus, although she may find it convenient to purchase Brand b in the current period, this does not mean she will continue to purchase Brand b into the indefinite future; she may or may not, depending on how convenient each brand is in each period. It follows that the introduction of secondary travel within a continuous geographic space, and the inconvenience cost associated with such travel, has somewhat complicated the determination of an optimal purchase policy. Nonetheless, an optimal policy can be found.

4.4.2 Policy B: A Policy of Postponement

In solving for an optimal policy, we can, as before, restrict our analysis to policies which choose either Brand a or Brand b in the current period, and choose optimally thereafter. Note that the only difference between these two policies, other than current utility value, is due to the timing of the purchase of Brand a and the concomitant receipt of information about its utility-value. Under Policy A, the purchase of Brand a

takes place in the current period and any information so obtained can then be immediately used in all future purchases. Policy B, in contrast, is a policy of postponement. Under Policy B, the purchase of Brand a is postponed for at least one more period, to a time stochastically determined by the consumer's travel pattern and the optimal purchase policy.

Ultimately, however, the purchase of Brand a will almost surely take place under Policy B as well. This is true since future purchases under Policy B are governed by the optimal policy, and $\lambda^* > \lambda^m > 0$ for $\sigma_a > 0$. In particular, let p^* denote the probability the consumer buys Brand a for the first time in any period under the optimal purchase policy. The relationship between p^* and λ^* is easily found. Given the consumer's uniform travel pattern, she will travel to a community served by Brand a with probability $c_a \equiv n_a/n_b$. Given she travels to such a community, the probability she finds herself at a location for which the purchase Brand a is dictated by the optimal policy is $2n_b\lambda^*$. The product of these two probabilities yields $p^* = 2n_a\lambda^*$, the probability she, as an inexperienced consumer, buys Brand a for the first time in any period under the optimal policy.[11]

Policy B is thus best viewed as one which postpones the purchase of Brand a to a future period. Obviously a policy of postponement might be optimal given her current location. This would surely be the case if the consumer's primary travel pattern takes her to a location in that range where Brand b is dominant (i.e. $\lambda > \lambda^h$, see Figure 4.4). However, once Brand a has been purchased under Policy B, the two policies, A and B, will yield identical values in all future periods, each having identical information sets and using those sets in an identical manner. Thus Policy B will ultimately yield information as well, it's just that the acquisition of the information under this policy is postponed to a future period, to be stochastically determined by the consumer's

[11]To verify that $2n_a\lambda^*$ does satisfy minimum requirements of a probability note that $0 \leq \lambda^* \leq 1/2n_b$. Multiplying through by $2n_a$ yields $0 \leq 2n_a\lambda^* \leq n_a/n_b \leq 1$.

primary travel pattern.

4.4.3 The Optimal Policy with Secondary Travel

In this section the job of finding the optimal purchase policy is carried out. As before, in finding an optimal policy we can restrict our attention to policies which choose either Brand a or b in the current period and optimally thereafter. Let $V_a(x_a, \lambda)$ and $V_b(x_a, \lambda)$ be, respectively, the expected present-values of these policies, given the consumer's beliefs x_a and her location λ. These values can be found as follows:

- **Policy** A: Choose Brand a in the current period, choosing optimally thereafter. Thus

$$V_a^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) = u_a\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) + F_a^*\left(1/2_{\mu_a \pm \sigma_a}\right) \tag{4.11}$$

where u_a is the current utility derived from a purchase of Brand a and

$$\begin{aligned} F_a^*\left(1/2_{\mu_a \pm \sigma_a}\right) &\equiv E_\lambda E_{x_a''|a,1/2_{\mu_a \pm \sigma_a}} V^*\left(x_a'', \lambda\right) \\ &= \frac{1}{i}\left(c_a E_{x_a''|a,1/2_{\mu_a \pm \sigma_a}} v_{\Lambda_a}^*\left(x_a''\right) + \left(1 - c_a\right) v_{\Lambda_a'}^*\right) \end{aligned} \tag{4.12}$$

is the expected present value of future purchases under an optimal policy which starts with the selection of Brand a. If the consumer travels to a community served by Brand a, she can, under an optimal purchase policy, expect to receive utility of

$$\begin{aligned} E_{x_a''|a,1/2_{\mu_a \pm \sigma_a}} v_{\Lambda_a}^*\left(x_a''\right) &= \frac{1}{2} v_{\Lambda_a}^l + \frac{1}{2} v_{\Lambda_a}^h \\ &= \mu_b - \frac{\tau}{4} + \frac{\tau}{2}\beta \end{aligned} \tag{4.13}$$

where $v^l_{\Lambda_a}$ and $v^h_{\Lambda_a}$ are as defined previously in Equations 4.6a and 4.6c, respectively. If, however, the consumer travels to a community served only by Brand b, she can, under an optimal purchase policy, expect to receive utility of

$$v^*_{\Lambda'_a} = \mu_b - \frac{\tau}{4} \tag{4.14}$$

(as previously derived in Equation 4.2). Substitution of Equations 4.13 and 4.14 into Equation 4.12 yields

$$\begin{aligned} F^*_a\left(1/2_{\mu_a \pm \sigma_a}\right) &= \frac{1}{i}\left[c_a\left(\mu_b - \frac{\tau}{4} + \frac{\tau}{2}\beta\right) + (1-c_a)\left(\mu_b - \frac{\tau}{4}\right)\right] \\ &= \frac{1}{i}\left(\mu_b - \frac{\tau}{4} + \frac{\tau}{2}\beta c_a\right) \end{aligned} \tag{4.15}$$

- **Policy** B: Choose Brand b in the current period, choosing optimally thereafter. Thus,

$$V^*_b\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) = u_b\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) + F^*_b\left(1/2_{\mu_a \pm \sigma_a}\right) \tag{4.16}$$

where $F^*_b\left(1/2_{\mu_a \pm \sigma_a}\right) \equiv E_\lambda E_{x''_a | b, 1/2_{\mu_a \pm \sigma_a}} V^*\left(x''_a, \lambda\right)$ is the expected present value of future purchases under an optimal policy which starts with the selection of Brand b. Under this policy, the beliefs of the consumer do not change, so $E_{x''_a | b, 1/2_{\mu_a \pm \sigma_a}} V^*\left(x''_a, \lambda\right) = V^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right)$. Thus,

$$\begin{aligned} F^*_b\left(1/2_{\mu_a \pm \sigma_a}\right) &= E_\lambda V^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) \\ &= c_a E_{\lambda \in \Lambda_a} V^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) + (1-c_a) E_{\lambda \in \Lambda'_a} V^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda\right) \end{aligned} \tag{4.17}$$

If the consumer travels to a community served by Brand a, she can, under an optimal purchase policy, expect to receive utility of

$$\begin{aligned} E_{\lambda\in\Lambda_a} V^*\left(1/2_{\mu_a\pm\sigma_a},\lambda\right) &= 2n_b \int_0^{\lambda^*} V_a^*\left(1/2_{\mu_a\pm\sigma_a},\lambda\right) d\lambda + 2n_b \int_{\lambda^*}^{1/2n_b} V_b^*\left(1/2_{\mu_a\pm\sigma_a},\lambda\right) d\lambda \\ &= 2n_b\lambda^*\left[\mu_a - \frac{t}{2}\lambda^* + F_a^*\left(1/2_{\mu_a\pm\sigma_a}\right)\right] \\ &\quad + (1-2n_b\lambda^*)\left[\mu_b - \frac{t}{2}\left(\frac{1}{2n_b}-\lambda^*\right) + F_b^*\left(1/2_{\mu_a\pm\sigma_a}\right)\right] \end{aligned} \tag{4.18}$$

Whereas if she travels to a community served only by Brand b, she can, under an optimal purchase policy, expect to receive utility of

$$\begin{aligned} E_{\lambda\in\Lambda_a'} V^*\left(1/2_{\mu_a\pm\sigma_a},\lambda\right) &= 2n_b \int_0^{1/2n_b} \left(\mu_b - t\lambda + F_b^*\left(1/2_{\mu_a\pm\sigma_a}\right)\right) d\lambda \\ &= \mu_b - \frac{\tau}{4} + F_b^*\left(1/2_{\mu_a\pm\sigma_a}\right) \end{aligned} \tag{4.19}$$

Substitution of Equations 4.18 and 4.19 into Equation 4.17 yields[12]

$$F_b^*\left(1/2_{\mu_a\pm\sigma_a}\right) = \frac{1}{(i+2n_a\lambda^*)}\left\{\mu_b - \frac{\tau}{4} + 2n_a\lambda^*\left(F_a^*\left(1/2_{\mu_a\pm\sigma_a}\right) + \frac{\tau}{2} - t\lambda^* - \Delta\right)\right\} \tag{4.20}$$

Remark 4.3 *With respect to these policies, note that F_a^* is not dependent on the unknown portion of the optimal policy λ^*. That is, F_a^* depends solely on that portion of an optimal policy which applies to an experienced consumer and not that applying to an inexperienced consumer. Thus, F_a^* is determined by λ^h and λ^l and not λ^*, so $F_a^* = F_a^k$, for all policies $k \in \mathcal{P}_2$. This is not the case with F_b^*. A policy of postponement leaves the consumer inexperienced—at least for one more period. Thus, future decisions under Policy B will rely on λ^*.*

[12] The details are contained in the appendix.

The model can now be solved in the traditional manner by finding the location $\lambda = \lambda^*$ which just leaves her indifferent between brands. Thus the optimal policy λ^* satisfies the following condition[13]

$$\begin{aligned} V_a^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda^*\right) &= V_b^*\left(1/2_{\mu_a \pm \sigma_a}, \lambda^*\right) \\ \mu_a - t\lambda^* + F_a^*\left(1/2_{\mu_a \pm \sigma_a}\right) &= \mu_b - t\left(\frac{1}{2n_b} - \lambda^*\right) + F_b^*\left(1/2_{\mu_a \pm \sigma_a}\right) \end{aligned} \tag{4.21}$$

After substituting for F_a^* (from Equation 4.15) and F_b^* (from Equation 4.20), this condition can be solved for a quadratic expression of the form[14]

$$p^{*2} + 2ip^* - \psi = 0 \tag{4.22}$$

where $p^* \equiv 2n_a\lambda^*$ denotes the probability an inexperienced consumer buys Brand a for the first time in any period and where $\psi = 2i\alpha^m c_a + \beta c_a^2$. Being a probability, $p^* \geq 0$. As such, p^* is the larger of the two roots satisfying Equation 4.22:

$$p^* = -i + \sqrt{i^2 + 2i\alpha^m c_a + \beta c_a^2} \tag{4.23}$$

where $\alpha^m = 1/2 - \Delta/\tau$ and $\beta \equiv (\alpha^m)^2 + (\sigma_a/\tau)^2$. This is the probability an inexperienced consumer buys from Brand a for the first time.

[13]This condition could also be obtained by setting $\partial F_b^*/\partial\lambda^* = 0$ and solving for λ^*.

[14]The details are contained in the appendix.

4.5 Statement and Explication of the Optimal Policy

In this section an explication of the optimal policy is performed, analyzing the role of various factors which determine the policy, and the implications that such a policy has for consumer-decision theory. However, before proceeding along these lines, a complete statement of the optimal policy is required.

To this end, recall that the optimal policy is a member of Policy-family $\mathcal{P}_1$ (see Definition 4.3 and Table 4.1). Given this fact, and the work of the previous section, the optimal purchase policy, stated in probability form, looks as follows:

$$P^*(x_a) = \begin{cases} p^h & \text{if} \quad x_a = 1_{\mu_a+\sigma_a} \\ p^* & \text{if} \quad x_a = 1/2_{\mu_a\pm\sigma_a} \\ p^l & \text{if} \quad x_a = 1_{\mu_a-\sigma_a} \end{cases} \tag{4.24}$$

where $p^h \equiv 2n_a\lambda^h = \alpha^h c_a$, $p^l \equiv 2n_a\lambda^l = \alpha^l c_a$, and where p^* is given by Equation 4.23.

Parametric Interpretation 4.2 *Recall from Table 4.1 and Parametric Interpretation 4.1 that the α's are the probabilities of purchase from Brand a, conditional on the consumer's primary travel taking her to a community served by Brand a. Since c_a is the probability she visits such a community, it follows that the p's, being the product of c_a and the various α's, give the unconditional probabilities of purchase from Brand a. For example, the product $p^h \equiv \alpha^h c_a$ is the unconditional probability an experienced consumer who has found Brand a to be of high utility-value buys from Brand a in any future period. Thus, while the α's can be interpreted as the "local market probabilities" of purchase, the p's can be interpreted as the "global market probabilities" of purchase. The former probabilities are those directly relevant to a store within the chain, whereas the latter probabilities are those directly relevant to the chain as a whole.*

Quite naturally, emphasis is placed on that portion of the optimal policy which governs the behavior of an inexperienced consumer—i.e. p^*. The job of explication is done by separating the static effect of convenience from its dynamic effect.

4.5.1 The Static Effect of Convenience

Under a myopic purchase policy, the intertemporal nature of the consumer's choice problem plays no role, for the consumer does not take into account the information-value contained in the current purchase. Nonetheless, even under such a policy, the probability that the consumer will purchase from Brand a is still increasing in the size of its chain. The reason is simple availability. A necessary condition for the consumer to purchase from Brand a is that it be available in the community which the consumer visits. Thus the greater the number of communities Brand a serves, the greater the likelihood that it will satisfy this necessary condition. Formally, the probability of first-purchase from Brand a is $p^m = \alpha^m c_a$, where α^m is a constant under a myopic policy. Thus, $\partial p^m / \partial c_a = \alpha^m > 0$.

Since α^m is a constant, however, so are the returns to scale. That is, if consumer's behave myopically, chain size will not affect the probability of first purchase from any particular store within the chain. As such, this probability exhibits constant returns to scale. One might expect, therefore, that chain size holds little import for brand value when consumers behave myopically. However, such is not the case. In particular, the size of a brand's retail chain will affect the present value of each store in the chain. Thus, chain size still matters, even if consumers use a myopic purchase-policy. Unfortunately a demonstration of this result will have to wait, for it requires some additional scaffolding which will not be erected until the next chapter.

4.5.2 The Dynamic Effect of Convenience

When consumers are forward looking, taking into account the value of information contained in their purchase, chain size becomes all the more important, affecting not only the probability of sale to inexperienced consumers by the chain as a whole, but also that by each store in the chain through the information value chain-size creates. Uncertainty and convenience are complementary factors in the production of information value. Due to this dynamic effect, increasing returns to scale prevail under forward-looking policies, like the optimal policy considered in this chapter.

The gross effect of chain size on the probability of first purchase can be found from that portion of the optimal policy which governs the behavior of an inexperienced consumer. In particular, the first partial of 4.23 with respect to chain size yields:

$$\frac{\partial p^*}{\partial c_a} = \frac{i\alpha^m + \beta c_a}{i + p^*} \geq 0$$

Thus, like a myopic policy, an increase in chain size will increase the probability of first-purchase from Brand a. However, unlike the constant returns to scale implied by a myopic policy, when consumers use an optimal forward-looking policy, returns to scale are increasing. This is apparent from the second-order partial:

$$\frac{\partial^2 p^*}{\partial c_a^2} = \frac{(i\sigma_a/\tau)^2}{(i + p^*)^3} \geq 0 \tag{4.25}$$

Equation 4.25 states that the probability an inexperienced consumer buys from Brand a is convex in c_a, implying that there are increasing returns to chain size. Since p^* is convex in market-coverage c_a, it follows that p^*/c_a is increasing in c_a, so the probability of first purchase *for each store* in Brand a's chain is increasing with the size of the chain.[15]

This result can also be found directly. Dividing p^* by c_a gives the probability of purchase from an individual store in Brand a's chain by an inexperienced consumer under an optimal policy. For symmetry of notation, let this value be denoted by α^*; i.e. $\alpha^* \equiv p^*/c_a$. Unlike the local-market probabilities α^h, α^l, and α^m, α^* is not a constant, independent of the size of Brand a's retail chain, as can be seen from the

[15]Somewhat more formally, the elasticity of scale is

$$\varepsilon \equiv \frac{\partial p^*/\partial c_a}{p^*/c_a} = \frac{i\alpha^m c_a + \beta c_a^2}{ip^* + p^{*2}}$$

which exhibits increasing returns if $p^* > p^m$. That is, if $p^* > p^m$ then $\varepsilon > 1$, so a one percent increase in coverage results in a greater than one percent increase in the probability of first purchase. The algebra looks as follows:

$$\begin{aligned} i\alpha^m c_a + \beta c_a^2 &> p^{*2} + ip^* \\ i\alpha^m c_a + \beta c_a^2 &> ip^* - 2ip^* + 2c_a\alpha^m i + \beta c_a^2 \\ p^* &> \alpha^m c_a = p^m \end{aligned}$$

The inequality holds so long as $\sigma_a > 0$. Formally, since $p^m = \alpha^m c_a$ and $\beta \equiv (\alpha^m)^2 + \sigma_a^2/\tau^2$ we can write $p^* = -i + \sqrt{(i + p_m)^2 + (\sigma_a c_a/\tau)^2}$. From which it follows that

$$p^* - p^m = \sqrt{(i + p_m)^2 + (\sigma_a c_a/\tau)^2} - (i + p_m) \geq 0$$

which says that the probability of first-purchase is never any less under an optimal policy. (The inequality follows from squaring both sides.) Naturally, chain size has the effect of increasing this differential in the probability of purchase through its effect on the value of information:

$$\frac{\partial}{\partial c_a}(p^* - p^m) = \frac{\alpha^m i + \beta c_a}{i + p^*} - \alpha^m > 0$$

This inequality can be proved by expanding and rearranging as follows:

$$\begin{aligned} i + \alpha^m c_a + (\sigma_a/\tau)^2 (c_a/\alpha^m) &\geq \sqrt{(i + \alpha^m c_a)^2 + (\sigma_a c_a/\tau)^2} \\ (c_a/\alpha^m)^2 + 2i(c_a/\alpha^m) + c_a^2 &\geq 0 \end{aligned}$$

This is strictly positive for $c_a > 0$.

following first partial:[16]

$$\frac{\partial \alpha^*}{\partial c_a} = \frac{p^* - p^m}{p^* + i}\frac{i}{c_a^2} > 0 \tag{4.26}$$

Thus, an increase in the size of Brand a's retail chain will increase the probability of sale to inexperienced consumers by each store in the chain. These results are summarized in the following proposition:

Proposition 4.4 ***Increasing Returns to Scale.*** *The probability an inexperienced consumer tries Brand a in any period under an optimal purchase-policy is convex in the size of Brand a's retail chain (Equation 4.25). As such, the probability of first-time sale under such a policy exhibits increasing returns to scale—i.e. the probability of first-time sale by each store in Brand a's chain is increasing in the size of its chain (Equation 4.26).*

The difference between an optimal policy and a myopic policy is the consideration the consumer gives to the information contained in the current purchase and the use such information will find in future purchase decisions. Thus the effect of convenience on the value of information can be measured by the change in the probability of purchase under these two policies; i.e. $\alpha^* - \alpha^m$. A feel for the relative importance of information within the context of the present model can be obtained graphically

[16]The following partials are also implied:

$$\frac{\partial \alpha^*}{\partial i} = -\frac{p^* - p^m}{(i + p^*)\, c_a} < 0 \qquad \frac{\partial \alpha^*}{\partial \Delta} = -\frac{i + p^m}{\tau\,(i + p^*)} < 0 \qquad \frac{\partial \alpha^*}{\partial \sigma_a} = \frac{1}{(i + p^*)}\frac{\sigma_a}{\tau^2}c_a > 0$$

Note, as well, that the present model, being more general then the model of discrete geographic space, is also capable of generating a minimum informative scale. In particular, if $\mu_a + F^* > \mu_b - t/2n_b > \mu_a$, the purchase of Brand a would never take place under a myopic policy, but could take place under an optimal policy which considers the value of information F^* contained in the purchase. However, to avoid the use of additional piecewise functions, the simplifying Parametric Restrictions 4.2 where imposed which rule out this possibility. Despite the absence of a minimum informative scale in the continuous-space model, increasing returns to scale still prevail.

using particular numerical values for the parameters. Figure 4.7 graphs $100\,(\alpha^* - \alpha^m)$ against the size of Brand a's chain c_a, using $v \equiv \sigma_a/\tau$ as a shift parameter which measures secondary convenience.[17]

As can be seen, the value of information is positive and increasing in both measures of convenience, implying that an increase in both types of convenience can substantially improve the probability of first-purchase from Brand a.[18]

[17] Recall that $\alpha^h = \alpha^m + \sigma_a/\tau$ and $\alpha^l = \alpha^m - \sigma_a/\tau$ are the conditional probabilities of purchase from Brand a for an experienced consumer who has found Brand a to be of high and low utility-value, respectively. Fix σ_a at some positive constant and let $v \equiv \sigma_a/\tau$. To provide a point of reference, assume the brands would split the market for inexperienced consumers evenly under a myopic purchase policy; i.e. $\Delta \equiv \mu_b - \mu_a = 0$ so $\alpha^m = 0.5$. For $\alpha^m = 0.5$, the restriction $0 \leq \alpha^h < \alpha^l \leq 1$ allows v to range between 0 and 1/2. The graph assumes $i = 5\%$.

[18] If the initial distribution of consumers over geographic space is uniform, then, since the travel pattern of consumers' is also uniform, so too will be the expected distribution of consumers at any point in time. (See also Chapter 5, Section 5.2). As a result, the probability of purchase by inexperienced consumers can also be interpreted as Brand a's expected market share of such consumers. Thus, under this interpretation, Figure 4.7 shows the increase in local market share of inexperienced consumers enjoyed by each store in Brand a's chain.

The effect of primary convenience, c_a, and secondary convenience, τ, on the value of information be found more formally from the following partial derivatives:

$$\frac{\partial\,(\alpha^* - \alpha^m)}{\partial c_a} = \frac{\partial \alpha^*}{\partial c_a} > 0$$

$$\frac{\partial\,(\alpha^* - \alpha^m)}{\partial \tau} = -\left(\frac{c_a\sigma_a^2/\Delta\tau + (p^* - p^m)}{(i + p^*)}\right)\frac{\Delta}{\tau^2} < 0$$

The latter sign holds for $\Delta \geq 0$.

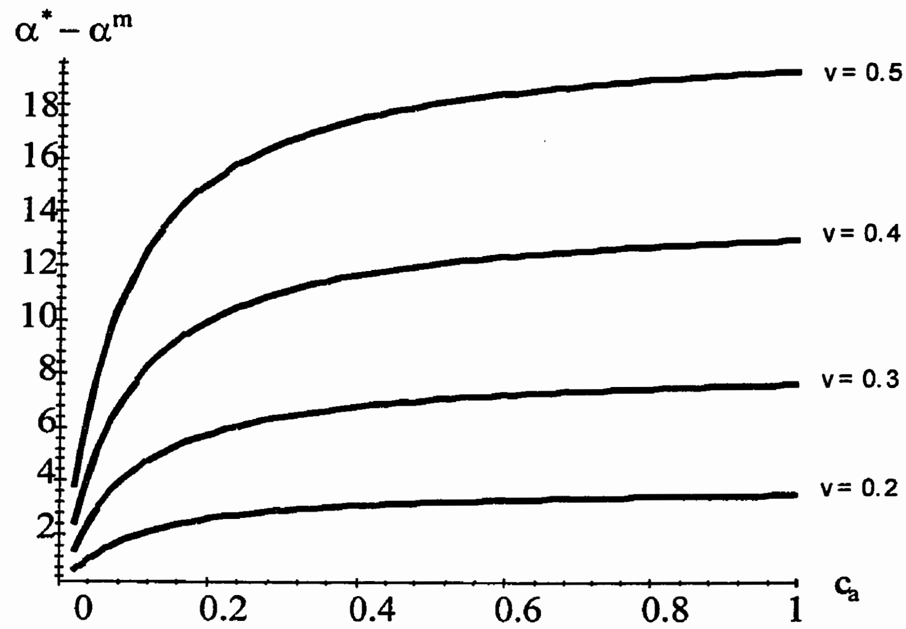


Figure 4.7: Increase in Local Probability of Sale to Inexperienced Consumers due to the Value of Information. The graphs show the value of information as measured by the increase in the probability of purchase by inexperienced consumers enjoyed by each store in Brand a's chain, i.e. $\alpha^* - \alpha$. The level of primary convenience is measured by c_a. The level of secondary convenience is measured by v.

4.6 Appendix: Derivation of Equations

4.6.1 Derivation of Equations 4.6a, 4.6b and 4.6c

Substitution of $\lambda^l \equiv \alpha^l/2n_b$ and $\bar{x}_a = \mu_a - \sigma_a$ into Equation 4.1 yields Equation 4.6a as follows:

$$\begin{aligned}
v_{\Lambda_a}\left(1_{\mu_a-\sigma_a}, \lambda^l\right) &= 2n_b\left(\frac{\alpha^l}{2n_b}\right)\left(\mu_a - \sigma_a - \frac{t}{2}\left(\frac{\alpha^l}{2n_b}\right)\right) \\
&\quad + \left(1 - 2n_b\left(\frac{\alpha^l}{2n_b}\right)\right)\left(\mu_b - \frac{t}{2}\left(\frac{1}{2n_b} - \left(\frac{\alpha^l}{2n_b}\right)\right)\right) \\
&= \alpha^l\left(\mu_a - \sigma_a - \frac{\tau}{4}\alpha^l\right) + \left(1-\alpha^l\right)\left(\mu_b - \frac{\tau}{4}\left(1-\alpha^l\right)\right) \\
&= \mu_b - \frac{\tau}{4} + \frac{\tau}{2}\left(\alpha^l\right)^2
\end{aligned}$$

Substitution of $\lambda^m \equiv \alpha^m/2n_b$ and $\bar{x}_a = \mu_a$ into Equation 4.1 yields Equation 4.6b as follows:

$$\begin{aligned}
v_{\Lambda_a}\left(1/2_{\mu_a\pm\sigma_a}, \lambda^m\right) &= 2n_b\left(\frac{\alpha^m}{2n_b}\right)\left(\mu_a - \frac{t}{2}\left(\frac{\alpha^m}{2n_b}\right)\right) \\
&\quad + \left(1 - 2n_b\left(\frac{\alpha^m}{2n_b}\right)\right)\left(\mu_b - \frac{t}{2}\left(\frac{1}{2n_b} - \left(\frac{\alpha^m}{2n_b}\right)\right)\right) \\
&= \alpha^m\left(\mu_a - \frac{\tau}{4}\alpha^m\right) + \left(1-\alpha^m\right)\left(\mu_b - \frac{\tau}{4}\left(1-\alpha^m\right)\right) \\
&= \mu_b - \frac{\tau}{4} + \frac{\tau}{2}\left(\alpha^m\right)^2
\end{aligned}$$

Substitution of $\lambda^h \equiv \alpha^h/2n_b$ and $\bar{x}_a = \mu_a + \sigma_a$ into Equation 4.1 yields Equation

4.6c as follows:

$$\begin{aligned}
v_{\Lambda_a}\left(1_{\mu_a+\sigma_a},\lambda^h\right) &= 2n_b\left(\frac{\alpha^h}{2n_b}\right)\left(\mu_a+\sigma_a-\frac{t}{2}\left(\frac{\alpha^h}{2n_b}\right)\right) \\
&\quad +\left(1-2n_b\left(\frac{\alpha^h}{2n_b}\right)\right)\left(\mu_b-\frac{t}{2}\left(\frac{1}{2n_b}-\frac{\alpha^h}{2n_b}\right)\right) \\
&= \alpha^h\left(\mu_a+\sigma_a-\frac{\tau}{4}\alpha^h\right)+\left(1-\alpha^h\right)\left(\mu_b-\frac{\tau}{4}\left(1-\alpha^h\right)\right) \\
&= \mu_b-\frac{\tau}{4}+\frac{\tau}{2}\left(\alpha^h\right)^2
\end{aligned}$$

4.6.2 Derivation of Equation 4.20

For brevity, the arguments of $F_a^*\left(1/2_{\mu_a\pm\sigma_a}\right)$ and $F_b^*\left(1/2_{\mu_a\pm\sigma_a}\right)$ will be dropped. Substitution of $E_{\lambda\in\Lambda_a}V^*\left(1/2_{\mu_a\pm\sigma_a},\lambda\right)$ (from Equation 4.18) and $E_{\lambda\in\Lambda_a'}V^*\left(1/2_{\mu_a\pm\sigma_a},\lambda\right)$ (from Equation 4.19) into the expression for F_b^* (Equation 4.17) and multiplying through by $1+i$ yields the following expression:

$$\begin{aligned}
(1+i)\,F_b^* &= c_a\left\{2n_b\lambda^*\left(\mu_a-\tfrac{t}{2}\lambda^*+F_a^*\right)+\left(1-2n_b\lambda^*\right)\left[\mu_b-\tfrac{t}{2}\left(\tfrac{1}{2n_b}-\lambda^*\right)+F_b^*\right]\right\} \\
&\quad +\left(1-c_a\right)\left(\mu_b-\tfrac{t}{4n_b}+F_b^*\right) \\
&= 2n_a\lambda^*\left\{\left(\mu_a-\tfrac{t}{2}\lambda^*+F_a^*\right)-\left[\mu_b-\tfrac{t}{2}\left(\tfrac{1}{2n_b}-\lambda^*\right)\right]\right\}+c_a\left[\mu_b-\tfrac{t}{2}\left(\tfrac{1}{2n_b}-\lambda^*\right)\right] \\
&\quad +\left(1-c_a\right)\left(\mu_b-\tfrac{t}{4n_b}\right)+\left[\left(c_a-2n_a\lambda^*\right)+\left(1-c_a\right)\right]F_b^* \\
&= 2n_a\lambda^*\left(\mu_a-\mu_b+F_a^*+\frac{t}{2n_b}-t\lambda^*\right)+\mu_b-\frac{t}{4n_b}+\left(1-2n_a\lambda^*\right)F_b^*
\end{aligned}$$

Solving for F_b^* yields Equation 4.20:

$$F_b^*=\frac{1}{\left(i+2n_a\lambda^*\right)}\left(\mu_b-\frac{\tau}{4}+2n_a\lambda^*\left(F_a^*+\frac{\tau}{2}-t\lambda^*-\Delta\right)\right)$$

where $\Delta=\mu_b-\mu_a$ and $\tau=t/n_b$.

4.6.3 Derivation of Equation 4.22

Substitution of F_b^* (from Equation 4.20) into the optimality condition (given by Equation 4.21) yields:

$$\mu_a - t\lambda^* + F_a^* = \mu_b - t\left(\frac{1}{2n_b} - \lambda^*\right) + F_b^*$$
$$\mu_a - t\lambda^* - \mu_b + t\left(\frac{1}{2n_b} - \lambda^*\right) + F_a^* = \frac{1}{(i+2n_a\lambda^*)}\left(\mu_b - \frac{\tau}{4} + 2n_a\lambda^*\left(F_a^* + \frac{\tau}{2} - t\lambda^* - \Delta\right)\right)$$
$$(i+2n_a\lambda^*)\left(F_a^* + \frac{\tau}{2} - 2t\lambda^* - \Delta\right) = \mu_b - \frac{\tau}{4} + 2n_a\lambda^*\left(F_a^* + \frac{\tau}{2} - t\lambda^* - \Delta\right)$$

$$i\left(F_a^* + \frac{\tau}{2} - 2t\lambda^* - \Delta\right) + 2n_a\lambda^*\left(F_a^* + \frac{\tau}{2} - 2t\lambda^* - \Delta - \left(F_a^* + \frac{\tau}{2} - t\lambda^* - \Delta\right)\right) = \mu_b - \frac{\tau}{4}$$
$$i\left(F_a^* + \frac{\tau}{2} - 2t\lambda^* - \Delta\right) - 2tn_a\lambda^{*2} = \mu_b - \frac{\tau}{4}$$

Substitution for F_a^* (from Equation 4.15) yields

$$\begin{aligned} i\left(\frac{1}{i}\left(\mu_b - \frac{\tau}{4} + \frac{\tau}{2}\beta c_a\right) + \frac{\tau}{2} - 2t\lambda^* - \Delta\right) - 2tn_a\lambda^{*2} &= \mu_b - \frac{\tau}{4} \\ 2tn_a(\lambda^*)^2 + 2it\lambda^* - \frac{\tau}{2}(2i\alpha^m + \beta c_a) &= 0 \end{aligned}$$

Multiplying through by $2n_a/t$ yields Equation 4.22:

$$p^{*2} + 2ip^* - \psi = 0$$

where $p^* \equiv 2n_a\lambda^*$, $\psi = 2i\alpha^m c_a + \beta c_a^2$ and where $\alpha^m = 1/2 - \Delta/\tau$ and $\beta = (\alpha^m)^2 + (\sigma_a/\tau)^2$.

Chapter 5

Chain Size and the Speed of Convergence to Long-Run Market Share

5.1 The Two Effects of Chain Size: Policy-Dependent and Policy-Independent Effects

Up to now this thesis has focused on the decisions of inexperienced consumers, and how chain size can affect those decisions by determining the value of information contained in a purchase from a spatial brand. This analysis has required rather minimal assumptions. In particular, the reader may have noticed that nothing has been assumed about the actual value of Brand a. The reason for this is rather simple. By definition, an experience good has attributes which are unknown prior to its purchase and consumption. As such, the attributes of the product are irrelevant in explaining the behavior of inexperienced consumers (i.e.an explanation of the first-purchase of an experience good). As was shown, what is relevant is the consumer's purchase policy—specifically, whether this policy considers the value of information contained in a purchase from Brand a, and how the size of its chain can affect that value.[1]

The time has come, however, to consider the role of experienced consumers. So long as inexperienced consumers have a positive probability of purchase, eventually they will become experienced, learning the quality of Brand a. It is the actual quality of Brand a which determines the long-run local-market share enjoyed by each store in its chain. This chapter presents the second half of the theory of spatial branding by showing how chain size can affect the value of a spatial brand. In particular, it is shown that chain size affects the value of a spatial brand in two ways: *(i)* by affecting the value of information contained in a purchase by new or inexperienced consumers (this is the "policy-dependent effect" since it depends on the purchase policy used by consumers), *(ii)* by affecting the speed with which local-market share converges to its long-run value (this is the "policy-independent effect" since this effect remains even if consumers use a myopic policy). A fundamental result derived from *(ii)* is that not all brands will benefit from being big, some brands are better off remaining independent single-store operations. The remainder of this section briefly introduces the analysis to follow.

[1] One might think that such effects require a high level of sophistication on the part of consumers. There is, however, an easy test to determine whether information value is relevant. Suppose, for example, you find yourself in an unfamilar place—perhaps on business—and you need a bite to eat, or a place to sleep, etc. You have a choice between a known brand and one which you have not seen before. Which one would you choose? Is your decision in any way affected by the possibility of returning to this area? In particular, would you be more likely to try the unfamiliar brand if you know you will be returning to this area frequently in the future? If so, then the value of information is relevant—otherwise, it is not.

5.1.1 The Policy-Dependent Effect: The Value of Information

The expected value of any consumer to Brand a is dependent on the likelihood of that consumer's expenditure being directed to Brand a, rather than some rival brand. Under an optimal policy, the probability of purchase from Brand a by an inexperienced consumer is greater than that under a myopic policy since the consumer takes into account the value of information contained in such a purchase and, therefore, is willing to travel farther to make the purchase (recall Figure 4.5). This effect of information-value on the purchase decisions of inexperienced consumers increases the expected value of inexperienced consumers to Brand a and, as will be shown in Section 5.3, increases the present value of each store in Brand a's chain, independent of the quality of its product.

5.1.2 The Policy-Independent Effect: The Speed of Convergence

Under an optimal policy, the probability of purchase by inexperienced consumers exhibits increasing returns to scale. As the size of Brand a's chain increases, so too does the value of information contained in a purchase from Brand a and, therefore, the probability of first-purchase for each store in Brand a's chain. In contrast, under a myopic purchase-policy, returns to scale are constant: the size of Brand a's chain will not affect the probability of first purchase from any store in the chain. One might expect, therefore, that if consumers behave myopically, chain size will hold little import for the value of Brand a. However, such is not the case.

Even if consumers behave myopically, chain size still matters to the value of a spatial brand, for the size of a brand's retail chain determines the speed with which inexperienced consumers are converted into experienced ones. To see this, suppose consumers follow a myopic purchase policy. When an inexperienced consumer buys from *any store* in Brand a's chain she becomes experienced, learning the value of Brand a and using this knowledge in all subsequent purchase decisions involving any other store from that chain. However, since $\lambda^h > \lambda^m > \lambda^l$, this conversion of an inexperienced consumer into an experienced one creates a positive externality for other stores in a high-value chain, but a negative externality for other stores in a low-value chain. Section 5.4 shows that the size of Brand a's retail chain affects the speed of this conversion and, therefore, the present value of each store in the chain. Quite naturally, the relationship between chain size and the present value of each store in the chain is determined by the quality of the brand. In particular, not all brands serving a mobile population benefit from being spatially branded; low-value brands are better off remaining single-store operations. Finally, Section 5.5 extends this result to show why consumers may believe bigger chains are better chains, even before they have tried the chain's product.

5.2 The Expected Time-Path of Sales and the Present-Value of a Chain Store

Before the policy-dependent and policy-independent effects of chain size introduced in the previous section can be demonstrated, the expected time-path of sales for Brand a must be derived. To this end, let t denote the number of periods which have elapsed since Brand a first entered the geographic area. Assume that consumers spend a fixed amount per period, allocating this fixed expenditure to whichever brand is dictated by their purchase policy, location, and beliefs. Under this assumption, the time-path of Brand a's sales will simply be a fixed multiple of the time-path of its market share, where "market share" is defined by the proportion of consumers who allocate their expenditure to Brand a.

Brand a's market share at any point in time will depend on the quality of the Brand a and the following two distributions:

1. The distribution of consumers across geographic space—in particular, the proportion of the population at time t who are in a community served by Brand a. In general, the expected distribution of consumers over geographic space at any future time is determined by the initial distribution of consumers and their travel pattern. However, given the primary travel pattern of consumers is uniform, so too is the expected distribution of consumers at any future time (independent of their initial distribution). As a result, the expected proportion of consumers who find themselves in a community served by Brand a will be equal to the proportion of communities served by Brand a. This proportion is c_a.

2. The distribution of consumers across experience. Experienced consumers choose differently than inexperienced consumers. Thus, the proportion of the population that has tried Brand a prior to time t will be a determinant of the market share of Brand a at time t. The distribution of consumers across experience is endogenous, being derived from the initial distribution of consumers, their travel pattern, and their purchase policy. At any time t there will exist an expected proportion of the population who are inexperienced, having not tried Brand a prior to period t. When consumers follow purchase policy $\pi^k \in \{\pi^m, \pi^*\}$, this proportion is $\left(1 - p^k\right)^{t-1}$, where $p^k \in \{p^m, p^*\}$ denotes the probability an inexperienced consumer buys Brand a under policy k.[2] Of this population, p^k will

[2] Recall that if consumers follow a myopic policy, the expected proportion of inexperienced consumers who buy Brand a is $p^m = \alpha^m c_a$, whereas if consumers follow an optimal policy, this proportion is $p^* = \alpha^* c_a$.

try Brand a for the first time in period t. Thus the expected proportion of the population who are experienced after t periods is

$$\begin{aligned}\epsilon_t^k &= p^k \sum_{s=0}^{t-1} \left(1-p^k\right)^s \\ &= 1-\left(1-p^k\right)^t\end{aligned} \tag{5.1}$$

The market share of Brand a is a convex combination of its share of experienced and inexperienced consumers in the market.[3] As such, if Brand a is of high quality, the expected proportion of the total population who purchase Brand a in period t is

$$\begin{aligned}H_t^k &= \epsilon_{t-1}^k p^h + \left(1-\epsilon_{t-1}^k\right) p^k \\ &= p^h - \left(1-p^k\right)^{t-1}\left(p^h - p^k\right)\end{aligned} \tag{5.2}$$

Whereas if Brand a is of low quality, the expected proportion of the total population which purchase Brand a in period t is

$$\begin{aligned}L_t^k &= \epsilon_{t-1}^k p^l + \left(1-\epsilon_{t-1}^k\right) p^k \\ &= p^l + \left(1-p^k\right)^{t-1}\left(p^k - p^l\right)\end{aligned} \tag{5.3}$$

Equations 5.2 and 5.3 describe alternative time-paths for the expected global-market share of Brand a under alternative assumptions about its quality and the purchase policy of consumers.

Our interest lies mainly in the relationship between the size of chain and the sales of a representative store in the chain (i.e. with scale effects). To this end, let $h_t^k \equiv H_t^k/c_a$ and $l_t^k \equiv L_t^k/c_a$ denote the expected local-market shares for a representative outlet of

[3]Recall that the probability an experienced consumer buys Brand a in any period is p^h if the consumer finds Brand a to be of high quality, but p^l if she finds it to be of low quality.

Brand a under alternative assumptions about its quality and the purchase policy of consumers.[4]

The expected present value of a representative store in the chain of Brand a is found by capitalizing the time-paths given by h_t^k and l_t^k. Let W_h^k and W_l^k denote these present values for Brand a under alternative assumptions about its quality and the purchase policy of consumers. These values are[5]

$$\begin{aligned} W_h^k &= \sum_{t=1}^{\infty} \frac{h_t^k}{(1+i)^t} \\ &= \left(\frac{i+p^h}{i+p^k}\right)\left(\frac{p^k/c_a}{i}\right) \end{aligned} \tag{5.4}$$

$$\begin{aligned} W_l^k &= \sum_{t=1}^{\infty} \frac{l_t^k}{(1+i)^t} \\ &= \left(\frac{i+p^l}{i+p^k}\right)\left(\frac{p^k/c_a}{i}\right) \end{aligned} \tag{5.5}$$

Multiplication of these values by the aggregate expenditure of consumers per period in the local market yields the expected present value of a representative store in Brand a's chain. Note that the upper and lower bounds on these values are, respectively, $W_h = \alpha^h/i$ and $W_l = \alpha^l/i$. These are the values which would be earned if consumers had perfect information prior to purchase and are, therefore, independent

[4]Note that for a myopic policy these equations can be simplified somewhat. Since $p^h = \alpha^h c_a$ and $\alpha^h = \alpha^m + v$, we have

$$h_t^m = \alpha^m + \left(1 - (1-\alpha^m c_a)^{t-1}\right) v \qquad l_t^m = \alpha^m - \left(1 - (1-\alpha^m c_a)^{t-1}\right) v$$

[5]These equations are derived in the appendix to this chapter. For simplicity it is assumed that producers and consumers have the same discount rate. This assumption is only relevant, however, when consumers use an optimal policy since a myopic policy is independent of the consumer's rate of discount.

of the purchase policy used.

5.3 The Policy-Dependent Value of Spatial Branding

The qualitative nature of the time paths given by h_t^k and l_t^k is illustrated in Figure 5.1.[6] The solid lines in Figure 5.1 represent the time-paths under the myopic policy. These curves start on the vertical axis at a local market share of $\alpha^m = 50\%$, and fall or rise from there depending on the underlying quality of Brand a. While the height of the vertical intercept for the myopic time-paths is independent of the size of Brand a's chain, such is not the case for the time-paths under the optimal policy. These time-paths are represented by the dashed lines. The vertical intercept for these curves is positively related to the size of Brand a's chain through the value of information contained in a purchase from Brand a. This information-value induces an inexperienced consumer to travel farther to buy from Brand a under an optimal policy than under a myopic policy. As such, the curves for an optimal policy start at a higher market share of $\alpha^* > \alpha^m$.[7]

[6] The graphs assume Brand a and Brand b would evenly split the market for inexperienced consumers when such consumers behave myopically; i.e. $\Delta \equiv \mu_b - \mu_a = 0$ so $\alpha^m = 50\%$. Note that for $\alpha^m = 50\%$, the simplifying restriction $0 \leq \alpha^l < \alpha^h \leq 1$ imposed in the Hotelling model of the last chapter allows v to range between 0 and 1/2. The graph assumes $c_a = 10\%$, $i = 1\%$ and $v = 1/2$, but is representative of graphs under alternative parameter values.

[7] For Figure 5.1 $\alpha^* = 68\%$. Under the assumed parameter values $\alpha^m = 1/2$ and $v = 1/2$, we have $\beta = (\alpha^m)^2 + v^2 = 1/2$. Along with the other values $c_a = 10\%$ and $i = 1\%$ it follows that

$$\begin{aligned} \alpha^* &\equiv \frac{p^*}{c_a} \\ &= \frac{-0.01 + \sqrt{0.01^2 + 2\,(0.01)\,(0.5)\,(0.10) + (0.5)\,(0.10)^2}}{0.1} = 0.681 \end{aligned}$$

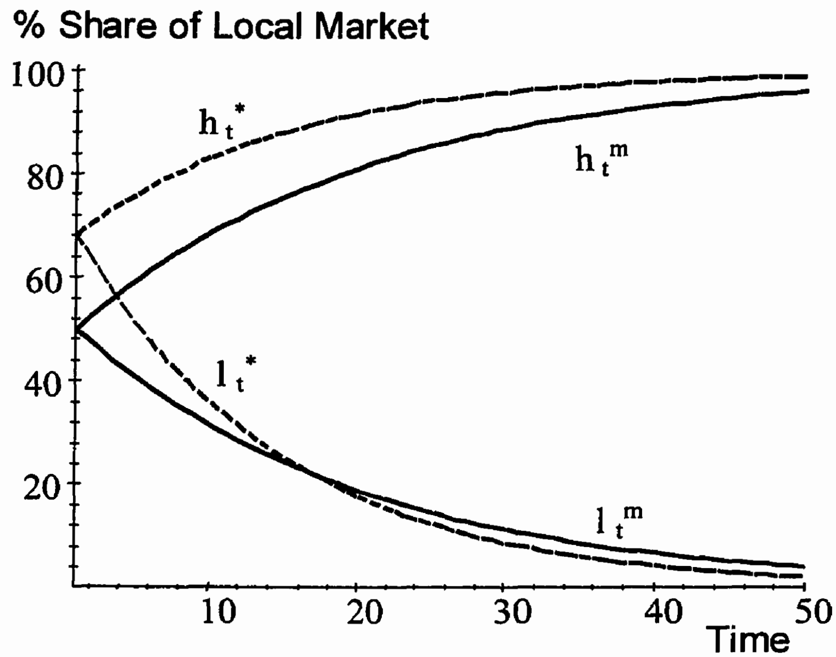


Figure 5.1: The Effect of Information Value on the Expected Time Path of Local Market Share. The solid line is under a myopic policy; the dotted line is under an optimal policy. If Brand a is of high quality, the upward-sloping lines are relevant; if Brand a is of low quality, the downward-sloping lines are relevant.

If Brand a is of high quality, its long-run local-market share of experienced consumers is $\alpha^h = \alpha^m + v$; whereas if it is of low quality, this share is $\alpha^l = \alpha^m - v$, where $v \equiv \sigma_a/\tau$. These long-run shares are independent of the purchase policy of consumers. That is, both h_t^* and h_t^masymptotically approach their common upper bound of α^h, while both l_t^* and l_t^m asymptotically approach their common lower bound of α^l.[8] However, while these long-run shares are independent of the purchase policy of consumers,

[8]Under the parameter values assumed for Figure 5.1 these long-run market shares are $\alpha^h = 1$ and $\alpha^l = 0$. More moderate values of secondary convenience, v, would squeeze the upper and lower bounds together.

the time it takes for the local-market shares to converge to these long-run values is not.

The speed of convergence is a measure of the speed of consumer learning, and is primarily determined by two factors: *(i)* the purchase policy used by consumers, and *(ii)* the size of Brand a's retail chain. When consumers take into account the value of information, they are more likely to purchase Brand a sooner rather than later. Sales earlier in time are especially important for it improves liquidity in the early years of entry into a new geographic area when a good credit history has yet to be established. This should not only improve the brand's income statement, but its balance sheet as well. Formally, the percentage increase in present value is as follows:[9]

$$\frac{W_h^*}{W_h^m} - 1 = \frac{W_l^*}{W_l^m} - 1 = \left(\frac{i + p^m}{i + p^*}\right) \frac{p^*}{p^m} - 1 > 0 \tag{5.6}$$

Thus, an optimal policy increases Brand a's market share in the early periods, independent of its quality. This result is summarized by Proposition 5.1 and is illustrated by the graphs of Equation 5.6 in Figure 5.2 for various levels of primary and secondary convenience.[10]

Proposition 5.1 ***Information Value and Brand Value.*** *Given the size of Brand a's retail chain, the present value of each store in its chain is higher if consumers purchase under an optimal policy than if they behave myopically. Further, this result is independent of the quality of Brand a's product.*

[9]Since $p^* > p^m$, the inequality can be easily shown to follow directly.

[10]Note that the value of information is not just significant when small discount rates are used in the consumer's optimal purchase policy. For example, this graph assumes $i = 20\%$ yet the difference in present values due to the value of information is still significant. Obviously, this difference would be even greater under smaller rates of discount since such rates would increase the probability of first-purchase. Finally, note that this graph provides an alternative measure of the value of information to that presented in Figure 4.7 of the last chapter

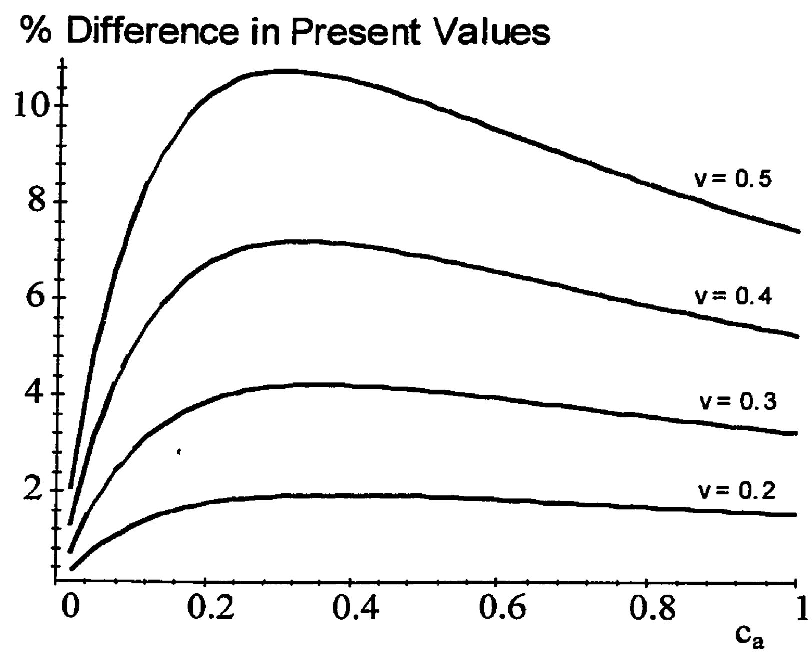


Figure 5.2: The Value of Information. These graphs (of Equation 5.6) show the expected increase in the present value for each store Brand a's chain if consumers use an optimal rather than a myopic purchase policy.

While the purchase policy of consumers can affect the value of a spatial brand, so too can chain size. Chain size has both a direct and an indirect effect on the speed with which local-market share converges to its long-run value. The indirect effect of chain size on the speed of convergence is policy-dependent, and is a direct implication of the analysis above. In particular, as was noted above, if consumers are forward-looking (e.g. following an optimal policy), the greater the size of Brand a's retail chain, the greater the value of information, and the sooner inexperienced consumers purchase this brand. But the sooner consumers try Brand a, the sooner its market share will converge to its long-run value. Thus, under an optimal policy, consumers try Brand a sooner, so the time taken for convergence to the long-run market share is decreasing in chain size.

The direct effect of chain size, however, is policy-independent, coming through the simple availability of brand. Given the travel patterns of consumers are primarily exogenous to the demand for convenience goods, and given consumers will not travel far off the primary travel routes to obtain such goods, more stores means greater availability of the brand and, therefore, a greater likelihood of purchase by all consumers, including inexperienced ones. The greater the size of Brand a's chain, the better the coverage it provides for the consumer's primary travel pattern. As a result, the consumer is more likely to visit a community served by Brand a and, therefore, purchase its product. As the probability of purchase in any period increases, the expected time between purchases falls. Thus, much like an optimal policy, an increase in chain size also gets consumers to purchase the brand sooner rather than later.

One might expect that sooner would be preferred to later and, in particular, that the present value of each store in Brand a's chain should be increasing in the size of its chain. Such, however, is not necessarily the case, as will be shown in the next section.

5.4 The Policy-Independent Value of Spatial Branding

Under an optimal policy, the probability of purchase by inexperienced consumers exhibits increasing returns to scale. As the size of Brand a's chain increases, so too does the value of information contained in a purchase from Brand a. Inexperienced consumers are thus willing to travel farther off their primary travel route to try Brand a and, as a result, the probability of first-purchase for each store in Brand a's chain rises with the size of the chain. In contrast, under a myopic purchase-policy, returns to scale are constant: the size of Brand a's chain will not affect the probability of first purchase from any store in the chain. Does this mean, therefore, that if consumers behave myopically, the value of each store in Brand a's chain is independent of the size of the chain?

Certainly not. To see this, suppose consumers follow a myopic policy. Under a myopic purchase policy the consumer does not take into account the information-value associated with the current purchase. Nonetheless, even under such a policy, as the size of Brand a's chain increases, so too does its availability and, therefore, the likelihood of its purchase.[11] Further, when an inexperienced consumer buys from *any store* in Brand a's chain she becomes experienced, learning the value of Brand a and using this knowledge in all subsequent purchase decisions involving any other store from the chain. Naturally, whether this purchase is beneficial to other stores in the chain depends on the quality of Brand a.

If, for example, Brand a is of high quality, the sooner the consumer tries Brand a the better; for the sooner she tries it, the sooner her decision rule dictates its purchase for all locations $\lambda \leq \lambda^h$, rather than for all locations $\lambda \leq \lambda^m$.[12] Since $\lambda^h > \lambda^m$, a purchase from any store in Brand a's chain conveys a positive externality on all other stores in the chain. Chain size affects the speed of consumer-learning. The greater the size of Brand a's chain, the greater the speed with which inexperienced consumers are converted into experienced ones and, therefore, the greater the speed with which the local-market share of each store in Brand a's chain converges from α^m to its upper bound of α^h. Thus, if Brand a is of high quality, each store in its chain has a vested interest in getting consumers to try the brand as soon as possible, and greater market coverage does just that—even if consumers behave myopically. For a high-quality

[11] Formally, $p^m = \alpha^m c_a$ and $\partial p^m / \partial c_a = \alpha^m > 0$, thus the probability of first-purchase from Brand a increases with the size of its chain.

[12] Recall Figure 4.3.

brand, a bigger chain is a better chain.[13]

As one might expect, however, the reverse is the case if Brand a is of low quality. Now the expected value of inexperienced consumers is greater than that of experienced ones. Their expected value is higher, since their probability of purchase is higher: purchasing for all locations $\lambda \leq \lambda^m$, rather than for all locations $\lambda \leq \lambda^l$. Since $\lambda^l < \lambda^m$, a purchase from any store in Brand a's chain conveys a negative externality on all other stores in the chain. While experienced consumers will still purchase from Brand a even if its quality is lower than that of Brand b, they will not go as far out of their way to make such a purchase; they will not deviate as much from their primary travel pattern.

The greater the size of Brand a's chain, the greater the speed with which inexperienced consumers are converted into experienced ones. However, if Brand a is of low quality, this means the greater the speed with which the local-market share of each store in Brand a's chain converges from α^m to its lower bound of α^l. When Brand a is of low quality, it is better off having just one store through which inexperienced consumers come to learn the value of the product. Inexperienced consumers are virgins: they can only be converted once, and once converted, they will never view Brand a in the same light. Best not to spread that bad reputation around by sharing a common brandname; better to let each store develop a bad reputation on its own under a different brandname. Spatial branding means spreading the high-valued inexperienced consumers over a greater number of stores, thereby reducing the present value of each store. It follows that a low-quality brand does not benefit from being spatially branded; it would be far more profitable to remain an independent

[13]This result holds when consumers use a myopic policy; however, it is even stronger when they use an optimal policy, for an increase in chain size has the additional effect of increasing the probability of first-purchase by inexperienced consumers. As a result, inexperienced consumers are converted into experienced ones even sooner under an optimal policy. If Brand a is of high quality, the sooner an inexperienced consumer becomes an experienced one, the better.

single-store operation, or have each store under common ownership use a different brandname.

The effect of chain size on the speed with which local-market share converges to its long-run value can be seen in the panels of Figure 5.3. These panels plot the time-paths of local market share for Brand a under alternative assumptions about its quality. Panel A plots the graphs of h_t^m for various sizes of the chain, while Panel B does the same for l_t^m. The graphs assume consumers behave myopically, but the effects of chain-size are even more pronounced under an optimal purchase policy (recall Figure 5.1). As can be seen, in both cases a larger chain means consumers try and learn the value of Brand a sooner. Since consumers learn the value of Brand a sooner, the local market share enjoyed by each store in Brand a's chain converges sooner to its long-run asymptotic value.[14] However, while sooner is better than later when Brand a is of high quality, such is not the case when Brand a is of low quality.

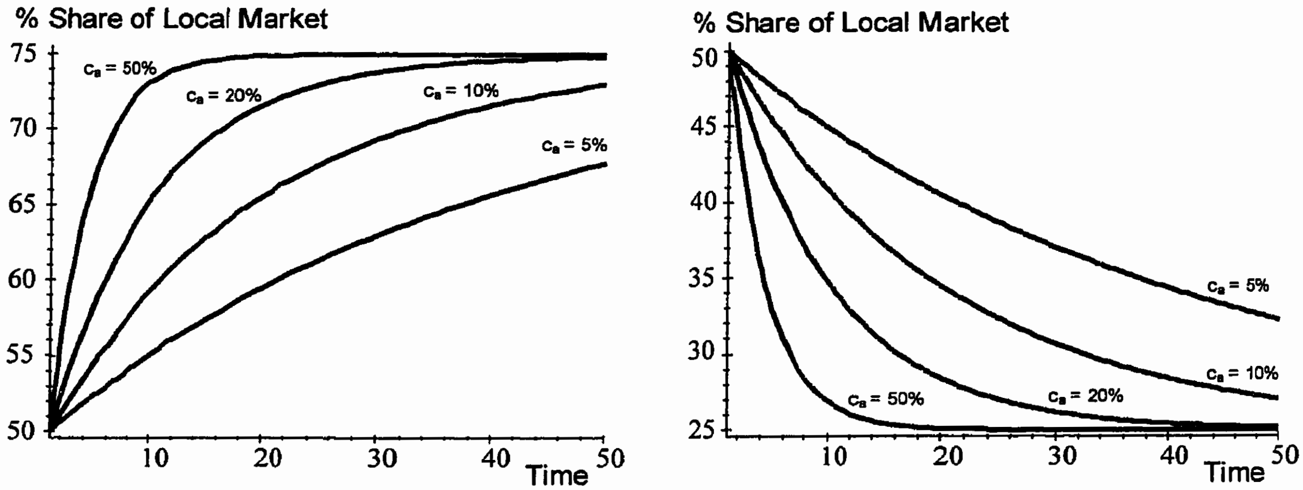
It is useful to state the present value of an individual store in Brand a's chain on a standard metric. Since we are interested in the role of the chain in consumer learning (i.e. the conversion of inexperienced consumers into experienced ones), a good metric to use is the value which would be obtained if consumers never learned (i.e. forget after purchase). This value is $W_0 \equiv \alpha^m/i$. Thus, the standardized present values are $\omega_h^k = W_h^k/W_0 - 1$ and $\omega_l^k = W_l^k/W_0 - 1$. These values for optimal and myopic policies are given by Equations 5.7 and 5.8, respectively, and are plotted in Figure 5.4.[15]

$$\omega_h^* = \left(\frac{i+p^h}{i+p^*}\right)\frac{p^*}{p^m} - 1 \qquad \omega_l^* = \left(\frac{i+p^l}{i+p^*}\right)\frac{p^*}{p^m} - 1 \tag{5.7}$$

[14]As in Figure 5.1, the graph assumes the market for inexperienced consumers is split evenly between brands a and b, i.e. $\alpha^m = 50\%$. Here, however, $v = 25\%$ so the long-run market shares are $\alpha^h = 75\%$ and $\alpha^l = 25\%$.

[15]For illustrative purposes, values of $i = 0.20$ and $v = 0.5$ were used.

Figure 5.3: Chain Size and the Speed of Convergence



Panel A: Brand a is of High Quality **Panel B:** Brand a is of Low Quality

$$\omega_h^m = \frac{p^h - p^m}{i + p^m} = \frac{vc_a}{i + \alpha c_a} \qquad \omega_l^m = \frac{p^l - p^m}{i + p^m} = -\frac{vc_a}{i + \alpha c_a} \tag{5.8}$$

The graphs demonstrates the common-sense result that a high-quality brand benefits from consumers learning the value of the product, whereas a low-quality brand does not. Greater chain size increases the speed with which consumers learn. As such, chain size is a "good" for a high-quality brand, but a "bad" for a low-quality brand, regardless of whether consumers use a myopic or an optimal purchase policy.

The results of this section are formally summarized by the following proposition and its corollary:[16]

[16]Proposition 5.2 can be more formally demonstrated as follows. The time t_c it takes for the local-market share of a store from Brand a's chain to converge within an ϵ (i.e. a small fraction) of its long-run value can be found as follows:

$$\begin{aligned} h_{t_c}^m &\geq (1-\epsilon)\,\alpha^h \\ \alpha^m + \left(1 - (1-\alpha^m c_a)^{t_c - 1}\right) v &\geq (1-\epsilon)\,(\alpha^m + v) \\ (1-\alpha^m c_a)^{t_c-1} v &\leq \epsilon\,(\alpha^m + v) \end{aligned}$$

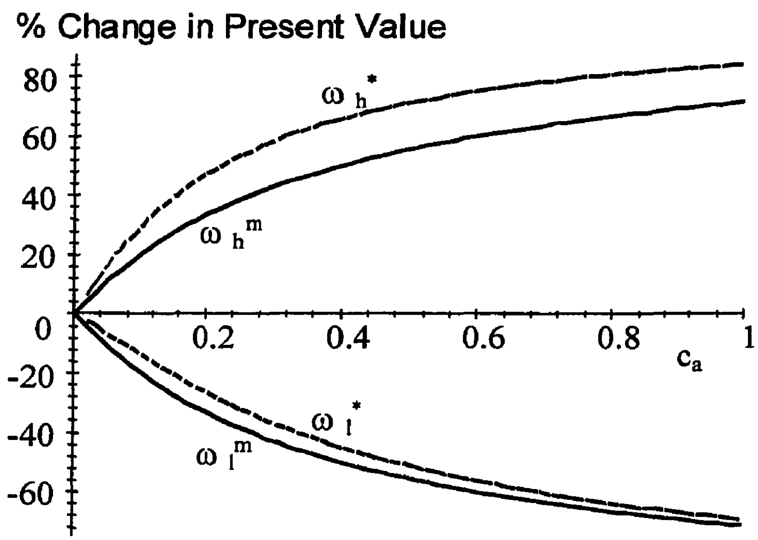


Figure 5.4: The Effect of Chain Size on the Present Value of a Member Store. The graphs measure the percentage change in present value due to consumer learning. The present value of the store is stated relative to that which would be earned if consumers never learned (i.e. forget after purchase). The solid line is under a myopic policy; the dotted line is under an optimal policy. The upward sloping lines are relevant for a high-quality brand; whereas, the downward sloping lines are relevant for a low-quality brand.

Proposition 5.2 ***Speed of Convergence.*** *The time taken for the local-market share of a spatially-branded experience good to converge to its long-run value is decreasing in the size of the brand's chain.*

From Proposition 5.2, the following corollary is implied:

This can be solved to obtain

$$t_c \geq 1 + \frac{\ln\left(\frac{\epsilon(\alpha^m + v)}{v}\right)}{\ln\left(1 - \alpha^m c_a\right)}$$

which is obviously decreasing in c_a. The algebra is similar if Brand a is of low quality.

Corollary 5.1 ***The Value of Spatial Branding.*** *The present value of a store from a high-quality (low-quality) experience-brand is increasing (decreasing) in the size of the brand's chain. Thus, spatial branding has value to high-quality brands, but not to low-quality brands.*

Given the analysis above, it should be clear that chain size is a variable that is chosen differently by high-quality and low-quality brands. Consumers can observe chain size and seem to understand that a brand would not be big unless it is good. This idea is the topic of the next section.

5.5 Signalling Quality with Chain Size

Up to now, the consumer's prior belief about the quality of Brand a has been taken as fixed, believing high and low quality to be equally likely, irrespective of the size of its chain. Obviously, in light of the discussion of the previous sections, such an assumption presumes consumers are somewhat unsophisticated for, as was shown in those sections, spatial branding only has value to Brand a if it is of high quality.

The present section goes at least part way to relaxing the assumption of exogenous beliefs by acknowledging a fundamental constraint which the size of Brand a's chain imposes on its quality. In particular, given the size of its chain, Brand a is only able to earn nonnegative profits if its quality exceeds some minimum. This minimum quality puts a lower bound on the distribution of possible qualities for that chain size. Further, since this minimum quality is an increasing function of chain size, it follows that, the greater the size of Brand a's chain, the greater the lower bound on the distribution of possible qualities and, therefore, the greater the prior expected quality of Brand a.

To see this, let the set of possible qualities (surplus-values) for Brand a be continuous, rather than discrete, as has been assumed up to this point. Let q denote the true quality of Brand a and assume the consumer's prior belief about q is uniform over support $[q_{\min}(c_a), 1]$, where $q_{\min}(c_a) \in [0,1]$ is the minimum quality able to earn non-negative profits at chain size c_a.[17] Under these assumptions, Brand a's local-market share of experienced consumers is then equal to its quality q, while its local-market share of inexperienced consumers is $\alpha^m = (1+q_{\min})/2$.[18] The minimum quality $q_{\min}$ is implicitly defined as a function of c_a by the following zero-NPV (net-present-value) condition:

$$\begin{aligned} \Pi(c_a, q_{\min}) &= 0 \qquad (5.9) \\ \left[\left(\frac{i + q_{\min}c_a}{i + (1+q_{\min})\,c_a/2}\right)\left(\frac{(1+q_{\min})/2}{i}\right)\rho - 1\right]K &= 0 \end{aligned}$$

where $\rho \equiv R/K$, R is rents (revenue less avoidable costs) per period, and K is the sunk capital costs.[19] This identity can be solved for $q_{\min}(c_a)$, the graph of which is presented in Figure 5.5 along with that for average quality $\alpha^m(c_a)$.[20][21]

Figure 5.5 shows the minimum quality which is financially viable at each chain size.

[17] The simplifying restrictions of the Hotelling model require $q_{\min} \in [\mu_b - \tau/2, \mu_b + \tau/2]$, so the assumption of $q_{\min} \in [0,1]$ requires $\mu_b = 1/2$ and $\tau = 1$. As noted in the text, Brand a's local-market share of experienced consumers is then equal to its quality q. Formally,

$$\alpha^q = \frac{1}{2} + \frac{q - \mu_b}{\tau} = q$$

[18] Since the qualitative-nature of the results are insensitive to the purchase policy of inexperienced consumers, a myopic policy is used.

[19] Expression 5.9 is a generalization of present-value expressions 5.4 and 5.5 to the case of continuous quality. Note that R is the rents which would be earned by a representative store of Brand a if it captured the entire market. For simplicity, both R and K are assumed to be independent of quality.

[20] The equations for these curves are as follows:

$$q_{\min}(c_a) = \frac{1}{2\rho c}\left((i-\rho)\,c - i\rho + \sqrt{(\rho i + 6ic - 2\rho c)\,\rho i + (\rho + i)^2 c^2}\right)$$

$$\alpha^m(c_a) = \begin{cases} \frac{1}{2} + \frac{q_{\min}(c_a)}{2} & \text{if } q_{\min}(c_a) \geq 0 \\ \frac{1}{2} & \text{if } q_{\min}(c_a) < 0 \end{cases}$$

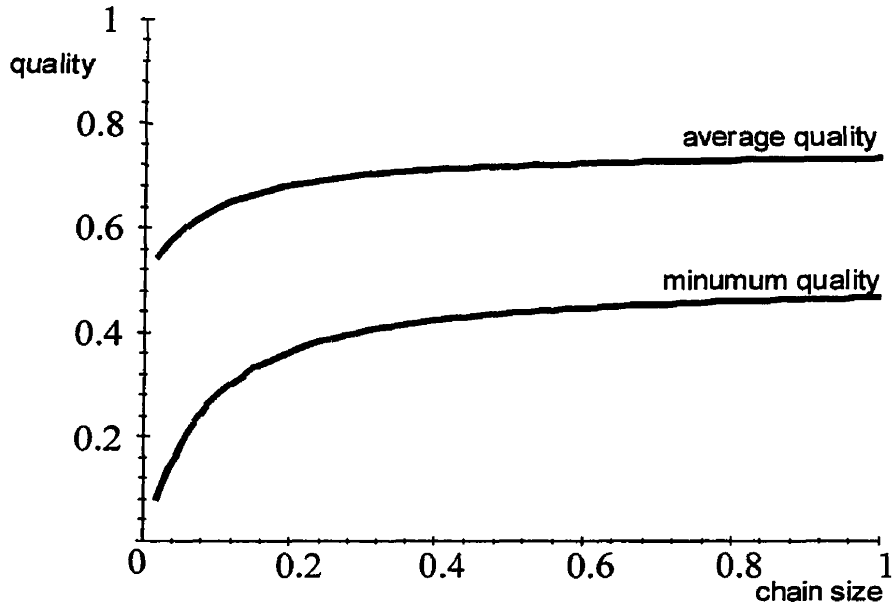


Figure 5.5: Chain Size and the Prior Beliefs of Consumers. The graph shows how the zero-profit minimum quality increases with chain size. Since qualities below the minimum are not feasible this puts a lower bound on the prior beliefs of consumers. Since this lower bound is increasing in chain size, so too is prior expected (or average) quality.

This minimum quality puts a lower bound on the distribution of possible qualities for each chain size. Since the minimum quality is increasing in chain size, so too is the

---

The graph was constructed using values of $\rho = 0.2$ and $i = 0.1$.

[21] It is important to note that the argument of this section can also be made more formally, using the concept of a Bayesian-Nash mixed-strategy equilibrium. In particular, equivalent results are obtained if Brand a is modelled as choosing a profit maximizing quality and chain size given the prior belief of the (representative) consumer, and the probability distribution representing the consumer's prior belief is such that all qualities yield the same profits to Brand a given the size of its chain. In such an equilibrium, Brand a randomly chooses a quality which may be different from what consumer expects given the size of its chain, but the consumer's belief is correct "on average" (i.e. in equilibrium, the probability distribution representing Brand a's mixed strategy is identical to the distribution representing the consumer's prior belief). Again, the equilibrium expected quality in such a model is increasing in chain size and maps a graph very similar to that supported by the much simpler argument of this section.

prior expected (or average) quality.

It is important to note that chain size is here conveying pre-purchase information about quality to an inexperienced consumer, who has yet to try the brand. This is an entirely different role for chain size than that argued in previous chapters and, in particular, is not dependent on the forward-looking nature of consumers. This argument formalizes the idea that "A brand would not be big, unless it is good." However, the argument also contains within it a form of the minimum-informative scale result of Proposition 3.2. In particular, if a brand producing the lowest possible quality (i.e. $q = 0$) can earn a nonnegative NPV at some chain size, then chain size will not be informative, at least over some range. The minimum-informative scale for a chain is found by setting $q_{\min} = 0$ in Equation 5.9 and solving for c_a. This yields $c_a^{\min} = \rho - 2i$, which is strictly positive for $i < \rho/2$.

The minimum-informative scale for $\rho = 0.3$ and $i = 0.1$ is depicted in Figure 5.6, where the curve labelled "$q = 0\%$" is gives all combinations of quality and chain size yielding a zero NPV for a brand of minimum quality (i.e. the zero iso-NPV curve for $q = 0$). Since a minimum quality brand can earn a nonnegative return for $c_a \leq c_a^{\min}$, small chains are uninformative over this range, so the prior expected quality is 1/2 (i.e. uniform prior over $[0, 1]$).

The results of this section are summarized by the following proposition:

Proposition 5.3 ***Big Chains are Better Chains.*** *For chains above minimum-informative size, the minimum quality which permits nonnegative profits is increasing in the size of the chain. As such, so too is the average quality expected by consumers.*

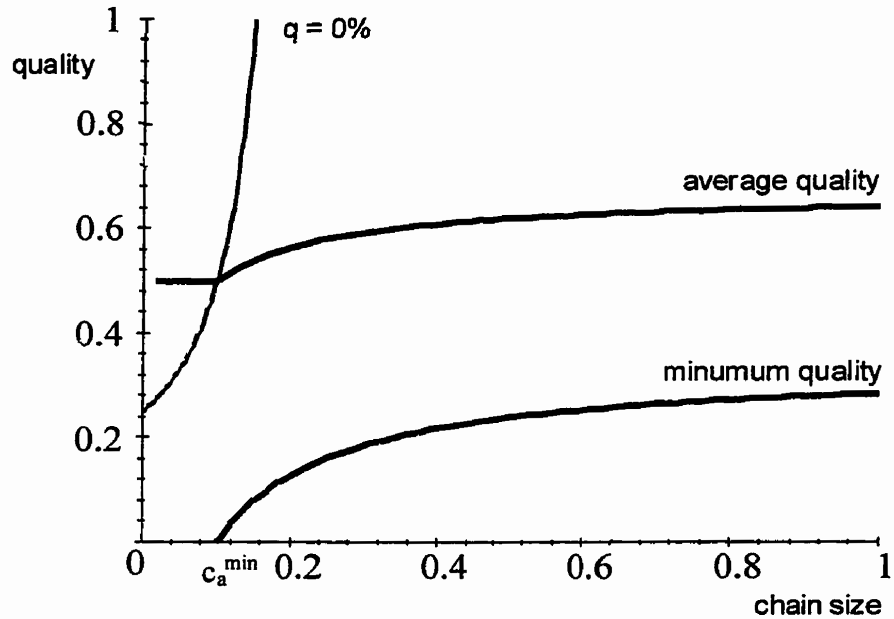


Figure 5.6: Minimum-Informative Scale, Again. Only chain sizes greater than that at which the lowest-quality brand can survive are informative.

5.6 Appendix

5.6.1 Derivation of Equations 5.4 and 5.5

Here W_h^k will be derived; W_l^k can be derived similarly. Let $\delta^t \equiv 1/\left(1+i\right)^t$. Substitution of h_t^k into the first line of Equation 5.4 yields:

$$\begin{aligned} W_h^k &= \frac{1}{c_a}\sum_{t=1}^{\infty}\delta^t\left(p^h - \left(1-p^k\right)^{t-1}\left(p^h - p^k\right)\right) \\ &= \frac{1}{c_a}\left(p^h\sum_{t=1}^{\infty}\delta^t - \left(p^h - p^k\right)\sum_{t=1}^{\infty}\delta^t\left(1-p^k\right)^{t-1}\right) \end{aligned}$$

But

$$\sum_{t=1}^{\infty} \delta^t = \delta \sum_{t=0}^{\infty} \delta^t = \frac{\delta}{1-\delta} = \frac{1}{i}$$

and

$$\sum_{t=1}^{\infty} \delta^t \left(1-p^k\right)^{t-1} = \delta \sum_{t=0}^{\infty} \delta^t \left(1-p^k\right)^t = \frac{\delta}{1-\delta\left(1-p^k\right)} = \frac{1}{i+p^k}$$

Thus, by substitution, we get the second line of Equation 5.4:

$$\begin{aligned} W_h^k &= \frac{1}{c_a}\left(\frac{p^h}{i} - \frac{p^h - p^k}{i+p^k}\right) \\ &= \left(\frac{i+p^h}{i+p^k}\right)\left(\frac{p^k/c_a}{i}\right) \end{aligned}$$

Chapter 6

Testing the Theory of Spatial Branding: Rents for McDonald's Franchisees?

McDonald's Corporation, with over 13,000 restaurants in 65 countries, is the prevailing food and franchising organization in the world. Serving 96 percent of American consumers every year, and 7 percent of the U.S. population on a daily basis, it is the most popular eatery in the country, drawing more customers than its closest competitors, Burger King, Wendy's and Hardee's combined. McDonald's, with total worldwide sales in 1992 over $20 billion, has succeeded beyond anyone's dreams with a method consisting of a simple menu, a speedy food preparation and delivery system based on assembly-line techniques, and above all, a high level of consistency that is found in each of its restaurants around the world.(Shook and Shook 1993, p. 139)

6.1 Introduction

Of all retail chains, McDonald's is, without a doubt, one of the most successful. In an article appearing in the *Journal of Law and Economics* (1994) entitled "Costs of Control: The Source of Economic Rents for McDonald's Franchisees", Patrick J. Kaufmann and Francine Lafontaine (hereafter K-L) used operating data from a large sample of McDonald's outlets to argue that the net present value of obtaining a franchise is positive and substantial. In their words:[1]

> [W]e find that the present value of the amount of ex ante rents that the owner of a new McDonald's franchise can expect to earn, after taxes, over the 20-year period covered by the franchise contract is between $300K and $455K in 1982 dollars. The amount of ex post rents is, of course, even larger as it includes not only the above amounts but also all up-front fees and specific investments. (Kaufmann and Lafontaine 1994, p. 420)

They go on to state,

> This conclusion, that the majority of franchisees earn ex ante rents, is not sensitive to reasonable changes in our assumptions.(Kaufmann and Lafontaine 1994, p. 427)

At stake is not just pure empiricism. K-L rationalize their findings by appeal to the theory of principal-agent relations. They state:

> Following Mathewson & Winter (Mathewson and Winter 1985), we propose that the ex ante rents represent the cost that McDonald's must

[1] "Rents" are the difference between revenues and avoidable costs. "Ex ante" and "Ex post" in this context mean, respectively, "prior to" and "posterior to" the incurrence of up-front, sunk investments necessary to get an outlet up and running. The costs associated with these up-front investments are amortized over the life of the investment and represent the difference between ex ante and ex post rents.

> incur to maintain the stream of ex post rents necessary to achieve strict control over its system in the face of downstream liquidity constraints. In other words, ex ante rents arise here because the franchisor needs to leave a stream of ex post rents downstream to create franchisee incentives, and franchisees' wealth constraints prevent the up-front extraction of the full net present value of these ex post rents.(Kaufmann and Lafontaine 1994, p. 420)

In collecting the necessary data and making this argument, K-L have produced one of the most informative articles on franchising to date. Nonetheless, this chapter has two purposes. First, to show that K-L's conclusion is, in fact, highly sensitive to reasonable changes in their assumptions. In fact, by replacing just one of their assumptions with an alternative supported by the Theory of Spatial Branding, all ex ante rents disappear and an entirely different implication follows from their data. Second, to show that their own data contradicts their assumption and supports the suggested alternative. As such, this chapter tests the Theory of Spatial Branding. It will be shown that the theory passes this test, in that it is not refuted by data independently collected for exogenous purposes.[2]

6.2 The Data and Assumptions

K-L obtained cost-data for 1982 operating year on 1,283 U.S. company-owned units open for 13 months or more.[3] Based on this data, they calculate ex ante and ex post

[2]Note that the argument presented herein should not be interpreted as leading to the conclusion that McDonald's franchisees do not earn ex ante rents. McDonald's franchisees may in fact earn such rents. Rather, the present argument simply shows that the existence of such rents does not necessarily follow from the argument presented by K-L.

[3]They obtained their data from the *Franchise Offering Circular*, a publication of the McDonald's Corporation (1983).

rents by first grouping outlets into three different investment classes: low, medium and high, and then assigning each investment-class a level of sales, with small establishments being assigned small sales levels, medium establishments being assigned medium sales levels, and large establishments being assigned large sales levels.[4] Data on sales was obtained from a different source. They state:

> According to the disclosure documents, of more than 5,400 McDonald's restaurants (franchised and company owned) opened 13 months or more in the United States as of December 1982, 76 percent had sales over $900K, 49 percent had sales above $1,100K, and 24 percent had sales greater than $1,300K.(Kaufmann and Lafontaine 1994, p. 422)[5]

The rents calculated by K-L in this manner are presented in the bottom two lines of Table 6.1, using alternative discount rates of 5% and 7%.[6]

K-L then make the following additional assumptions:

a) The distribution of sales levels across stores in a given year is the same for company-owned stores and for franchised stores,[7]

---

[4]They implicitly make this assumption when they state:

> Presumably the variation in all these figures is a function of the size of the restaurant. Hence the total investment required for a smaller restaurant would be around $282K, while for a larger outlet it would be about $410K. (Kaufmann and Lafontaine 1994, p. 423)

[5]They go on to note (*fn.* 13, p. 422) that the lowest and highest sales levels were $306K and $3,223K, respectively, with an average sales level of $1,123K.

[6]This table presents the data contained in Table 2 (p. 426) of their paper. The method by which sales were assigned seems to beg the question: Why would small outlets ever be constructed if rents can be earned on larger ones? Further, if sales and investment levels are not perfectly correlated (as they implicitly assume), the variance of the distribution of rents will be substantially increased, resulting in greater risk to the franchisee. Such risk will require compensation, thus providing yet another explanation of their data.

[7]This assumption allows K-L to argue from data collected on company-owned stores to conclusions

Table 6.1: Present Values for a Single McDonald's Restaurant

	Real Rate of 5 Percent			Real Rate of 7.5 Percent		
	Low	Mid	High	Low	Mid	High
Yearly sales	900	1,100	1,300	900	1,100	1,300
Yearly (ex post) rents	1.9	64.3	112.2	(8.4)	(53.0)	(97.2)
Present value of rents	24.9	841.4	1,468.4	(109.9)	693.5	1,271.9
Ex ante costs:						
Equipment	329	362	493	329	362	493
Franchise Fee	12.5	12.5	12.5	12.5	12.5	12.5
Training	50	50	50	50	50	50
Total	391.5	424.5	555.5	391.5	424.5	555.5
Ex ante rents:						
Before taxes	(366.6)	416.9	912.9	(501.4)	269.0	716.4
After taxes	(366.6)	291.8	639.0	(501.4)	188.3	501.5

(Thousands of 1982 Dollars)

b) This distribution is normal with a mean of $1,100K and a standard deviation of $300K.[8]

c) Sales for an individual store are flat over the 20-year life of the agreement,

Applying these assumptions to Table 6.1, K-L concluded "that despite our very conservative estimates of rents, about 65 percent of all new McDonald's restaurants earn rents ex ante" and "more than 75 percent earn rents ex post."(Kaufmann and

---

about rents for franchisees. In support of this assumption KL state that:

> [Plotkin (1991) presents] data describing the average sales levels of all the 178 McDonald's restaurants operating in three television markets ... between 1983 and 1985. ... In these data the means of average yearly sales for the 148 franchised and 30 company-owned restaurants were $1,430.4K and $1,447.2K respectively.(Kaufmann and Lafontaine 1994, footnote 13, p. 422)

I presume Plotkin presents data which shows that the variance of the distributions of annual sales for franchised and company-owned stores is also similar over these years, although such information is not offered by K-L.

[8]A normal distribution with these parameters would have approximately 25% of its mass below 900, which is quite consistent with their data.

Lafontaine 1994, p. 427 and p. 424, respectively)

K-L find support for assumptions *(a)* and *(b)* from an auxiliary data set and from work done by Plotkin (1991). However, they *offer no evidence at all* to support assumption *(c)*—that sales are flat over the 20-year life of the agreement. Therefore, for the purpose of brevity, this chapter will accept their data as given and grant them assumptions *(a)* and *(b)*, limiting the discussion simply to the conclusions they derive from assumption *(c)*.

6.3 An Alternative Conjecture and Its Implication

Without assumption *(c)* the conclusion that a large portion, or any portion, of McDonald's franchisees earn rents both ex ante and ex post, simply does not follow from their data and the other assumptions. To see this, let me offer the following alternative hypothesis:

Conjecture 6.1 *The time path of sales for a new McDonald's restaurant outlet is not flat, but rather follows a life-cycle consistent with that predicted by the Theory of Spatial Branding: starting off low when the outlet is young and consumers are inexperienced, and gradually increasing towards an upper bound as the outlet matures and consumers come to learn the value of the McDonald's product.*

An example of the conjectured time-path is presented in Figure 6.1. Under their assumptions, the time path for ex-post *rents* would also take this form, being proportional to sales, though shifted down by the amount of fixed operating expenses.[9] It should be clear that such a time path is *not* inconsistent with ex-post rents for 75% of McDonald's outlets in 1982, but it simply does not imply that even one such outlet

[9]Under their assumptions, operating costs increase proportionally with sales, so sales and ex-post rents move together—see their Table 1, p. 421.

will earn rents ex ante, contrary to the claim made by K-L that 65% of outlets earn such rents.

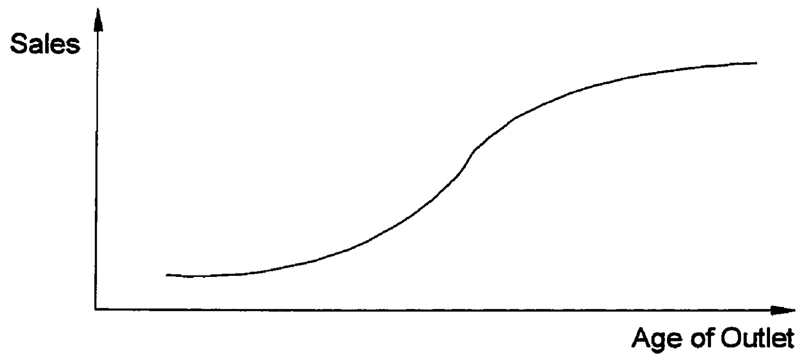


Figure 6.1: A Life-cycle Time Path for Sales

To see this note that in 1982 there existed a distribution of McDonald's establishments of various ages. For simplicity assume, á la K-L, this distribution is over the following discrete support: young, middle age, and mature. In Figure 6.2 the time-paths for outlets of different ages are superimposed to generate a cross-sectional distribution of sales in 1982. Note that the present value of sales and ex post rents for each of these classes of outlets will be identical. Further, ex ante rents may be positive negative or zero depending upon the size of initial up-front costs.

Clearly, the data presented by K-L is simply a snapshot of firms at 1982, providing just one point on the time-paths of sales and ex-post rents. As such, the data simply puts one constraint on the entire distribution of time-paths for this sample of outlets. No doubt the time-path of sales varies somewhat from outlet to outlet, depending not only on age and the size of the initial investment, but a variety of other factors as well.[10] Nonetheless, *one* point on a curve does not tell us the entire shape of the curve.

[10] As mentioned earlier, if all that is necessary to obtain large ex-ante rents is a large establishment, then why would small establishments be constructed? However, if rents depend on factors other than

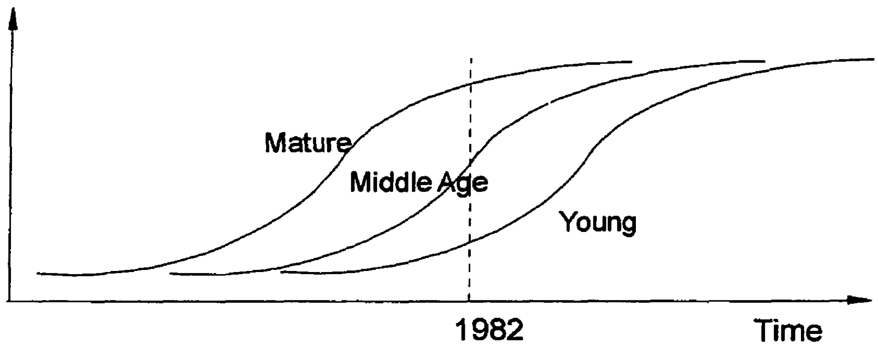


Figure 6.2: Distribution of Sales in 1982 for Outlets of Various Ages

In contrast to ex post rents, ex ante rents is necessarily an intertemporal concept, requiring capitalization of the flow of ex post rents over the life of the franchised operation. As such, evidence for the existence of ex ante rents requires not just a demonstration of ex post rents for the 1982 operating year, but knowledge of the entire time path of rents, and on this matter no evidence is provided at all.

We are thus left with the following conclusion:

Proposition 6.1 *If the flat time path for sales (assumed by K-L) is replaced with the life-cycle time-path (predicted by the Theory of Spatial Branding), the 1982 operating-data presented by K-L can be explained by the ages of establishments at that time. Thus, while the conclusion that* 75% *of McDonald's franchisees earn rents ex post in 1982 is not refuted, the conclusion that* 65% *of McDonald's franchisee's earn rents ex ante simply does not follow.*

---

the size of investment, than franchisees bear substantial risk, which may account for any ex-ante rents—if such rents could be shown to exist.

K-L find that 75% of McDonald's franchisees earn rents ex post, implying that 25% earn no such rent. If Proposition 6.1 is correct, then 25% of McDonald's establishments in 1982 should be young. Indeed, a simple calculation using other data presented by K-L (Table 3, p. 428—also presented in Table 6.2 of this chapter) shows that approximately 28% of the franchises which existed worldwide at the end of 1982 were less than 4 years old, thus providing some preliminary support for the Proposition.[11] This argument, however, is subject to the same criticism just leveled against that presented by K-L. Just because the distribution of ages for McDonald's establishments in 1982 is consistent with the life-cycle time-path of sales, does not mean that this is the actual time path—one point on the curve does not imply anything about the shape of the curve. Thus, although the life-cycle time-path for sales is a possible explanation for their data, it has yet to be shown that there is any more support for this conjecture than that of a flat time-path assumed by K-L. However, further evidence supporting the life-cycle time-path is available, and this evidence again comes from K-L's own data.

Evidence in Support of the Life-cycle Conjecture

The debate over alternative time paths would obviously be resolved if a large random-sample of time-series data from McDonald's franchisees could be obtained. Presumably K-L have attempted this, hence the need for their assumptions. Fortunately, a large sample of outlet-level data is not necessary to settle this issue since the two hypothesized time paths have different implications for the firm-level data. In particular, if outlet-level sales follow the life-cycle time path, then the average sales level across all outlets should increase with the average age of the outlets; whereas if a flat time path for outlet sales is correct, the average sales level should bear no relationship

[11] Similar calculations show that approximately 48% were less than 7 year old and approximately 69% were less than 10 years old.

with the average age of outlets.

Fortunately, the firm-level data necessary to test these implications is not nearly so difficult to obtain—it is contained within Table 3 (p. 428) of the K-L article. This data along with some new calculations derived therefrom are presented in Table 6.2.

Table 6.2: Time Series for McDonald's Outlets Worldwide

Year	Number of Outlets	Average Age of Outlets	Average Sales per Outlet	% Change in CPI	Real Average Sales per Outlet
1964	688	3.00	188,372	–	188,372
1965	760	3.68	224,868	1.6	221,327
1966	862	4.29	253,480	2.9	242,457
1967	967	4.82	275,491	3.1	255,588
1968	1,125	5.14	297,778	4.2	265,129
1969	1,345	5.30	335,167	5.5	282,861
1970	1,592	5.48	368,719	5.7	294,396
1971	1,944	5.49	403,344	4.4	308,469
1972	2,272	5.70	514,877	3.2	381,558
1973	2,784	5.65	563,218	6.2	393,015
1974	3,037	6.18	677,642	1.1	467,715
1975	3,756	6.00	696,619	9.1	440,709
1976	4,225	6.33	724,970	5.8	433,502
1977	4,736	6.65	789,274	6.5	443,148
1978	5,185	7.07	882,353	7.6	460,417
1979	5,747	7.38	936,837	11.3	439,216
1980	6,262	7.77	994,251	13.5	410,690
1981	6,739	8.22	1,057,872	10.3	396,164
1982	7,259	8.63	1,075,768	6.2	379,347
1983	7,778	9.06	1,116,804	3.2	381,606
1984	8,304	9.48	1,205,034	4.3	394,778
1985	8,901	9.85	1,235,917	3.6	390,826
1986	9,410	10.31	1,317,747	1.9	408,933
1987	9,911	10.79	1,445,909	3.6	433,113
1988	10,513	11.18	1,528,013	4.1	439,680

Data on number of outlets, average sales per outlet, and percentage change in the CPI are as contained in Table 3 of K-L. Average age of outlets and average real sales per outlet were calculated from their data as follows.

The average age of outlets was calculated on the basis of the following assumptions:

1. In 1964 the average McDonald's outlet has been in business for 3 years. This assumption seems reasonable since by 1964 McDonald's had only been franchising seriously for about 8 years, with the vast majority of the 688 outlets existing in 1964 opening after 1960.[12]

2. The number of discontinued outlets in any year is zero. This assumption is made for simplicity, but it is not entirely unrealistic. Franchisors often advertise that few, if any, outlets are shut down. In fact, K-L report (p. 434, *fn.* 43) an annual discontinuance rate of about 0.3% for McDonald's franchisees. Under this assumption the growth in the number of outlets is equal to the growth in new outlets.

From these assumptions it follows that the average age of outlets at time t is then given by:[13]

$$age_t = \frac{n_{t-1}}{n_t} age_{t-1} + 1$$

where n_t is the number of outlets at time t and $age_{64} = 3$. After adjusting for inflation, the time path for real sales (i.e. in 1964 dollars) presented in Figure 6.3 is obtained.[14]

[12]Shook and Shook (1993, p. 144) report that at the end of 1962 there were over 400 outlets, implying that approximately 42% of the 688 outlets existing at the end of 1964 were less than 2 years old.

[13]This follows from the following equation:

$$age_t = \frac{n_{t-1}}{n_t} (age_{t-1} + 1) + \frac{n_t - n_{t-1}}{n_t}(1)$$

The data only permit the calculation of differences for $t \geq 1965$.

[14]Real sales is calculated from nominal sales for period t as follows:

$$rsales_t = \frac{nsales_t}{\prod_{s=1964}^{t} (1 + cpi_s)}$$

Clearly this time path is not flat, but rather consistent with that predicted by the Theory of Spatial Branding: starting off low when the outlet is young and consumers are inexperienced, and gradually increasing towards an upper bound as the outlet matures and consumers come to learn the value of the McDonald's product.

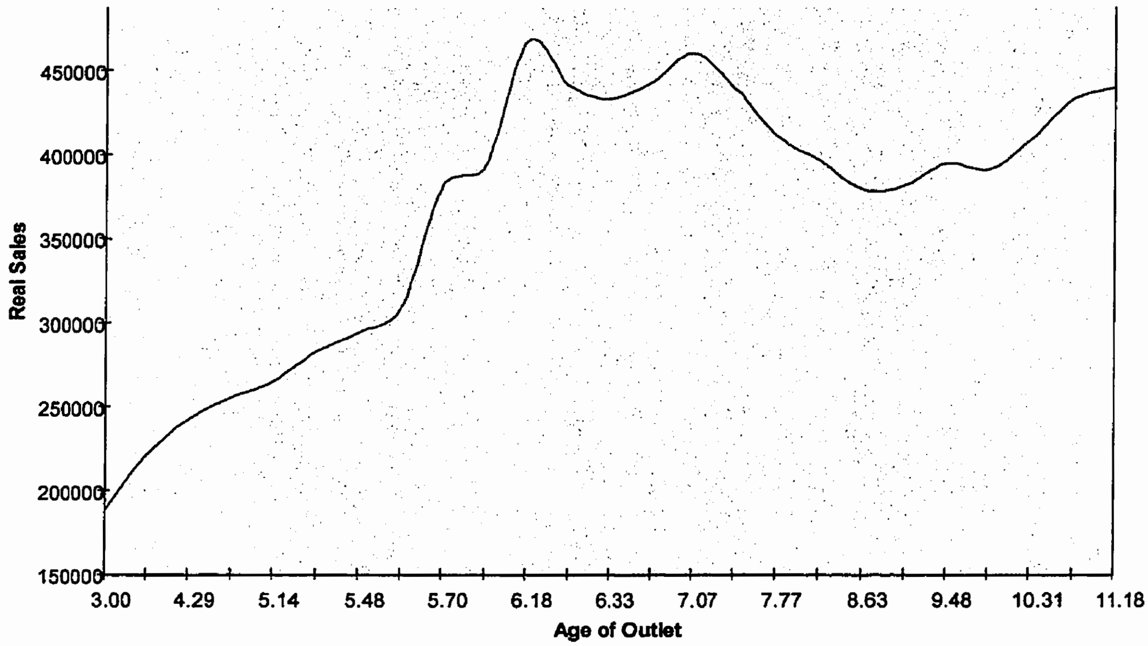


Figure 6.3: Time Path of Real Sales for an Average McDonald's Establishment

Given the time-path derived from K-L's data, we can conclude this chapter with the following two propositions:

Proposition 6.2 *Sales are positively related to the age of the outlet, so the assumption of a flat time path is contradicted by K-L's own data. Further, as previously*

---

where cpi_s is the percentage increase in the consumer price index (i.e. a measure of inflation) in year s.

demonstrated, without this key assumption, the conclusion that McDonald's franchisees earn rents ex ante is not implied.

Proposition 6.3 *The time path derived from K-L's data is consistent with that predicted by the Theory of Spatial Branding.*

Chapter 7

Conclusions and Implications for Franchising

7.1 Towards a Theory of Franchising: An Application of the Theory of Spatial Branding

This chapter attempts to summarize some of the fundamental results derived from of the theory of spatial branding by exploring their implications for the theory of franchising, with particular emphasis being placed on the life-cycle hypothesis. A franchise contract is essentially a fixed-term licencing agreement that entitles a franchisee to use the business concept and other intellectual capital of the entrepreneur/franchisor to distribute products and/or services in a manner specified in the contract. In exchange for the right to use the franchisor's intellectual capital, the franchisee typically makes both capital and current payments to the franchisor.[1] The capital payment,

[1]Other fees are also sometimes imposed. Some franchisor's will require franchisees to purchase inputs at a mark-up, though this practice runs up against tying-regulations and continues to be used in only a few industries. However, many franchisors charge for the performance of specific services, such as advertising, bookkeeping, management consultation, employee training, location selection

a franchise fee, varies substantially across industries and firms, but is typically non-refundable and paid at the time of signing. The current payments, periodic royalties, also vary somewhat across industries and firms, but are typically levied on gross revenues and usually less than 10 percent.[2]

7.1.1 The Life-cycle Hypothesis

What is the role of franchising within retail chains? The life-cycle hypothesis, inspired by discussions in trade journals and surveys of franchisors, was first introduced in the Journal of Retailing by Oxenfeldt and Kelly (1969). The hypothesis states that franchising is used by young retail chains to provide the financial and managerial capital necessary "to penetrate the market as widely and rapidly as possible, thus preempting valuable territory from competitors. Once the desired initial coverage is attained, their emphasis usually shifts toward operating efficiencies and market development—both of which can best be attained through the tight control permitted by ownership." (Oxenfeldt and Kelly 1969, p. 74) Their hypothesis leads them to make the following bold prediction:

> We will contend in this article that most *successful* franchise systems will end up as almost wholly-owned chains; we will argue that franchising is advantageous to a successful franchisor mainly during the infancy and adolescence of the enterprise and even thereafter for the exploitation of marginal locations. (Oxenfeldt and Kelly 1969, p. 69)

---

and provision, etc.

[2]The royalty-tax on revenues is often viewed by the franchise literature as presenting a bit of a puzzel, but is easily explained by the theory of spatial branding. In particular, this theory has shown that the revenues earned by a spatially branded store serving a mobile population of consumers are, for the most part, determined by the chain and not by the performance of the store being taxed. As such, these revenues constitute a rent to the store, which is unlikely to adversely affect the incentives of the franchisee if taxed.

However, despite much empirical support, the hypothesis seems to have been largely dismissed by academics.[3] There appear to be two reasons for this dismissal.

First, the hypothesis simply presumes quick expansion and market preemption are goals of retail chains. Oxenfeldt and Kelly, and later proponents of the hypothesis, provide no formal support or theoretical rationale for these goals. Almost any productive activity is more costly if done quickly, and expansion of a retail chain is no different. Unless the hypothesis can explain the source of additional benefits generated by quick expansion, it is little more than an interesting and insightful conjecture. Likewise, market preemption is founded in imperfectly defined property rights over a market. To preempt a market is to capture it by entering prior to the competition. Not all markets can be preempted: rights to the market must be founded in a resource which is being allocated on a first-come first-serve basis. Moreover, entering a market prior to the competition often means entering prior to the time which would be optimal in the absence of such competition, and this is costly. The competition

---

[3] Using data from a previous study done by Ozanne and Hunt (1971), Hunt (1973) found that in the fast-food industry "the percentage of total units that was company-operated increased steadily from a low of only 1.2 percent in 1960 to 9.5 percent in 1970 ..." (p. 8) Further, that the increase in company ownership at that time was due to larger franchise systems and older franchise systems "disproportionately increasing their percentage of company-operated units." (p. 9) Similarly, Caves and Murphy II (1976) found that from "1968-1971 large fast-food franchisors increased their percentages of owned establishments significantly faster than small ones, consistent with the hypothesis of capital rationing." (p. 583) They concluded:

> The most obvious mechanism to explain this cycle [of an increasing proportion of company-owned outlets] rests on the availability of capital to the franchisor. He may be quite constrained by lender's risk in the early days before his intangible asset [brandname] is well established. For financing outlets the capital supplied by franchisees has no ready substitute. If and when his operation becomes successful, however, the franchisor no longer needs the franchisee as a source of funds.(p. 580)

In a survey of franchisors, Ronald L. Tatham and Bush (1972) found that when selecting franchisees the number one consideration is financial capital. McDonald's franchisees, for example, must contribute at least 40 percent of the initial capital and obtain financing for the rest.(Kaufmann and Lafontaine 1994, footnote to Table 1, p. 421) Additional empirical support for franchising as a method of quick expansion and market preemption is given by Lafontaine (1992b), Dant (1994), Lillis and Gilman (1976), and Gilman (1990).

to be first often dissipates the rents which will be earned by the winner (Eaton and Lipsey 1979). Unless the hypothesis can explain how markets typically served by retail chains are "captured" by early entry (i.e. the source of the rents), the hypothesis does not constitute a complete explanation.

Second, even if theoretical support for the goals of quick expansion and market preemption can be provided, why would retail chains choose franchising as a method of financing the attainment of such goals? Here too the hypothesis is silent, it simply presumes franchising is the best method of acquiring the necessary capital. In fact, one of the primary reasons why the life-cycle hypothesis seems to have been dismissed is an oft-cited critique made by Rubin (1978) that franchising as a method of financing runs counter to modern portfolio theory. In particular, he argues that the theory of diversification implies that "A risk averse franchisee would clearly prefer to invest in a portfolio of shares in all franchise outlets, rather than confining his investment to a single store. This means, essentially, that the franchisee will require a higher rate of return on his capital if he is required to invest in one outlet rather than in a portfolio." (Rubin 1978, p. 225) Rubin concludes that franchising could not possibly be used as a method of financing the growth of a retail chain.

Rubin's critique has led many researchers to dismiss the view proposed by the life-cycle hypothesis that franchising is simply a short-run solution to the immediate problem of chain growth faced by an entrepreneur starting a new chain. Many researchers have instead chosen to view franchising as an organizational form which is observed because it is efficient: providing the proper incentives for downstream effort by giving the operator of the store a claim to the residual profits of the operation. While space precludes a complete analysis of both sides of these arguments, this chapter will attempt to demonstrate that the theory of spatial branding does provide theoretical foundation for the life-cycle hypothesis and, therefore, that the hypothesis should not be dismissed so quickly. Further, that franchising may be observed, not because it is long-run efficient, but rather because it solves the immediate needs of an entrepreneur for quick expansion and market preemption. After these goals have been achieved, the entrepreneur switches to company-run stores which permit tighter control over the operations.[4]

7.1.2 Rebuttals to Rubin's Critique

Rubin's critique is that franchisees will demand additional return for bearing additional risk, making franchising a more costly method of financing than the alternative of share ownership. Two rebuttals to this critique will be presented here.[5]

First, this argument overlooks the fact that franchising may offer nonpecuniary returns not present in share ownership of the chain. These nonpecuniary returns may leave the pecuniary cost of financing unchanged or even lower than that which would be demanded through share ownership. One such nonpecuniary return enjoyed by many franchisees is business experience. For example, Diaz and Gurnick (1969, p. 12) argue that franchisees are purchasing a door to the business world:

> It is apparent that they were attracted to franchised businesses because the "franchise package" (what they were getting for their investment) was recognized as superior to the "package" that the franchisee as an individual could create through his singular efforts. Perhaps the most difficult business ingredients for an individual business to develop are the very

[4]Dow (1987) and Dow (1993) have made similar arguments with respect to labor-managed vs. capital-managed firms. In particular, even though labor-managed firms may be more efficient, capital managed firms may be observed since they best protect the entrepreneur against approporiation of rents.

[5]Lafontaine (1992a) offers a third rebuttal to Rubin's critique. In particular, she notes that if, according to the Incentive Hypothesis of franchising, company managers have less incentive to exert effort than franchisees, the expected yield on shares of a portfolio of all outlets is likely to be lower than if the system were franchised. Further, this reduction in yield may more than offset any gains in risk reduction, making the net cost of financing lower under franchising than under share ownership.

> components of the "franchise package" which often includes all or most of the following: established name and reputation, widely advertised brands, popular store design, carefully chosen location, standardized procedures and operations, and initial and continuing assistance (e.g. franchisor training, financing, and research).[6]

No doubt many individuals become franchisees to acquire business experience, but Rubin's critique can be rebutted on much more fundamental grounds. In particular, this critique completely ignores the role of financing within the context of the life-cycle hypothesis. The hypothesis argues that franchising as a method of financing is *required to develop the chain.* To argue that financing cannot explain franchising because *shares to an already developed chain* are less risky than ownership of a single store in that chain is simply a non sequitur. His argument presupposes the existence of the chain which the financing is required to develop. For the law of large numbers to apply we need both large numbers and independent risks, neither of which holds true for an immature system with one or just a few stores. In failing to appreciate the role which the financing argument plays within the context of the life-cycle hypothesis, Rubin's argument puts the system-cart before the financing-horse.[7]

---

[6]Even Rubin apparently acknowledges this as a reason to purchase a franchise. He states:

> "... to the extent that franchisees are closer to being employees than entrepreneurs, they may simply lack the requisite human capital to open businesses without the substantial assistance of franchisors." (Rubin 1978, p. 230)

[7]This type of fallacious reasoning is responsible for many other misunderstandings about retail chains. For example, the argument that franchisees have a lower failure failure rate than independent operators is usually supported by anecedotal evidence from franchisees who are members of large well-established chains. Similarly, the argument that consumers choose to purchase from brands with big chains because they are well known, presumes the consumer is familar with the brand. Such an argument does not explain how consumers came to be familar with the big chain, or how the big chain got to be big.

7.1.3 The Threat of Internal Appropriation

Rebutting Rubin's critique simply means that franchising as a method of financing cannot be ruled out on portfolio-theoretic grounds. However, while franchising *could* be used as a method of financing, the life-cycle hypothesis provides no theoretical rationale why entrepreneurs attempting to grow their retail operations would come to select franchising. Yet there are good reasons why franchising may come to be preferred to the alternatives.

First, the goals of quick expansion and market preemption rule out internal financing through retained earnings; that method of growth is simply too slow. Second, it may be difficult for an entrepreneur to obtain large amounts of debt financing. As Caves and Murphy note," He may be quite constrained by lender's risk in the early days before his intangible asset [brandname] is well established. For financing outlets the capital supplied by franchisees has no ready substitute." (Caves and Murphy II 1976, p. 580) This is especially true since much of the investment made by retail chains is highly brand-specific (or sunk) and offers little or no value to creditors in case of default. As Hadfield notes:

> For the franchisee, the most significant economic feature of franchising is the allocation of capital investments. Franchisees are distinct from ordinary employees because they have made capital investments in the business. These investments, however, are normally highly idiosyncratic ... Consequently, the costs of establishing a franchise are effectively sunk costs, which, once expended, are not easily recovered if the franchisee goes out of business. (Hadfield 1996, p. 951)

Yet venture capitalists provide many credit-rationed businesses with the equity capital necessary to finance their sunk investments, why should retail chains choose franchising?

First, as a financing instrument the franchise contract is a hybrid, having attributes which lie between standard equity and debt contracts. Like the equity supplied by a venture capitalist, the equity supplied by a franchisee does not come with fixed-date interest and maturity obligations typical of debt contracts. These obligations create liquidity problems for a young chain which may push it into bankruptcy.

Second, and much more importantly, like a debt contract but unlike venture-capital financing, franchising does not dilute ownership and control over the intellectual capital which is the heart of the retail chain. While franchising does relinquish ownership of the individual outlet to the franchisee, ownership and control over the business format and the intellectual capital of chain remains in the hands of the entrepreneur. Such control is maintained since, unlike both debt and venture-capital contracts, the primary restrictions within a franchise contract are imposed on the behavior of the individual suppling the capital—i.e. the franchisee.

Third, the divisible nature of retail chains (i.e. into standardized stores) provides the entrepreneur with a unique opportunity to obtain capital through a method which is simply unavailable to many other businesses. Retail chains are not distinguished by their need for large-scale financing, but they are distinguished by the divisible nature of the capital such financing is used to finance. Many an entrepreneur has envisioned a product or service which can only be produced efficiently at large scale. Economies of scale often have their source in large and indivisible production capital and/or processes. However, the production capital of a retail chain is divisible and standardized. Because it is divisible, the capital itself can be divided and owned by separate individuals, much like share ownership in a piece of indivisible capital. Further, because the complementarity resides in the sharing of the brandname across stores and not in the stores themselves, dividing ownership of the capital does not expose the franchisor to hold-up by any one franchisee. Finally, unlike other teams composed of idiosyncratic and indispensable members, each member of the chain is highly standardized and dispensable.

Economies of scale are limited by the extent of the market. The extent of the market for a retail chain is determined by the extent to which the chain can get inexperienced and mobile consumers to try the product. As the theory of spatial branding has shown, it is the sharing of a common brandname across spatially distributed stores which creates the market by inducing mobile consumers to try the brand for the first time. Until the market is created, no sales can be made, and all other scale economies commonly attributed to the success of retail chains cannot be had. The increasing returns generated by brandname sharing are associated with the locational choice of retail chains and the mobility of the consumers they have evolved to serve. The divisible nature of the operation provides the entrepreneur with a unique opportunity to have her cake and eat it too: through franchising, she can obtain the quick financing necessary to achieve the goals of quick expansion and market preemption without risking internal appropriation that often comes with venture capital.

7.1.4 The Theory of Spatial Branding: Theoretical Foundations for the Life-cycle Hypothesis

The previous section has attempted to show that, as conjectured by the life-cycle hypothesis, franchising may in fact be used as a method of financing quick expansion and market preemption; that this conjecture cannot be ruled out on purely theoretical grounds; and that a number of researchers have found empirical evidence to support it. Yet until theoretical foundations for the goals of quick expansion and market preemption are supplied, the life-cycle hypothesis is really just that: a conjecture. The present section attempts to provide the required theoretical foundations for the life-cycle hypothesis by building upon the theory of spatial branding.

The theory of spatial branding has shown that the comparative advantage of retail chains lies in serving convenience goods to a mobile population. When consumers travel for reasons exogenous to the demand for the good, they will take their demand for the good with them and find the information conveyed by spatial branding useful in their spatially-distributed purchase decisions. So as to exploit their comparative advantage in serving mobile populations, retail chains locate their stores along major roads and highways, in bus stations and airports, and in many other locations where such populations can be found.[8]

Spatial branding has little value unless stores are located in areas of high consumer-mobility. Since such locations are scarce, retail chains will want to be first into newly developing areas so as to capture these prime locations. The competition for such locations is severe since it includes all retail chains, not just those brands within a specific product group. As the best locations are taken, subsequent entrants will have to settle for less visible locations, further off the primary travel routes of consumers, with poorer access and egress making the cost of secondary travel all that much greater. The inconvenience of such locations means consumers are less likely to try brands occupying these locations, not only because the current-cost of acquiring the brand is higher, but also because the value of information is lower (Proposition 4.2). There is little value to experimenting with brands poorly located, for such information is unlikely to find future use.

Locations of high mobility are not the only resource that can be captured by early

[8]As Luxenburg notes,

> "Franchise companies tell prospective franchisees that site selection is a difficult art that no novice can hope to master. The chains trot out pages of computer printouts showing convincing research. Site selection is crucial for the success of retail outlets, which is why many chains are conservative and choose obvious locations clustered near competing units. There are two basic rules for site selection: (1) the best locations are near the interchanges of major highways; (2) long secondary highways connecting downtowns with interstates and neighboring towns are likely sites for franchise rows." (Luxenburg 1985, p. 190)

entry. Being first in the market may also allow a high-quality brand to preempt the beliefs of consumers (Proposition 3.3). Even if complete preemption of beliefs is not possible, being first into a market can still have substantial returns. Being first means consumers have a low opportunity cost of trying the product; they are, therefore, more willing to experiment, and this improves the revenue-side of things. Getting consumers to try the brand sooner means more cash flow in the early years, and quicker convergence of the local-market share to its long-run value (Proposition 5.2). On the cost-side of things, economies of scale are limited by the extent of the market, so unless inexperienced consumers can be induced to try the brand, traditional production economies will not be realized. Finally, being first will make subsequent entry by competing brands more difficult. When consumers can buy a high-quality incumbent brand, experimenting with a new brand comes at a cost—even in the absence of risk aversion. Consumers will be reluctant to try a new brand unless such a brand enters the market with a chain of a size sufficient to make the value of experimentation commensurate with its cost (Proposition 3.2).

The theory of spatial branding makes it clear that entry into a new market is a dynamic process. It takes time for consumers to learn the value of the brand, and the time taken will be directly related to the chain's coverage of primary travel routes used by consumers. Once consumers have come to learn the value of the brand, subsequent stores established in the area will enter the market at or about their long-run sales levels, and will not have to go through this period of consumer learning. At this point, franchisors may switch from franchising to company-owned operations as conjectured by the life-cycle hypothesis.

7.2 Explaining Franchise Failures

It is common belief that aspiring entrepreneurs who dream of owning their own business can improve their chances of survival by becoming a franchisee. Popular media frequently point to large franchised chains as evidence of the success of business-format franchising and chain-distribution generally. While there can be little doubt that these big chains are successful, they by no means constitute systematic evidence for the success of chain distribution. To point to Tiger Woods as evidence of the profitability of golf as a profession, or Wayne Gretzky as evidence of the profitability of hockey as a profession is to simply ignore all those who have failed attempting to do what these successful individuals have accomplished.

A systematic study of the factors which affect the survival rates of small businesses requires detailed firm-level data. In the past, such data was difficult to obtain. However, Bates (1995) has recently published a study which uses the Characteristics of Business Owners Database from the U.S. Bureau of Census. From this database, Bates was able to obtain detailed information on some 20,554 small firms started between 1984 and 1987. He tracked their survival to late 1991. Surprisingly, he found that of "retailing firms that were operating in 1987, 45.1 percent of the young franchises had gone out of business by 1991, versus 23.4 percent of the independent young retail firms." (Bates 1995, p. 26)[9]

Due to the detailed nature of his data, Bates was able to use regression analysis to

---

[9]While the work of Bates is probably the most systematic work done to date, others have found confirming ancedotal evidence. For example, Luxenburg (1985, p258-59) reports that in May 1981 there was a U.S. Congressional investigation into defaults by small businesses on loan guarantees made to them by the Small Business Administration. He states:

> [T]he General Accounting Office determined that in the SBA program 10 percent of franchises defaulted, while only 4 percent of the independent businesses receiving similar assistance were failing. And some franchises were defaulting at rates considerably higher than the average. (Luxenburg 1985, p. 258)

control for the effect of other factors likely to contribute to firm failure—e.g. owner and firm traits such as education, captialization, etc. The result is especially interesting since, in effect, his finding is that a store which is a member of a chain (i.e. a franchisee) is more likely to fail than one which is not. Since he is only looking at independently owned operations, he has eliminated company-owned operations from the data set, thereby controlling for lack of residual-claimancy incentives as a possible explanatory factor. Thus, in the end, his finding is that some factor, or set of factors, associated with the affiliation with a chain is primarily responsible for failure. Further, while he finds the relationship between franchising and failure to be statistically significant across the entire database, the relationship is found to be strongest in the retail sector. He states: "[T]he franchise characteristic is the strongest predictor of firm discontinuance: the retail franchise firms are *much more* likely to go out of business, other factors constant, than independent young retail businesses."(Bates 1995, his italics, p. 32) He goes on to state: "[R]esearch on franchise behavior is properly viewed as being at an early stage. Future research is likely to extend our understanding of the underlying causes of the lower survival rates that typify the young franchise small businesses." (Bates 1995, p. 33)

By itself, the theory of spatial branding offers a possible explanation for the higher failure rate. If a brand enters a new geographic area with just a few stores and fails to expand quickly, it will take much longer for consumers to learn the value of the brand and longer for each store's local-market share to converge to its long-run value. Over this period of consumer-learning, sales will be low and the likelihood of failure greater. Independent single-store operations, in contrast, find it difficult to compete with spatially-branded chains for the custom of a mobile population travelling along major thoroughfares. These independent operations tend to locate off the beaten trail, serving the needs of their local community. Such locations give rise to repeat sales, so the time-path of sales converges rapidly to its long-run value. Hence independent operations are typically not exposed to a prolonged period of consumer learning and lower sales which can ultimately result in failure.

While the theory of spatial branding can explain the higher failure rate on its own, better explanations can be had by combining this theory with other awkward facts of the market which retail chains serve. In this section two alternative but complementary explanations are briefly outlined.

The Threat of External Appropriation

While the very existence of the franchise contract presumes well-defined rights over the intellectual capital embodied in the contract, this presumption is often unjustified. The laws of intellectual property include the laws of patents, copyright, tradesecrets and trademarks. In the interest of promoting competition in the market place, the laws of intellectual property provide only minimal protection to good business concepts. Unfortunately, it is the business concept which is of primary value to a entrepreneur who has yet to grow her chain and establish her brandname. The vast majority of novel business concepts used by retail chains are not grounded in new and useful machinery or processes. As such, these concepts are typically not of a form which can be protected by the laws of patent. Similarly, while the laws of copyright can protect the expression of a business concept in the operation manuals of the chain, such laws provide little or no protection for the concept itself. Further, good business concepts are often apparent to customers—that is what makes them good. As such, they typically cannot be protected by the laws of tradesecrets.

Often a new business concept for a retail chain is simply the quick and convenient provision of an old product or service dressed up in a new and interesting package. This "packaging" is called a chain's tradedress. A chain's tradedress is usually a significant part of the whole business concept or format, constituting the chain's "look and feel". Unfortunately, a chain's tradedress can only find protection under the laws of trademark if such dress is *(i)* nonfunctional and *(ii)* inherently distinctive or has acquired secondary meaning.[10] However, much of what constitutes a chain's tradedress is typically considered by the courts to be of a functional nature, being a component of the product or service being sold, or affecting its cost or quality. Functional tradedress can find no protection in law of trademark for fear of lessening competition in the market place.

Nonfunctional tradedress can find protection if the court considers such dress to be inherently distinctive. A dress is inherently distinctive if it is "fanciful or arbitrary", serving to uniquely identify the source of the product to consumers. Even if a tradedress is found to be nonfunctional and inherently distinctive as a whole, many of the individual components of the dress are free to be copied by potential and existing competitors. However, much of a chain's tradedress will typically be "suggestive or descriptive" of the goods or services being provided and, therefore, not serve to inherently distinguish the source of the goods. Much like the reasoning which underlies the inability to protect functional tradedress, the courts believe that to prevent competitors from using the same suggestive or descriptive tradedress would put them at a disadvantage and serve to lessen competition in the market place.

The law does, however, provide for one exception. Even suggestive or descriptive tradedress can find protection under the laws of trademark if it is nonfunctional and can be shown to have acquired a "secondary meaning" in the minds of consumers. This means that the tradedress has, through its use in commerce, come in fact to distinguish the source of the product or service.[11] Once secondary meaning for a tradedress can be shown, courts will usually extend protection, for to allow others to use the same dress would serve to confuse consumers and allow competitors to

[10] See, e.g., McCarthy (1984).

[11] The "primary meaning" of the tradedress is the meaning usually attached to the object. For example, golden arches have the primary meaning of "golden arches" and the secondary meaning of "McDonald's restaurant."

appropriate any goodwill embodied in such a dress. Nonetheless, protection under the requirement of secondary meaning is often restricted to the region of commerce—i.e. where the dress has acquired its source-identifying meaning.[12]

These are tough requirements. It means, for example, that novel theme-concepts, such as a turn-of-the-century motif for a restaurant, although franchisable, will not be protectable under the laws of intellectual property. In the interest of competition, the courts will typically let consumers choose which restaurant best fills this market niche. Many an entrepreneur has lost their good idea to one or more copy-cat chain. For example, Luxenburg describes the problems Wendy's Hamburger Restaurants faced when its business format was copied by competitors. He states:

> "After Wendy's opened, Judy's, Cindy's, and Andy's began offering variations on the large-patty theme. The Wendy's concept consisted of thicker hamburgers served in restaurants featuring turn-of-the-century design motifs. Cindy's sought to bring higher-quality fast food to small towns in the Southeast that had been overlooked by the national chains. Signs over Wendy's stores read "Wendy's Old-Fashioned Hamburgers" and displayed a picture of a girl with pigtails tied with ribbons. Cindy's sign said "Cindy's Ole Time Hamburgers" and featured a picture of a girl whose hair was tied with a ribbon."(Luxenburg 1985, p. 24)

In part, copy-cat chains are to be expected, for they are a by-product of the market which retail chains have evolved to serve. Retail chains serve a mobile population, and it is the very mobility of the population which allows good ideas to be spread

[12]There presently exists a debate over a controversial doctrine called "secondary meaning in the making". This doctrine would allow trade dress to be protected prior to establishing a secondary meaning for the dress so long as it can be shown that such a meaning is "in the making." The Theory of Spatial Branding may find useful application in resolving some issues surrounding the debate. Of particular relevance, Spatial Branding provides a theoretical link between the size of the chain and the speed of consumer learning.

between geographic areas. If people were not mobile, there would be little threat of an individual taking a good idea observed in one region, copying it, and trying to make it work in another.

The threat of external appropriation by copy-cat chains forces many chains to adopt a strategy of broad geographic expansion. Yet the theory of spatial branding has shown that large scale entry and concentrated expansion is required to be successful in a new area. This provides perhaps the best explanation for the higher failure rate observed by Bates. If franchisors are more concerned about broad geographical expansion than developing a particular region sufficiently, the few franchisees that establish stores in one area will experience a prolonged period of low sales due to lack of market penetration, while the franchisor attempts to establish additional franchise operations in an entirely unrelated area. Speed of convergence is, therefore, being traded off against breadth of expansion.

Bibliography

Akerlof, George A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, August 1970, *84* (3), 488–500.

Allen, Franklin, "Reputation and Product Quality," *Rand Journal of Economics*, 1984, *15* (3), 311–327.

Arrow, K. J., *The Limits of Organization*, New York: W. W. Norton Co., 1974.

Bagwell, Kyle, "Informational Product Differentiation as a Barrier to Entry," *International Journal of Industrial Organization*, 1990, *8*, 207–223.

Bates, Timothy, "Analysis of Survival Rates Among Franchise and Independent Small Business Startups," *Journal of Small Business Management*, April 1995, *33* (2), 26–36.

Bather, J. A., "Randomized Allocation of Treatments in Sequential Experiments," *Journal of the Royal Statistical Society*, 1981, *43* (3), 265–292.

Berry, D.A. and B. Fristedt, *Bandit Problems: Sequential Allocation of Experiments*, New York: Chapman and Hall, 1985.

Bradach, Jeffrey L., *Franchise Organizations*, Boston, Mass.: Havard Business School Press, 1998.

Caves, R. E. and W. F. Murphy II, "Franchising: Firms, Markets, and Intangible Assets," *Southern Economic Journal*, 1976, *42*, 572–586.

Dant, Rajiv P., "Motivations for Franchising: Rhetoric Versus Reality," *Working paper, Boston University*, 1994.

DeGroot, Morris H., *Optimal Statistical Decisions*, New York: McGraw-Hill, 1970. Ch. 14: Sequential Choice of Experiments.

Diaz, Robert M. and Stanley Gurnick, *Franchising*, New York: Hastings House, 1969.

Dicke, Thomas, *Franchising in America*, Chapel Hill, N.C.: University of North Carolina Press, 1992.

Dnes, Antony W., "The Economic Analysis of Franchise Contracts," *Journal of Institutional and Theoretical Economics*, 1996, *152*, 297–324.

Dow, Gregory, "The Appropriability Critique of Transaction Cost Economics," in Christos Pitelos, ed., *Transaction Costs, Markets and Hierarchies*, Basil Blackwell, 1993, pp. 101–132.

Dow, Gregory K., "The Function of Authority in Transaction Cost Economics," *Journal of Economic Behavior and Organization*, 1987, *8*, 13–38.

Eaton, B. Curtis and Richard G. Lipsey, "The Introduction of Space Into the Neo-Classical Model of Value Theory," in M.J. Artis and A. R. Nobay, eds., *Studies in Modern Economic Analysis*, Oxford: Basil Blackwell, 1977, pp. 59–96.

——— and ———, "The Theory of Market Pre-Emption: The Persistence of Excess Capacity and Monopoly in Growing Spatial Markets," *Economica*, 1979, *46* (182), 149–158.

—— **and** —— , "Product Differentiation," in Richard Schmalensee and Robert D. Willig, eds., *Handbook of Industrial Organization*, Vol. 1 of *Handbooks in Economics*, Amsterdam; Oxford and Tokyo: North-Holland, 1989, pp. 723–68.

Gilman, Alan L., "Competitive Growth with Franchising," *Chain Store Age Executive*, January 1990, p. 160.

Gittins, J. C., "Bandit Processes and Dynamic Allocation Indices," *Journal of the Royal Statistical Society*, 1979, *41* (2), 148–177.

Gosh, Avijit and Sarah L. McLafferty, *Location Strategies for Retail Service Firms*, Lexington, Mass.: Lexington Books, 1987.

Hadfield, Gillian K., "Problematic Relations: Franchising and the Law of Incomplete Contracts," *Stanford Law Review*, 1996, *42*, 927–991.

Hart, Albert Gailord, "Risk, Uncertainty, and the Unprofitability of Compounding Probabilities," in "Studies in Mathematical Economics and Econometrics," University of Chicago Press, 1942, pp. 110–118.

Hollander, Stanley C. and Glenn S. Omura, "Chain Store Developments and Their Political, Strategic, and Social Interdependencies," *Journal of Retailing*, 1989, *65* (3), 299–323.

Hotelling, H., "Stability in Competition," *Economic Journal*, 1929, *39*, 41–57.

Hunt, Shelby D., "The Trend Toward Company-Operated Units in Franchise Chains," *Journal of Retailing*, 1973, pp. 3–12.

Jones, Robert A., "Decreased Confidence as Increased Information: Bayesian Representation with Application to Option Value," January 1995.

——— **and Joseph M. Ostroy**, "Flexibility and Uncertainty," *Review of Economic Studies*, January 1984, pp. 13–32.

Kaldor, N., "Market Imperfections and Excess Capacity," *Economica*, 1935, *2*, 33–50.

Kaufmann, P. J. and F. Lafontaine, "Costs of Control: The Source of Economic Rents for McDonald's Franchisees," *Journal of Law and Economics*, 1994, *37*, 417–453.

Keller, Kevin Lane, *Strategic Brand Management: Building, Measuring, and Managing Brand Equity*, New Jersey: Prentice-Hall, 1998.

Kihlstrom, Richard E. and Michael H. Riordan, "Advertising as a Signal," *Journal of Political Economy*, 1984, *92* (3), 427–450.

Klein, Benjamin and Keith B. Leffler, "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 1981, *89* (1), 615–641.

Lafontaine, F., "Agency Theory and Franchising: Some Empirical Results," *Rand Journal of Economics*, 1992, *23*, 263–283.

——— , "How and Why Do Franchisors Do What They Do: A Survey Report," in P. J. Kaufmann, ed., *Franchising: Passport for Growth and World of Opportunity, Sixth Annual Proceedings of the Society of Franchising*, International Center for Franchise Studies University of Nebraska 1992.

Lancaster, K. J., *Variety, Equity, and Efficiency*, New York: Columbia University Press, 1979.

Luxenburg, Stan, *Roadside Empires: How the Chains Franchised America*, New York: Viking Press, 1985.

M., Narayana C. L. Lillis Charles and John L. Gilman, "Competitive Advantage Variation Over the Life Cycle of a Franchise," *Journal of Marketing*, 1976, *40*, 77–80.

Mathewson, G. F. and Ralph A. Winter, "The Economics of Franchise Contracts," *Journal of Law and Economics*, 1985, *28*, 503–526.

McCarthy, J. Thomas, *Trademarks and Unfair Competition*, 2nd ed., Rochester, N.Y.: Lawyers Co-operative Publishing Co., 1984.

Nelson, P., "Information and Consumer Behavior," *Journal of Political Economy*, 1970, *78*, 311–329.

Oxenfeldt, A. R. and A. O. Kelly, "Will Successful Franchise Systems Ultimately Become Wholly-Owned Chains?," *Journal of Retailing*, 1969, *44*, 69–87.

Oxenfeldt, Alfred R. and Donald N. Thompson, "Franchising in Perspective," *Journal of Retailing*, 1968, *3* (4), 3–15.

Ozanne, U. B. and S. D. Hunt, "The Economic Effect of Franchising," Technical Report, U. S. Government Printing Office, Washington, D. C 1971. A report prepared for the U.S. Senate, Select Committee on Small Business.

Plotkin, Irving H., "On the Nature of Intangible Assets in a McDonald's Franchise, Expert's Report, Cantebury V. Commissioner," Technical Report 38037-87, U.S. Tax Ct. 1991.

Radner, Roy and Joseph E. Stiglitz, "A Nonconcavity in the Value of Information," in M. Boyer and R. E. Kihlstrom, eds., *Bayesian Models in Economic Theory*, Elsevier Science Publishers B.V., 1984, pp. 33–52.

Rothschild, Michael, "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 1974, *9*, 185–202.

Rubin, Paul, "The Theory of the Firm and the Structure of the Franchise Contract," *Journal of Law and Economics*, 1978, *21*, 223–234.

Salvaneschi, Luigi, *Location, Location, Location: How to Select the Best Sight for Your Business*, Grants Pass: Oregon: The Oasis Press, 1996.

Schmalensee, R., "Product Differentiation Advantages of Pioneering Brands," *American Economic Review*, 1982, *72* (3), 349–365.

Shook, Carrie and Robert L. Shook, *Franchising: The Business Strategy That Changed the World*, Englewood Cliffs, New Jersey: Prentice Hall, 1993.

Smith, Adam, *The Wealth of Nations*, Chicago: University of Chicago Press, 1976. Originally published 1776.

Tatham, Robert Douglass Ronald L. and Ronald F. Bush, "An Analysis of Decision Criteria in Franchisor/Franchisee Selection Processes," *Journal of Retailing*, 1972, *48* (1), 16–24.

Thompson, Donald N., *Franchise Operatons and Antitrust*, D. C. Heath and Company, 1971.

Wilson, Robert, "Informational Economies of Scale," *Bell Journal of Economics*, 1975, *6*, 184–195.