

Essays on Pricing and Advertising

by

Eric T. Anderson

B.S. Electrical Engineering, Northwestern University (1988)
M.S. Engineering Economic Systems, Stanford University (1989)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy in Management
at the
Massachusetts Institute of Technology

May 1995

© 1995 Massachusetts Institute of Technology
All Rights Reserved

Signature of Author:

Certified by:

John D.C. Little
Institute Professor
Professor of Management Science
Thesis Advisor

Certified by:

Birger Wernerfelt
Professor of Management Science
Thesis Advisor

Accepted by:

Birger Wernerfelt
Chairmen, Department Committee

ARCHIVES
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 08 1995

Essays on Pricing and Advertising

by

Eric T. Anderson

Submitted to the Sloan School of Management, M.I.T.,
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Management

Abstract

The dissertation consists of three separate essays on pricing and advertising, each of which is described below.

Essay 1: Dynamic Competitive Promotion Policies

Sellers often prefer to promote similar products at different times rather than at the same time. For example, Coke and Pepsi tend to be promoted in alternating weeks in a supermarket. The question we ask is under what conditions do we tend to observe alternating promotions? What type of customer behavior can help explain this phenomenon? This paper demonstrates that as customers become more loyal, sellers are more likely to promote their products in alternating periods. We provide examples showing how the model applies to both competing retailers and competing manufacturers.

Essay 2: Why Advertise Both Regular and Sale Prices?

When an item is on sale, some retailers advertise both a regular price and a sale price while other retailers advertise just the sale price. This paper offers an explanation as to why some retailers may prefer to advertise in the regular/sale price format while others prefer the sale price only format. Our explanation relies on a retailer's incentive to signal information to deal prone consumers. These consumers want to buy at the lowest price but are not certain as to whether they should search at other retailers for a lower price or postpone a purchase to a future sale period. We show how including the regular price in an advertisement may signal information to deal prone consumers which influences their decision to search or wait for a lower price.

We highlight this intuition in two separate models. Our main result in both models is that the incentive to signal depends on the proportion of deal prone consumers in the market. We present empirical evidence in support of the main result. This paper contributes to the comparative literature on selling formats by evaluating the implications of bargain hunting (searching and waiting for lower prices) on the part of consumers. We also contribute to the literature on reference prices as we rationalize the use of a type of reference price by consumers.

Essay 3: Long-Term Effects of Trade Promotion:

Brand Advertising or Brand Erosion?

In the packaged goods industry, manufacturer's spend considerable amounts of money on both trade promotion and advertising. In recent years, increased spending on trade promotion relative to advertising has spurred industry debate. Trade promotion is often viewed as part of a short-term strategy focused on generating an immediate purchase. If trade promotion has a long-run benefit, it is only to the extent that today's purchase influences a customer to repeat purchase. In contrast, advertising is usually viewed as part of a long-term strategy and an investment in brand equity.

In this paper, we explore the issue of whether trade promotion has positive or negative long-term effects. Our theory is based on the information contained in a trade promotion. We claim that previous displays and features inform consumers about the presence and price of the product. These two pieces of information have a positive and a negative effect on future product purchase probability. Multiple exposures to the product are a weak form of advertising which has a positive effect on future purchase probability. In contrast, multiple exposures to price information increases consumer price sensitivity which has a negative effect on future purchase probability. We find evidence for these two effects using scanner data for laundry detergents.

Thesis Committee:

John D.C. Little, Institute Professor, Professor of Management Science, Co-Chair
Birger Wernerfelt, Professor of Management Science, Co-Chair
David Scharfstein, Dai-Ichi Kangyo Bank Professor Of Management

Acknowledgments

There are many people to whom I am very grateful over the last four years for their guidance and friendship. The close-knit nature of the marketing department has provided me with a unique academic experience as a doctoral student. I have benefited from my interaction with all of the faculty, students, and staff in the department. In particular, I wish to thank John Little and Birger Wernerfelt.

One of the main reasons I came to MIT was to work with John. When I arrived, I was introduced to two important pieces of John's life: scanner data and wine. I have now learned which is more important. John has always been supportive of my work and has given me the freedom to pursue all of my interests, even when they diverge from his own. I can only hope some of his wisdom has rubbed off on me.

During my first weeks at MIT, I sat down with Birger and discussed a course schedule for the first two years. This was just the first of many times that Birger would sit down with me and provide valuable advice. When I needed feedback on my work or hit a road block, Birger was always available. His sharp and direct comments were always on the mark. Birger encouraged me to do my best. One cannot ask for more in an advisor.

Most of all, I wish to thank my wife, Liz, for four years of support, patience, and unconditional friendship. Whether it was studying for generals, weekend work to prepare for the job market, or late night hours to finish this thesis, Liz has always been at my side and there to support me. I cannot thank you enough, but I have a lifetime to try.

Table of Contents

Abstract	3
Acknowledgments	5
Essay 1: Dynamic Competitive Promotion Policies	
Introduction	10
Section 1: Related Literature	14
Section 2: The Model	17
Section 3: Results	21
Section 4: Discussion	31
Conclusion	33
Appendix 1	34
Appendix 2	45
References	48
Essay 2: Why Advertise Both Regular and Sale Prices?	
Introduction	52
Section 1: Related Literature	56
Section 2: General Model	59
Section 3: Search Model	62
Section 4: Wait Model	70
Section 5: Empirical Test	77
Section 6: Discussion	83
Conclusion	87
Appendix 1	89
Appendix 2	97
Appendix 3	99
Appendix 4	102

Appendix 5	103
Appendix 6	108
References	124
Essay 3: Long Term Effects of Trade Promotion:	
<i>Brand Advertising or Brand Erosion?</i>	
Introduction	130
Section 1: Theoretical Framework	134
Section 2: Data	136
Section 3: Model	141
Section 4: Hypotheses	145
Section 5: Results	147
Section 6: Discussion	154
Conclusion	156
Appendix 1	157
Appendix 2	158
Appendix 3	159
References	161

Essay 1

Dynamic Competitive Promotion Policies

Introduction

In many markets, we tend to observe competing sellers promoting products in alternating periods rather than at the same time. To motivate the paper, let us begin with a few examples. An example that many people are familiar with is promotion of Coke and Pepsi¹. Within a supermarket, Coke and Pepsi tend to be promoted in alternating weeks. Another manufacturer example comes from Lal (1990). He provides empirical evidence from dish washing detergents which suggests that manufacturers tend to offer trade deals at different times rather than at the same time. At the retailer level, we also find evidence of alternating promotions. While competing supermarkets often promote the same products, they tend to do so in alternating periods. Consider refrigerated Tropicana Orange Juice which was promoted by any store in a particular market 102 out of 130 weeks. Analyzing two competing stores, one finds the product promoted at both stores simultaneously during only three weeks. In contrast, the product is promoted at only one store during twenty one weeks and thirty five weeks respectively². Thus we have examples of both competing manufacturers and retailers promoting similar (in one case identical) products in alternating periods. The question we wish to ask is why do sellers prefer to promote similar products in alternating periods rather at the same time? What type of customer behavior can help explain this phenomenon?

This paper demonstrates that as customers become more loyal, sellers are more likely to promote their products in alternating periods. We consider a simple market where some customers make a repeat purchase. When a customer offers potential repeat business, a seller is concerned with not only today's sale, but all future sales from that customer. For example, retailers exert considerable effort keeping customers satisfied and meeting their diverse needs with a goal of building a loyal base

¹ It should be noted that at one point Coke and Pepsi wrote contracts which allocated alternating weeks in a supermarket to each firm. This was contested by smaller manufacturers, such as Royal Crown, as an illegal practice. Whether or not explicit contracts are currently written, the de-facto practice today is for alternating promotion. The intuition in this model suggests that in the absence of a contract, one should still tend to observe Coke and Pepsi alternating promotions.

²The dataset contains three national brands carried by all stores: Tropicana, Citrus Hill, and Minute Maid. Analyzing each brand individually, $\text{Prob}(\text{retailer A promote}|\text{retailer B no promote}) > \text{Prob}(\text{retailer A promote}|\text{retailer B promote})$ has the right sign for all brands and is statistically significant ($p < .05$) for Tropicana Regular and Citrus Hill. The author thanks IRI, Pete Fader, and Bruce Hardie for the data.

of customers. A recent Harvard Business Review article notes that "Companies are willing to spend now to build customer loyalty for the long term. They typically look at the customer's lifetime value to the company, not the value of any single transaction." (Treacy and Wiersema 1993) Presumably a customer that is loyal to a seller is more likely to buy from that seller on a future purchase occasion. In this model, a customer that *ceteris paribus* prefers a seller after purchasing a product from that seller will be referred to as loyal. A customer is loyal to the extent that a competing seller must offer a lower price to get the customer to switch sellers. Every seller, whether it be a manufacturer or retailer, wants a loyal base of customers as this provides them with some degree of monopoly power. A natural question to ask is how do customers become loyal?

Manufacturers and retailers would like to believe that they have the ability to create loyal customers. For example, customers may learn which sellers they prefer through their purchase experience. Consider a customer's store preference. After purchasing an item from a store, a customer has more information about uncertain store attributes such as friendliness, level of salesperson assistance, after sales support, and availability of products. This new information enables a customer to update his beliefs about how well a particular store met his needs relative to other stores in the market. Similarly, after a customer purchases and then uses a product, (e.g. Coke or Pepsi), a customer has more information. Presumably if a customer likes a store or product, he is more likely to buy from that seller in the future. While resolving uncertain information is one possibility, our general definition of loyalty only implies a preference to return to a seller. Therefore we allow for a wide variety of subtle mechanisms which need not be as straight forward.

Next it seems logical to ask the sellers how evolving customer preferences (loyalty) will influence their pricing policy. To understand a seller's incentives, consider one customer that is indifferent between two retailers (a "switcher") and is going to purchase an item today as well as in at least one future period. A retailer that sells to this customer has an incentive to promote today (lower price) with the objective of getting him into the store. If the retailer provides our customer with a positive purchase experience, he may be more likely to return to that store on a future purchase occasion. Let's assume a retailer promotes, our customer makes a purchase at the retailer and he subsequently develops a preference for the retailer. The retailer would like our customer to return to the store on his next purchase occasion but wants to charge him as high a price as possible on that repeat visit. When prices are

equal across retailers, this customer prefers the store previously visited. This allows the retailer to slightly increase prices until our customer is indifferent between switching stores. Thus a retailer selling to our single customer has incentive to promote when the customer is indifferent to get our customer into the store and to generate store loyalty. To the extent our customer prefers this store (i.e., is loyal to the store), there is an incentive to increase prices in the future.

Moving away from our single customer, competing sellers generally face a trade off when selling to customers that make a varying number of future purchases (new and old) and that have varying degrees of store loyalty (loyal and switching). On the one hand, a seller may price slightly higher than the competition and attract mainly loyal customers. Alternatively, a seller may attempt to undercut the competition by promoting and attract both switching and loyal customers. Under either strategy, a seller would like to serve its loyal customers. However, when a retailer promotes to get both switching and loyal customers, there is an opportunity cost due to the fact that the retailer could have charged a higher price and served just the loyals. As the number of loyal customers increases, promoting becomes less attractive. In a market with both switching and loyal customers, retailers face a tradeoff between a larger market share (loyals + switching) and lower prices versus smaller market share (loyals) and higher prices.

Combining these basic elements in a dynamic setting completes the framework of our model. At any point in time, there are switching customers in the market some of whom will make a future purchase. Of these indifferent repeat purchasers, a portion may become loyal to the store where they last purchased. Finally, each customer's purchase stream is finite and eventually a customer leaves the market. Now consider how competing retailers will efficiently serve these customers. Suppose a retailer charges a higher price (regular price) and sells primarily to loyal customers. Because customers eventually leave the market, this retailer's loyal customer base will gradually erode. As the number of loyal customers diminishes, promotion becomes more attractive as a means to attract new customers and increase the loyal base of customers. However, our retailer must ask when is the optimal time to promote? To avoid price competition, it is more efficient to promote when the competing retailer is less likely to promote. But our retailer knows that the competition is less likely to promote when they have more loyals. When does the competitor's loyal base increase? When the competitor promoted in the previous period. Thus it is more desirable for our retailer to promote when a competitor promoted in the previous

period. These basic incentives to efficiently attract customers and generate loyalty create a situation where sellers tend to take turns promoting an item.

The paper is organized as follows. Section 1 reviews related literature. In Section 2, we present the model and in Section 3 we present results from the model. Section 4 is a discussion and interpretation of the results and we conclude with a summary.

Section 1: Related Literature

This paper contributes to the promotions literature on selling to heterogeneous consumers. There is a vast amount of literature in marketing and economics in this area which can be broadly classified into two groups: Static Models and Dynamic Models. We will address related work in both streams of research and highlight the contribution of this paper.

Static models of promotion that allow for customer heterogeneity typically consider segments of customers (e.g., loyals and switchers) or use some variant of a Hotelling line. A general intuition in these models is that firms charge higher prices when they have more loyal customers and lower prices when there are more switching customers. We see this intuition in Narasimhan (1988) who considers a duopoly with two customer segments: loyals and switchers. Loyals are "locked in" to one of the firms and switchers are indifferent. This assumption is later relaxed and switchers are allowed to be loyal in the sense of our paper. In this static game, Narasimhan concludes that the firm with more loyal customers will price higher. Simester (1993) looks at a market with both loyals and switching customers in a market where stores are pricing multiple products. Simester finds that it is more desirable for firms to discount a product that has fewer loyal buyers. As static models are a special case of our model where stores behave myopically and try to maximize current period profits it is not surprising that we find this same intuition in our dynamic model.

A related work of interest is Shilony's (1979) static oligopoly model. Shilony considers a market with n firms each with a unit mass of customers that has some degree of loyalty to the firm. Shilony finds a unique symmetric mixed strategy equilibrium for any oligopoly with n firms. Since firms are identical, they all have the same pricing strategy. Raju et al (1990) consider a duopolistic variant of this game with one loyal segment of customers that are locked in to a firm and one switching segment that are loyal but not locked in to the other firm. Raju et al find that the firm with the locked in loyals will tend to promote less often. However, given that this firm does promote, they charge a lower expected price than the competing firm. Our model will have similar properties because, as we will see, our model reduces to Raju et al in a particular case.

While static models provide us with important insights, retailers generally face a dynamic environment. We take the perspective that at any point in time, the retail

environment may be in one of multiple states. If the probability of ending up in any state was a random draw then static models would be sufficient. However, dynamic models offer deeper insight when a retailers past actions directly influence the current state. A key component of our model is that customer preferences are not static. In fact, store visits play a key role in the formation of preferences. Thus dynamic models allow us to capture some strategic behavior which may not be evident in the more simple static case.

An example of such a dynamic model is Farrell and Shapiro (1988). They consider a duopoly where customers enter the market indifferent between firms ("youngsters") and develop a cost of switching retailers ("oldsters") after the first period. A key assumption is that they only consider a Stackelburg (leader/follower) game where either firm may take the lead role in posting price. They find that the firm with oldsters always leads and sells to only his customer base and the competing firm follows and sells to "youngsters." While the firms take turns selling to oldsters and youngsters, there is no alternating price pulsing as the "following" firm always matches the leaders price. Wernerfelt (1991) also considers a model where customers become more and more locked in to one firm over time. This model incorporates a similar dynamic preference formation to our model, however, Wernerfelt's model does not have alternating price pulsing³. Rosenthal (1982) constructs a model where customers only search for prices if the firm they are loyal to raises prices, in which case customers switch to the firm offering the lowest price. While reasonable, this type of customer behavior does not lead to a pulsing equilibrium. Beggs and Klemperer (1992) construct a multi-period model of competition where a fraction of new customers enter and old customers leave each period. They show that firms with more old customers charge higher prices and sell to fewer new customers. Again, the authors do not show a price pulsing equilibrium.

All of these models address different variants of preference formation and price setting behavior. They all use the intuition that firms want to charge higher prices to loyal customers and lower prices to new customers. However, none of the papers can explain why firms would prefer to take turns promoting rather than compete head to head.

³Wernerfelt finds a steady state single price equilibrium. He suggests that there may be other equilibria as well.

In related work, Lal, Little, and Villas-Boas (1994) consider a dynamic game of manufacturer trade deals with and without retailer inventory constraints. Inventory constraints make it less desirable for a retailer to accept a trade deal from a manufacturer whose product has been stock piled in the retailer's inventory. Thus to get his trade deal accepted, the manufacturer would have to offer an even deeper trade deal. The authors show how this leads to manufacturer trade deal pulsing. In a recent paper by Villas-Boas (1993), he offers an explanation for why firms take turns advertising and finds an equilibrium where firms alternate between (advertise, high price) or (no advertise, low price). As price is negatively correlated with advertising spending there is price pulsing. The explanation in Villas-Boas centers around decay of advertising and a customer's evoked set. It is important to recognize that explanations such as these may also help explain promotion pulsing. Promotion pulsing in our model hinges on a different explanation that relies on dynamic preference formation or loyalty building.

Section 2: The Model

We consider a market with two retailers⁴, R_1 and R_2 , selling a single product in multiple periods to both new and old types of consumers. A new consumer is assumed to be indifferent between the two retailers and becomes an old consumer after a purchase occasion⁵. After a purchase, a new consumer becomes an old consumer that either remains indifferent, becomes loyal to a store, or leaves the market. Every new customer lives for two periods. Thus the evolution of one group of new customers is as follows: a unit mass of customers arrives in period 1. In period 2, γ of these customers remain indifferent, θ become loyal to the store at which they made a purchase, and $(1-\theta-\gamma)$ leave the market. At the end of period 2, all of these customers leave the market. Since customers continuously enter/exit the market, there is a unit mass of new customers, a γ mass of old indifferent customers, and a θ mass of loyal customers in every period.

We define the degree to which a customer is loyal as the additional amount a customer is willing to pay before switching from the preferred retailer. The parameter c represents the discount a competing firm must offer to get a loyal customer to switch stores. The parameter c is allowed to take on any positive value. For large values of c , loyal customers are "locked in" at the store they purchased at last period and for small values of c , loyal customers will switch retailers for a small difference in price. Customer loyalty has no effect on reservation prices as both new and old customers have a reservation price of r at both retailers. Loyalty simply enters the utility function as a cost of switching retailers⁶. Other consumer costs, e.g., transportation costs, are assumed to be equal across retailers and have no effect on consumer choice.

There are two stores in the market selling a single product in multiple periods. Our definition of product encompasses all dimensions of the delivered product including the actual product itself as well as retail service and support. The stores may or may not offer differentiated products. Each consumer has unit demand for this product every period. Stores compete by simultaneously posting prices for a single

⁴ For exposition, we consider two retailers. More generally, there are two sellers.

⁵Indifference of new consumers may be due to limited information.

⁶Without changing the main results, one could assume that loyalty shifts the reservation price from r to $r+c$. The main effect this has on the model is to shift the support of the distribution.

product and all types of customers have perfect knowledge of these prices before purchasing. New and Old Indifferent consumers will purchase from the store offering the lowest price. Loyal consumers will purchase from their preferred store if the difference in prices is less than or equal to c .

One can think of θ and c as store specific parameters. In any period, a store has the ability to make θ of the new customers loyal at a degree of loyalty equal to c . The parameters θ and c can have multiple interpretations. For example, θ may be the percentage of people who were satisfied with the service provided and c is the extent to which they were satisfied. This interpretation is consistent with the store visit revealing *new information* to the customer. Other possible interpretations of θ and c include *switching costs*, *risk aversion*, and *psychological perception*. Switching costs often arise when a transaction with a firm leads to a firm specific investment. For example, items in a supermarket are easier to locate after the first visit at a store. Loyalty may also exist because of risk aversion when there is fear of a bad outcome at a competing store. In other words, absence of bad information may also generate store loyalty. Finally, experimental evidence from Carpenter and Nakamoto (1989) suggests that an incumbent brand (i.e., a store visited first) may enjoy a perceptual advantage over later entrants. Thus what we model as loyalty may have multiple interpretations. Managers would like to take the perspective that a firm takes actions which generate store loyalty. The effectiveness at generating customer loyalty may be different across firms (different θ and c). However, for analytic tractability we restrict ourselves to a symmetric equilibrium where both firms have the same θ and c .

The parameter γ , which determines the proportion of customers that do not change their store preference after a purchase experience, may also have multiple interpretations. For example, type γ customers may be very price sensitive such that any difference in other store attributes will be dominated by small price differences. It is also possible that these customers remain indifferent because stores have positioned themselves in a very similar manner along attributes that are important to these customers and therefore stores are difficult to distinguish. Finally, if all new customers have the same prior expectations at both stores, customers that remain indifferent may be those whose expectations were met rather than exceeded. Thus, the stores continue to be indistinguishable to some customers.

These simple assumptions give us an environment with two states of the world which we label the focal state (FS) and the alternative state (AS). In the focal state, retailer 2 has a loyal segment of size θ and in the alternative state, retailer 1 has a loyal

segment of size θ . In either state there are $(1+\gamma)$ switching customers. The state is determined by the firm that announces the lowest price and sells to the new customers. Table 1 summarizes the description of the game.

Table 1	
<u>Retailers:</u>	
Retailer 1	R_1
Retailer 2	R_2
<u>Size of Customer Segments:</u>	
New	1
Loyal	θ
Old Indifferent	γ
Leave Market	$1-\theta-\gamma$
<u>States</u>	
Focal State (FS):	R_2 has θ loyal customers, R_1 no loyal customers
Alternative State (AS):	R_1 has θ loyal customers, R_2 no loyal customers
Summary of Game	

Let p_i be the price of retailer i in the focal state. The current period payoffs⁷ for the firms in the focal state are defined as:

$$\Pi_1(p_1, p_2 | \text{Focal State}) = \begin{cases} 0 & \text{if } p_1 > p_2 \\ p_1(1+\gamma) & \text{if } p_1 \leq p_2 - c \\ p_1(1+\theta+\gamma) & \text{if } p_1 < p_2 - c \end{cases}$$

⁷We assume that if $p_1 = p_2$, the firm with No Loyal customers gets all the new customers. Similarly, in the focal state, if $p_1 = p_2 - c$, the loyals continue to choose R_2 .

$$\Pi_2(p_1, p_2 | \text{Focal State}) = \begin{cases} p_2(1+\theta+\gamma) & \text{if } p_1 > p_2 \\ p_2\theta & \text{if } p_1 \leq p_2 - c \\ 0 & \text{if } p_1 < p_2 - c \end{cases}$$

If R_1 undercuts R_2 by more than c , R_1 gets the whole market. If R_1 undercuts R_2 by less than c , R_1 only gets the new customers and the old indifferent customers. Finally, if R_1 announces a higher price than R_2 , R_1 sells nothing. Given the current period payoffs, we can now compute the continuation payoffs in each state. Let V_i^{FS} and V_i^{AS} be the payoffs to retailer i in the focal state (FS) and the alternative state (AS). The distribution function $F_1(p)$ and $F_2(p)$ represent the CDF of R_1 and R_2 in the focal state and δ is the discount factor (≤ 1). Using this notation, we get the following continuation payoffs in the focal state:

$$V_1^{\text{FS}} = p[(1+\gamma)[1-F_2(p)] + \theta[1-F_2(p+c)]] + \delta[V_1^{\text{AS}}[1-F_2(p)] + V_1^{\text{FS}}F_2(p)] \quad (1)$$

$$V_2^{\text{FS}} = p[(1+\gamma)[1-F_1(p)] + \theta[1-F_1(p-c)]] + \delta[V_2^{\text{FS}}[1-F_1(p)] + V_2^{\text{AS}}F_1(p)] \quad (2)$$

To interpret these equations, consider V_1^{FS} . Fixing a price for retailer 1 at p , the probability that this is lower than retailer 2's price is $[1-F_2(p)]$. Retailer 1's marginal payoff from this event is $p(1+\gamma)$ which is the price charged times the incremental share of customers. Recall that retailer 1 has no loyal customers. If he undercuts retailer 2 by an amount c , then he also serves all of retailer 2's loyal customers. The expression $[1-F_2(p+c)]$ represents the probability of this event and now the marginal payoff to retailer 1 is $p\theta$. Finally, we have to consider what state we will be in next period. If retailer 1 has a higher price than retailer 2, he will remain in the focal state with no loyal customers. Thus with probability $F_2(p)$ retailer 1 will be in the focal state with expected payoff V_1^{FS} . Similarly, with probability $[1-F_2(p)]$, retailer 1 will have a lower price than retailer 2 and be in the alternative state next period with expected payoff V_1^{AS} . We employ the symmetry of our problem to simplify notation: $V_1 = V_1^{\text{FS}} = V_2^{\text{AS}}$ and $V_2 = V_2^{\text{FS}} = V_1^{\text{AS}}$. The mixing distributions, $F_i(p)$, are also symmetric with respect to the focal and alternative states. We will use these equations to calculate the equilibrium in Section 3.

Section 3: Results

We consider only symmetric Markov Perfect equilibria. By definition, in a Markov perfect equilibria firm's payoff functions are only a function of payoff relevant states of the world. Thus in our model, only the size of each firm's loyal segment determines the current state of the world and other historical information is not used in computing payoffs. The restriction to Markov perfect equilibria has several advantages over other equilibria refinements. For a brief discussion and references see (Villas-Boas 1993). The description of the equilibria is organized as follows. First we show that there are no pure strategy equilibrium. Next, we describe the mixed strategy equilibrium. Finally, we demonstrate that other equilibria are not possible.

Claim 1: There are no pure strategy equilibria.

Proof: See Appendix 1.

The intuition for no pure strategy equilibrium is that undercutting strategies prevent a single price point from being supported. At any price pair (p_1, p_2) firms will either have an incentive to raise prices or to undercut the opponent and steal share. Thus even for large⁸ c and θ , there must always be some mixing to prevent these undercutting strategies.

Claim 2: The following fully characterizes all symmetric mixed strategy equilibria where $(V_1, V_2, \bar{p}_1, \underline{p}_1, \bar{p}_2, \underline{p}_2)$ depend on the parameters (c, θ) and are the solution to a set of six equations for the three regions of parameter space. $\bar{p}_i, \underline{p}_i$ are the upper and lower bounds of the support of the distribution of R_i in the focal state.

⁸ But Finite

Probability Distribution in the Focal State

Price	$F_1(p)$	$F_2(p)$
$p < \underline{p}_1$	0	0
$\underline{p}_1 \leq p < \bar{p}_2 - c$	$1 - \left[\frac{V_2 - \delta V_1}{\theta(p+c)} \right]$	0
$\bar{p}_2 - c \leq p < \underline{p}_2$	$1 - \left[\frac{V_2 - \delta V_1}{\theta \bar{p}_2} \right]$	0
$\underline{p}_2 \leq p < \bar{p}_1$	$\frac{p(1+\gamma+\theta) - V_2(1-\delta)}{p(1+\gamma) + \delta(V_2 - V_1)}$	$\frac{p(1+\gamma) + \delta V_2 - V_1}{p(1+\gamma) + \delta(V_2 - V_1)}$
$\bar{p}_1 \leq p < \underline{p}_1 + c$	1	$\frac{\bar{p}_1(1+\gamma) + \delta V_2 - V_1}{\bar{p}_1(1+\gamma) + \delta(V_2 - V_1)}$
$\underline{p}_1 + c \leq p < \bar{p}_2$	1	$1 - \frac{1}{\theta} \left[\frac{V_1 - \delta V_2}{p-c} - (1+\gamma) \right]$
$p \leq \bar{p}_2$	1	1

Table 2

The support of the distributions comes from the following. First note that $\bar{p}_2 \geq \bar{p}_1$ in any equilibrium as all $p_1 > \bar{p}_2$ lead to zero profit for retailer 1. Second, $\underline{p}_2 \geq \underline{p}_1$ in any equilibrium as any $p_2 < \underline{p}_1$ lead to less profit for retailer 2 than $p_2 = \underline{p}_1$. Finally, consumers have a reservation price of 1 and, as we have shown, there are no pure strategy equilibrium, $1 \geq \bar{p}_2 > \underline{p}_2$ and $1 \geq \bar{p}_1 > \underline{p}_1$ must also hold. This leads to the following conditions:

$$\bar{p}_2 \geq \bar{p}_1 \quad (3)$$

$$\underline{p}_2 \geq \underline{p}_1 \quad (4)$$

$$r \geq \bar{p}_2 > \underline{p}_2 \quad (5)$$

$$r \geq \bar{p}_1 > \underline{p}_1 \quad (6)$$

Now consider the flat regions of the distribution. When $p_1 \leq \underline{p}_2$, R_1 has undercut R_2 and has gained the switching customers for sure. However, for all $\bar{p}_2 - c \leq p_1 \leq \underline{p}_2$, R_1 has zero probability of getting the loyal customers. Thus all prices in this range are dominated by $p_1 = \underline{p}_2$. A similar argument applies to the region $\bar{p}_1 \leq p_2 \leq \underline{p}_1 + c$. In this region, R_2 has lost the Switching customers for sure but has zero probability of losing the loyal customers. Thus, $p_2 = \underline{p}_1 + c$ dominates all prices in this range. Figure 2 provides a useful visual interpretation of the support of the probability distribution.

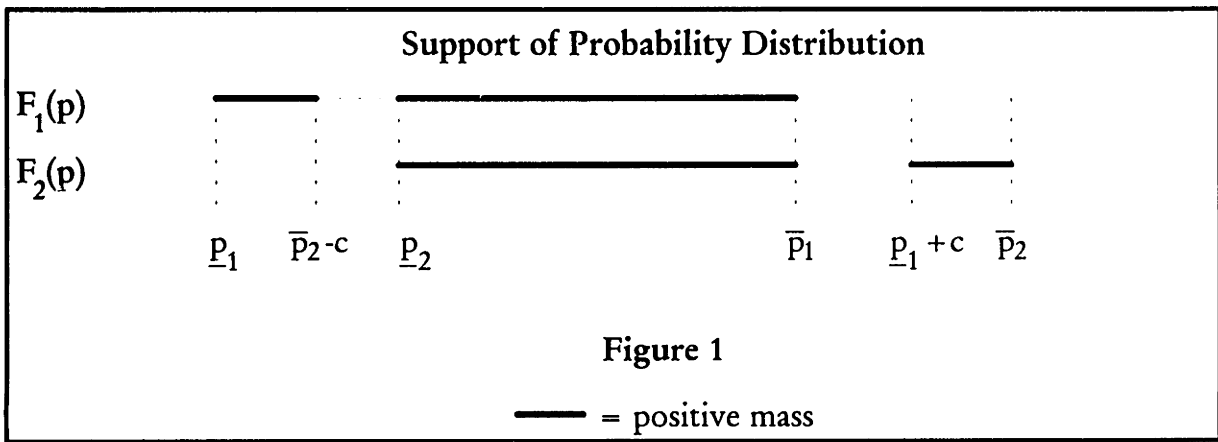
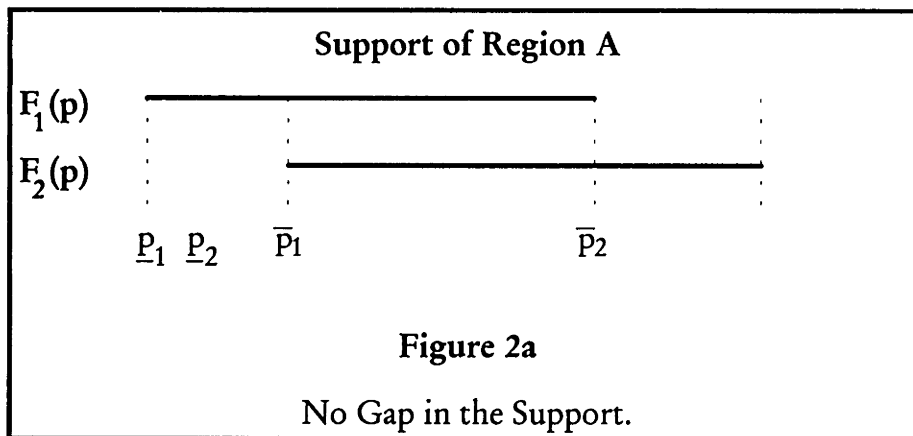


Figure 1 is a general description of the support of the equilibrium. For our specific model, the support will change depending on the parameters c and θ . We find that there are three relevant regions (labeled Regions A, B, and C) of parameter space for our model. We will specify the equations that define the payoffs and support points $(V_1, V_2, \bar{p}_1, \underline{p}_1, \bar{p}_2, \underline{p}_2)$ in each region as well as describe the intuition for the support of each region.

In region A, θ and c are "small" and R_1 and R_2 each mix over a support of size c . Neither firm is able to charge the reservation price $r=1$ because either θ is small and there are few loyal customers, or c is small and it is easier to steal customers from the opponent. There are no gaps in the support, but R_1 tends to charge a lower price because he has no loyal customers and therefore no opportunity cost. It is possible for R_1 to steal loyal customers, however, this only happens if R_1 charges a "low" price and R_2 charges a very "high" price.

$$\bar{p}_1 - \underline{p}_1 = c \tag{7}$$

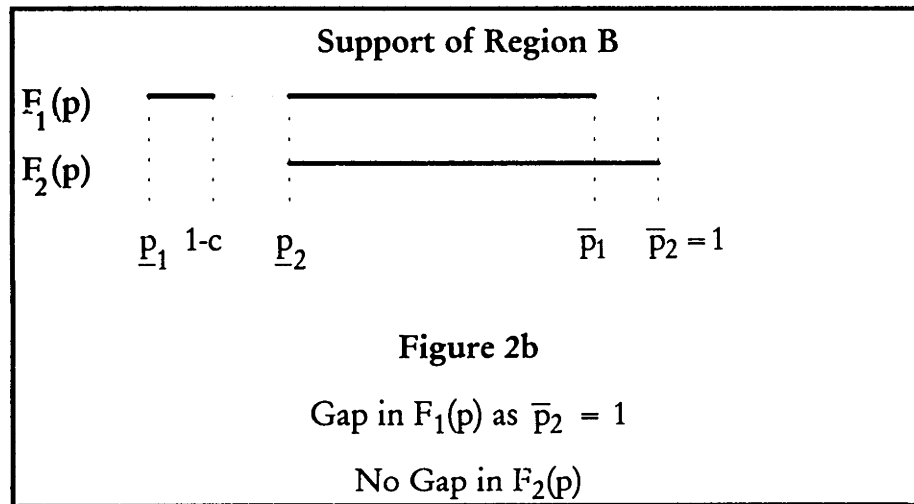
$$\bar{p}_2 - \underline{p}_2 = c \tag{8}$$



In Region B, θ and c are large enough such that R_2 can charge the reservation price of his loyal customers. However, c is small enough that the loyal customers are not locked in. Thus R_1 has an opportunity to steal the loyal customers from R_1 by charging prices in the range $p_1 \in [\underline{p}_1, r - c)$. R_2 responds and protects his loyal share by mixing over the range $p_2 \in [\bar{p}_1, r]$. Over the range $p_2 \in [\underline{p}_2, \bar{p}_1]$, R_1 and R_2 compete for the switching customers. There is a discontinuity in the support because R_2 does charge $p_2 = r$ some of the time. When R_1 undercuts, the first price that will steal share is $r - c$. The lower bound of R_2 's support will be above $r - c$ because charging a low price results in a large opportunity cost to the loyal share. Thus we have $\bar{p}_2 - \underline{p}_2 < c$.

$$\bar{p}_1 - \underline{p}_1 = c \tag{9}$$

$$\bar{p}_2 = 1 \tag{10}$$

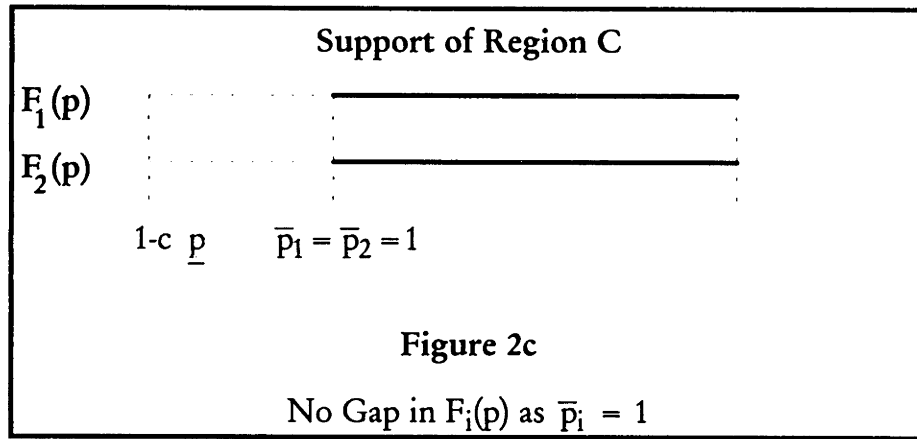


In Region C, both θ and c are large enough⁹ that R_1 cannot steal the loyals from R_2 . Both retailers charge a regular price $p=r$ some of the time. However, R_1 charges $p=r$ less frequently. Both firms mix over the range $p \in [\underline{p}, r]$ in an attempt to get the switching consumers. Intuitively, this follows from Claim 1 as the price $p_i=r$ is not an equilibrium. Even for large c and θ , R_2 will prefer a small price cut and the whole market to the loyal consumers and $p=r$. In a sense, R_1 must now mix to "protect" R_2 from undercutting. R_1 has no incentive to charge a lower price than R_2 thus both firms have the same lower bound on their support (\underline{p}). Large θ and c lead to a large opportunity cost to R_2 for $p < r$. Thus the lower bound of the support will be larger than $r-c$. This leads to the conditions $\bar{p}_1 - \underline{p}_1 < c$, $\bar{p}_2 - \underline{p}_2 < c$, and:

$$\bar{p}_1 = r = 1 \tag{11}$$

$$\bar{p}_2 = r = 1 \tag{12}$$

⁹The region θ large and $c > r$ is obviously part of this region. However, the region does exist for $c < r$. See Raju et al (1990) for an example.



In addition to each pair of region specific equations, four general equations apply to Regions A and B. This gives us a system of six equations and six unknowns.

$$[V_1 - \delta V_2 - (1+\gamma) \underline{p}_1][(1+\gamma) \bar{p}_1 + \delta(V_2 - V_1)] - V_1(1-\delta)\theta \underline{p}_1 = 0 \quad (13)$$

$$[-\theta \bar{p}_2 + (1+\gamma) \underline{p}_2 + \delta(V_2 - V_1)][V_2 - \delta V_1] = \theta^2 \bar{p}_2 \underline{p}_2 \quad (14)$$

$$V_2 - \delta V_1 = \theta(\underline{p}_1 + c) \quad (15)$$

$$V_1 - \delta V_2 = \underline{p}_2(1+\gamma) \quad (16)$$

The equations for Region C are slightly simpler as both R_1 and R_2 have the same support: $[\underline{p}, r]$.

$$V_1 = \underline{p}(1+\gamma+\delta\theta)/(1-\delta) \quad (17)$$

$$V_2 = \underline{p}(1+\gamma+\theta)/(1-\delta) \quad (18)$$

$$\underline{p} = \theta/[1+\gamma+\theta + \delta\theta] \quad (19)$$

Proof: See Appendix 1.

Details of the proof of Claim 2 are in Appendix 1. With a little algebra, one can rewrite equations 15 and 16 as:

$$V_2(1-\delta^2) = \theta(\underline{p}_1 + c) + \delta \underline{p}_2(1+\gamma) \quad (20)$$

$$V_1(1-\delta^2) = \delta\theta(\underline{p}_1 + c) + \underline{p}_2(1+\gamma) \quad (21)$$

These payoffs can be interpreted intuitively. First recognize that both retailers must be indifferent over the support of the distribution. At the highest price that retailer 1 charges ($\underline{p}_1 + c$), retailer 2 sells to only the loyal customers and earns payoff $\theta(\underline{p}_1 + c)$. Similarly, at the lowest price that retailer 2 charges (\underline{p}_2), retailer 1 sells to all the indifferent customers and earns payoff $\underline{p}_2(1 + \gamma)$. Recognizing that the problem is symmetric, the payoffs seem quite intuitive. Retailer 2's focal state expected payoff, is equivalent to alternating between selling to only the θ loyals today at a price ($\underline{p}_1 + c$) and selling to only the $(1 + \gamma)$ indifferent customers next period at a price \underline{p}_2 . Retailer 1's payoff have a similar interpretation. Thus while our equilibrium solution is somewhat complex, the payoffs for the firms are actually quite intuitive.

It is interesting that the support of the distribution is always less than or equal to c . This is a result first found by Shilony (1979) in looking at a static oligopoly with loyal segments of customers. Shilony finds that n symmetric firms each with a loyal segment of size c mix over a support of size less than or equal to $2c$. Our model has a similar setup and we find that firms mix over a support of size less than or equal to c . The intuition for this result is that in equilibrium, firms must mix over a range of prices that keep the opponent indifferent over his support. In equilibrium, for each point of the support, there is only one uncertain outcome! For R_2 , either the loyals are locked in, and competition is over the switching customers; or the switching customers are lost, and R_2 is trying to extract as much rent as possible from the loyals while protecting them, if necessary, from R_1 . If a firm extends the range of the support beyond c , it becomes impossible to mix over prices and still keep the other firm indifferent.

Our main result is stated below.

Proposition 1: The expected price of the firm with more loyal customers, is larger than the expected price of the firm with fewer loyal customers. In the focal state, $E[p_1] < E[p_2]$ and in the alternative state, $E[p_1] > E[p_2]$.]

Proof: See Appendix 2

Previous models with mixed strategy equilibrium have been interpreted as promotions and we apply this interpretation to our model. (Varian 1980) The implication of proposition 1 is that stores will tend to take turns promoting in a stochastic sense. Note that the mixed strategy equilibrium implies that we should

observe periods where both stores promote. This is consistent with the empirical evidence presented in the introduction. We now consider comparative statics in Region C.

Proposition 2: Comparative Statics for Expected Prices and Profits

Degree of Loyalty: c

Expected prices and profits for both firms are weakly increasing (nondecreasing) in c .

Proportion of New customers that become loyal: θ

Expected prices and profits for both firms are weakly increasing (nondecreasing) in θ .

Number of Indifferent Customers: γ

\underline{p} decreases, but profits increase

Patience of Retailers: δ

Expected Prices are decreasing in δ

Expected payoffs are increasing in δ .

Proof: See Appendix 2

Comparative statics on c and θ are intuitively obvious. The more effective retailers are at generating loyalty (c, θ), the greater the prices and profits. Note how the ability of a retailer to "lock in" customers allows the competing retailer to charge a slightly higher price to the switching customers. That is, increases in loyalty raise prices for both loyalists and switchers. This is a slightly counter intuitive result but is consistent with results in Farrell and Shapiro (1988). A firm that is able to increase loyalty has a positive marginal effect on competing firms and increases average market prices and profits.¹⁰

¹⁰One must be careful not to take this too far. One can imagine a situation where firms couldn't generate loyalty but could generate disloyalty (negative c). If competition leads to the Bertrand Outcome ($p=0$), generating disloyalty may allow both firms to price above marginal cost and earn positive profits. In the long run, this seems unlikely as reputation effects matter and such firms would leave the market.

A slightly less obvious result is the comparative static on γ . Holding prices constant, adding more switching customers increases profits. However, the presence of more switching customers decreases the expected price. The net effect of more type γ customers, despite the decrease in price, is still positive.

It is also clear that the difference in expected prices, $E(p_2) - E(p_1)$ in the focal state, is increasing in θ . To understand the intuition, consider $c > 0$ and $\theta = 1$. From Proposition 1, $E(p_2) - E(p_1) > 0$. As θ approaches 0, the model reduces to Bertrand competition ($E(p_2) - E(p_1) = 0$), since only switchers are present, and both prices equal marginal cost. From continuity, the difference in expected prices should decrease as θ decreases.

To illustrate another property of the model, consider a situation where each firm's ability to generate loyal customers differs ($\theta_j \neq \theta_i$). An extreme example is where R_2 generates θ loyal customers and R_1 generates $\theta=0$ loyal customers. In the focal state, R_2 has the same incentive to charge higher prices and extract rent from the loyal consumers. However, in the alternative state, neither firm has loyal customers and we end up in Bertrand. This decreases R_2 's continuation payoff. Therefore, R_2 benefits from R_1 's ability to generate loyal customers. R_2 builds a loyal base of customers in the alternative state at a price proportional to R_1 's ability to generate loyalty. Formally, let p_i be retailer i 's profit and θ_j be retailer j 's loyalty parameter, then $\frac{\partial \pi_i}{\partial \theta_j} > 0$.

Finally, consider two additional limiting properties of the model.

Property 1 (P1)

In the limit as either θ or c approach 0, $F_i(0) = 1$. (Bertrand Competition)

The intuition for property 1 is straight forward. In a market with only switching customers with identical reservation prices, prices will be driven to marginal cost. Plugging in $\theta = 0$ to Table 1, one can see that $F_i(0) = 1$. As c approaches 0, the support shrinks and eventually collapses to a mass point at $p = 0$.

Property 2 (P2)

If firms behave myopically, ($\delta = 0$), region C of our model is equivalent to Raju et al. (1990).

It is appealing that if retailers behave perfectly myopically, our model reduces to Raju et al. The Raju et al model considered competition between a an exogenously specified strong brand and weak brand. In our model, strength of a brand (number of loyal customers) is endogenous.

Section 4: Discussion

While our model focuses on a single product that is purchased in multiple periods it seems that this intuition should carry forward to a more general setting. Many customers purchase a basket of goods from a retailer on a regular basis. Stores tend to offer promotions on a small set of salient items to encourage customers to visit their store. An interpretation of this model is that stores should tend to promote different sets of salient items. Consider a single household and a store that promotes a set of items that encourage the customer to visit the store. In the next period, the customer is likely to have demand for that same basket of goods. However, if the store was successful at generating store loyalty, the customer should be more likely to return even if some of the items aren't promoted. Thus we might expect to see competing firms alternating the baskets of goods that they promote.

Previous models that have interpreted mixed strategy equilibria as promotions often include a mass point which is interpreted as the regular price. To accomplish this one includes a segment of completely loyal customers which leads to a mass point on the reservation price. In our model, a mass point can be found in Region C where customers are effectively locked in to a firm. It is clearly feasible to add loyal segments of customers to both firms and generate similar results with respect to those presented here. We conjecture that results would then be similar to existing papers in this area (Narasimhan 1988 and Raju et al 1990) with the added feature of taking turns.

The main result of this paper is that in a stochastic sense we expect to see retailers "taking turns" promoting their product. It is interesting to discuss interpretations and extensions of the current model. For example, during holiday weekends (e.g. Labor Day) we often observe many retailers having a "Blockbuster Sale!" Why do all retailers tend to promote on these busy weekends? When a large number of customers suddenly enter the market, as on a holiday weekend, there is increased desire to attract these customers. Our model assumes a steady flow of customers into and out of the market. In our notation, an increased number of new customers is equivalent to thinking of a market where a low proportion of customers are loyal (small θ). From proposition 3, the expected prices of both firms will decrease when θ is small. One can interpret this as both firms aggressively promoting an item to attract switching customers. Note that there is incentive for both firms to promote even if few of these customers may become loyal.

It seems interesting to think about the model from the perspective of the retailer's loyalty technology θ and c . In the model, we assume that these parameters are symmetric across firms. What if one firm was better at generating repeat business (higher θ and c)? If either firm increases θ or c , both firms are weakly better off. The intuition for the own-firm effect can be seen in the comparative statics. A firm with higher θ and c is going to tend to charge higher prices to their loyal customers which raises all prices in the market. The model now helps us understand the benefit of generating store loyalty. Managers have direct control over sales support, service, and many factors that influence customer satisfaction. The benefit of generating store loyalty is that customers are more likely to return to your store even when an item is not promoted. Thus a store that is able to get customers into the store and provide them with a positive purchase experience benefits by being able to charge higher prices (regular price) in future periods. Both firms benefit from this as average market prices are higher.

Furthermore, what really matters in the customer's selection of stores is relative satisfaction. In our model, customers live for a few periods for analytic convenience. However, if customers are allowed to live longer the loyalty parameters are now interpreted as relative satisfaction. Customers are only willing to pay a price premium if they receive incremental benefits from a store. Thus, in a more general model the parameters θ and c in our model could be interpreted as relative measures of satisfaction.

Conclusion

We have presented a model that helps explain how the goal of efficiently building a loyal base of customers influences sellers to alternate price promotions. We find that as sellers are able to generate more loyal customers (θ) and more loyalty per customer (c), both profits and prices increase.

The examples we provided at the outset suggest that alternating promotions is a phenomenon which is present for many types of sellers. While other explanations for alternating promotions have been offered, our explanation based on customer loyalty seems generally applicable to many types of competing sellers.

There are many possible extensions to the model. An obvious extension is to investigate retailers selling multiple products. Many of the empirical examples suggest that promotion occurs on a regular basis but for different products. It is possible to add a second product and show how sellers take turns promoting different products in every period. Finally, the degree of loyalty (c) and the number of new customers that become loyal (θ) are exogenous parameters in our model. In the discussion, we note that one result from our model is that a firm benefits from a competitor's ability to generate loyal customers. As firms often think of θ and c as control variables, it would be interesting to look at these issues in more detail. These extensions to the basic model are promising and may be explored in future research.

This paper offers an explanation for the common seller practice of alternating promotions. Our discussion highlights how a rather simple model helps one understand common selling situations. Our consideration of customers with dynamically formed preferences offers deeper insight into critical manufacturer and retailer issues.

Appendix 1: Proof of Claims

Claim 1: There is no pure strategy equilibria for any positive finite θ and c .

Proof: WOLOG consider the focal state. It cannot be optimal for R_1 to charge a price less than or greater than R_2 as R_1 can increase profits by charging $p_1 = p_2$. Thus only price pairs (p, p) are candidates for equilibrium. Now consider R_2 and an equilibrium of (p, p) . For all p , R_2 can earn strictly positive profits by either undercutting or raising price. Formally, at least one of the following inequalities will be true for all p .

1) Undercutting: profits from sell to loyals at current price $<$ profits from lower price by epsilon and sell to everyone. $p\theta < (p-\epsilon)(1+\gamma+\theta)$

2) Raising price: profits from sell to loyals at current price $<$ profits from raise price by c and still sell to loyals. $p\theta < (p+c)\theta$.

Therefore, strategies of (p, p) are not candidates for equilibrium and there is no pure strategy equilibrium.

QED

Claim 2: The distribution functions in Table 2 and equations 13-19 fully characterizes the only symmetric mixed strategy equilibria.

Lemma 1: Assume $\underline{p}_1 + c$ is in the support of R_2 and \underline{p}_2 is in the support of R_1 . In any symmetric mixed strategy equilibrium, there are no mass points at:

- (i) Points of differentiability
- (ii) \underline{p}_i in $G_i(p)$
- (iii) $\bar{p}_2 - c$ in $G_1(p)$, $\underline{p}_1 + c$ in $G_2(p)$

Proof of Lemma1:

(i) By construction, a mass point at a point of differentiability implies a discontinuity in $G_i(p)$ and is therefore impossible.

(ii) We show that given $G_i(p)$, mass points at \underline{p}_i implies that R_i is not indifferent over all possible values of his support and therefore is not possible. Consider a mass point at \underline{p}_1 of size k . By construction, $\underline{p}_2 \geq \underline{p}_1$. If $\underline{p}_2 = \underline{p}_1$, then R_2 will not be indifferent to \underline{p}_2 and $\underline{p}_2 + \epsilon$. Now consider $\underline{p}_1 + c$. At $(\underline{p}_1 + c + \epsilon)$, the probability of R_2 losing the loyal customers increases by k . As ϵ is small, changes in the marginal payoffs for the indifferent customers can be ignored. Looking at the loyal customers, the marginal payoffs are $\theta(\underline{p}_1 + c)(1 - G_1(\underline{p}_1 + c))$ and $\theta(\underline{p}_1 + c + \epsilon)(1 - G_1(\underline{p}_1 + c) - k)$. The only k for which R_2 will be indifferent is $k=0$. Similar arguments hold for \underline{p}_2 by considering R_1 's indifference to undercutting \underline{p}_2 strategies.

(iii) Consider a mass point at $\bar{p}_2 - c$ of size k in $G_1(p)$. Consider \bar{p}_2 for R_2 . Identical arguments as in (ii) apply for payoffs to R_2 . Consider a mass point at $\underline{p}_1 + c$ of size k in $G_2(p)$. Consider \underline{p}_1 for R_1 . Identical arguments as in (ii) apply for payoffs to R_1 .

QED

Table 3 characterizes all possible regions of parameter space.

Table 3

Region		
A	$\bar{p}_1 - \underline{p}_1 = c$	$\bar{p}_2 - \underline{p}_2 = c$
B	$\bar{p}_1 - \underline{p}_1 = c$	$\bar{p}_2 - \underline{p}_2 < c$
C	$\bar{p}_1 - \underline{p}_1 < c$	$\bar{p}_2 - \underline{p}_2 < c$
D	$\bar{p}_1 - \underline{p}_1 < c$	$\bar{p}_2 - \underline{p}_2 = c$
E	$\bar{p}_1 - \underline{p}_1 < c$	$\bar{p}_2 - \underline{p}_2 > c$
F	$\bar{p}_1 - \underline{p}_1 = c$	$\bar{p}_2 - \underline{p}_2 > c$
G	$\bar{p}_1 - \underline{p}_1 > c$	$\bar{p}_2 - \underline{p}_2 = c$
H	$\bar{p}_1 - \underline{p}_1 > c$	$\bar{p}_2 - \underline{p}_2 < c$
I	$\bar{p}_1 - \underline{p}_1 > c$	$\bar{p}_2 - \underline{p}_2 > c$

Lemma 2: Table 2 fully characterizes the mixing distribution of retailer 1 and 2 in regions A, B, and C.

Proof of Lemma 2:

By construction, \underline{p}_i is the lowest price charged by retailer i . Therefore, for $p < \underline{p}_i$ we know that $F_i(\underline{p}_i) = 0$. Similarly, for $p > \bar{p}_i$ $F_i(p) = 1$. We now consider regions of the support with mass other than 0 or 1. In each region, with probability 1 each player will either serve just the loyal or just the indifferent customers. Consider $\underline{p}_2 \leq p < \bar{p}_1$ such that R_2 definitely serves the loyal customers. That is, $F_2(p+c) = 1$ and $F_1(p-c) = 0$. We can write payoffs as:

$$V_1 = p(1+\gamma)[1-F_2(p)] + \delta[V_2[1-F_2(p)] + V_1F_2(p)] \quad (22)$$

$$V_2 = p[(1+\gamma)[1-F_1(p)] + \theta] + \delta[V_2[1-F_1(p)] + V_1F_1(p)] \quad (23)$$

Solving equation 22 for $F_2(p)$ we get:

$$F_2(p) = \frac{p(1+\gamma) + \delta V_2 - V_1}{p(1+\gamma) + \delta(V_2 - V_1)} \quad (24)$$

Equation 24 can be found in Table 1. Similarly, one can solve eqn 23 for $F_1(p)$. Next, consider $\underline{p}_1 + c < p < \bar{p}_2$ and $\underline{p}_1 < p < \bar{p}_2 - c$ such that R_1 definitely charges a lower price than R_2 and sells to the indifferent customers. That is, $F_2(p) = 0$ and $F_1(p) = 1$.

$$V_1 = p[(1+\gamma) + \theta[1-F_2(p+c)]] + \delta V_2 \quad (25)$$

$$V_2 = p\theta[1-F_1(p-c)] + \delta V_1 \quad (26)$$

Solving these equations for $F_i(p)$ and using Lemma 1 completes the remaining entries in Table 1.

QED

Although it is not shown formally for all regions, the regions A, B, and C can be characterized by conditions on θ , γ , and c . For example, Region C holds when $c \geq c^*$ where c^* is defined by R_1 's deviation constraint.

We look for a c such that R_1 does not undercut the price and steal all the share. The relevant condition is:

$$\underline{p}(1+\gamma) + \delta V_2 \geq p[(1+\gamma) + \theta[1-F_2(p+c)]] + \delta V_2$$

Substituting from Table 2:

$$\underline{p} \geq p \left[1 + \theta \frac{V_1(1-\delta)}{p(1+\gamma) + \delta(V_2 - V_1)} \right]$$

As the RHS is increasing in p , it is sufficient to check at the first undercutting price: $p = 1-c$.

$$\underline{p} \geq (1-c) \left[1 + \theta \frac{\underline{p}(1+\delta\theta)}{1+\delta\theta\underline{p}} \right]$$

$$c \geq 1 - \frac{\underline{p}(1+\delta\theta)}{1+\delta\theta\underline{p} + \theta\underline{p}(1+\delta\theta)} = c^*$$

We find it more convenient to characterize the regions as in Table 3.

Consider region A.

Region A: By construction $\bar{p}_1 - \underline{p}_1 = c$, $\bar{p}_2 - \underline{p}_2 = c$

Consider $p_1 = \underline{p}_1$. By Lemma 1, there is no mass point on \underline{p}_1 , thus $F_1(\underline{p}_1) = 0$. Taking the entry from Table 1 and substituting gives equation 15.

$$0 = 1 - \left[\frac{V_2 - \delta V_1}{\theta(p+c)} \right]$$

Now consider $p = \underline{p}_2$. By Lemma 1, there is no mass point on \underline{p}_2 , thus $F_2(\underline{p}_2) = 0$. This gives equation 16. We now consider deviations outside the support of the distribution and will show that we get the equations 13 and 14.

Eqn 13: We show that deviation constraints for $p_1 > \bar{p}_1$ and $p_1 < \underline{p}_1$ combine to give equation 13.

Consider $p_1 > \bar{p}_1$. In equilibrium, the expected payoff to retailer 1 must be weakly greater than charging $p_1 > \bar{p}_1$. The deviation condition is:

$$p_1(1+\gamma)[1-F_2(p_1)] + \delta[[1-F_2(p_1)]V_2 + F_2(p_1)V_1] \leq V_1$$

Substituting for $F_2(p)$ from Table 2:

$$[p_1(1+\gamma) + \delta(V_2 - V_1)] \frac{1}{\theta} \left[\frac{V_1 - \delta V_2}{p_1 - c} - (1+\gamma) \right] \leq V_1(1-\delta)$$

The LHS of this equation has a negative derivative, so it is sufficient to verify at \bar{p}_1 . This leads to:

$$[V_1 - \delta V_2 - (1+\gamma)\bar{p}_1][\theta(1+\gamma)\bar{p}_1 + \delta(V_2 - V_1)] - V_1(1-\delta)\theta\bar{p}_1 \leq 0 \quad (27)$$

Consider $p_1 < \underline{p}_1$. The deviation condition is:

$$p_1[(1+\gamma) + \theta[1-F_2(p_1+c)]] + \delta V_2 \leq V_1$$

$$p_1 \theta \left[\frac{V_1(1-\delta)}{(p_1+c)(1+\gamma) + \delta(V_2 - V_1)} \right] + \delta V_2 + p_1(1+\gamma) \leq V_1$$

The LHS of this equation has a positive derivative, so it is sufficient to verify at \underline{p}_1 . This leads to:

$$[V_1 - \delta V_2 - (1+\gamma)\underline{p}_1][\theta(1+\gamma)\underline{p}_1 + \delta(V_2 - V_1)] - V_1(1-\delta)\theta\underline{p}_1 \geq 0 \quad (28)$$

Combining (27) and (28), we get equation 13.

Eqn 14: We show that deviation constraints for $p_2 > \bar{p}_2$ and $p_2 < \underline{p}_2$ combine to give equation 14.

Consider $p_2 > \bar{p}_2$. The deviation condition is:

$$p_2\theta[1-F_1(p_2-c)] + \delta V_1 \leq V_2$$

$$p_2\theta\left[\frac{\theta(p_2-c) + V_2 - \delta V_1}{(p_2-c)(1+\gamma) + \delta(V_2-V_1)}\right] \leq V_2 - \delta V_1$$

The LHS of this equation has a negative derivative, so it is sufficient to verify at \bar{p}_2 . This leads to:

$$\theta^2 \bar{p}_2 \underline{p}_2 \leq [V_2 - \delta V_1][-\theta \bar{p}_2 + \underline{p}_2(1+\gamma) + \delta(V_2-V_1)] \quad (29)$$

Consider $p_2 < \underline{p}_2$. The deviation condition is:

$$p_2[\theta + [1-F_1(p_2)](1+\gamma)] + \delta[(1-F_1(p_2))V_2 + F_1(p_2)V_1] \leq V_2$$

$$[p_2(1+\gamma) + \delta(V_2-V_1)][1-F_1(p_2)] \leq V_2 - \delta V_1 - p_2\theta$$

$$[p_2(1+\gamma) + \delta(V_2-V_1)]\left[\frac{V_2 - \delta V_1}{\theta(p+c)}\right] \leq V_2 - \delta V_1 - p_2\theta$$

The LHS of this equation has a positive derivative, so it is sufficient to verify at \underline{p}_2 . This leads to:

$$[\underline{p}_2(1+\gamma) + \delta(V_2-V_1)][V_2 - \delta V_1] \leq [(V_2 - \delta V_1) - p_2\theta]\theta \bar{p}_2$$

$$[-\theta \bar{p}_2 + \underline{p}_2(1+\gamma) + \delta(V_2-V_1)][V_2 - \delta V_1] \leq \theta^2 \bar{p}_2 \underline{p}_2 \quad (30)$$

Combining (29) and (30) and using $\bar{p}_2 - \underline{p}_2 = c$, we get equation 14.

Equations 13-16 and the region specific equations ($\bar{p}_1 - \underline{p}_1 = c = c$, $\bar{p}_2 - \underline{p}_2 = c$) give us 6 equations to solve for the unknown parameters.

Similar arguments follow for region B.

Region B: By construction, $\bar{p}_1 - \underline{p}_1 = c$, $\bar{p}_2 - \underline{p}_2 < c$

Similar to Region A, we will show that the six equations are Equations 13-16, ($\bar{p}_1 - \underline{p}_1 = c$, $\bar{p}_2 = 1$)

Equations 15 and 16 follow immediately as in Region A.

Eqn 13: Deviation constraints for $p_1 > \bar{p}_1$ and $p_1 < \underline{p}_1$ combine to give equation 13 as in Region A.

Eqn 14: We show that deviations by R_2 for $p_2 < \underline{p}_2$ and continuity in $F_1(p)$ combine to give equation 14.

Consider $p_2 < \underline{p}_2$ such that p_2 lies in the flat region of $F_1(p)$. Deviation condition implies:

$$p_2[\theta + [1 - F_1(p_2)](1 + \gamma)] + \delta[[1 - F_1(p_2)]V_2 + F_1(p_2)V_1] \leq V_2$$

$$[p_2(1 + \gamma) + \delta(V_2 - V_1)][1 - F_1(p_2)] + p_2\theta \leq V_2 - \delta V_1$$

$$[p_2(1 + \gamma) + \delta(V_2 - V_1)] \left[\frac{V_2 - \delta V_1}{\theta \bar{p}_2} \right] + p_2\theta \leq V_2 - \delta V_1$$

The LHS of this equation has a positive derivative, so it is sufficient to verify at \underline{p}_2 . This leads to:

$$[\underline{p}_2(1 + \gamma) + \delta(V_2 - V_1)][V_2 - \delta V_1] \leq [V_2 - \delta V_1 - \underline{p}_2\theta][\bar{p}_2\theta]$$

$$[-\bar{p}_2\theta + \underline{p}_2(1 + \gamma) + \delta(V_2 - V_1)][V_2 - \delta V_1] \leq \theta^2 \underline{p}_2 \bar{p}_2 \quad (31)$$

Continuity in $F_1(p)$ at \underline{p}_2 implies:

$$F_1(\underline{p}_2) \geq F_1(\bar{p}_2 - c)$$

$$\frac{\underline{p}_2(1 + \gamma + \theta) - V_2(1 - \delta)}{\underline{p}_2(1 + \gamma) + \delta(V_2 - V_1)} \geq \left[\frac{\theta \bar{p}_2 - (V_2 - \delta V_1)}{\theta \bar{p}_2} \right]$$

$$\left[\underline{p}_2(1 + \gamma + \theta) - V_2(1 - \delta) \right] \theta \bar{p}_2 \geq [\theta \bar{p}_2 - (V_2 - \delta V_1)] \left[\underline{p}_2(1 + \gamma) + \delta(V_2 - V_1) \right]$$

$$\left[\underline{p}_2(1 + \gamma + \theta) - V_2(1 - \delta) \right] [V_2 - \delta V_1] \geq \theta \bar{p}_2 \left[\underline{p}_2(1 + \gamma + \theta) - V_2(1 - \delta) \right]$$

$$[-\bar{p}_2 \theta + \underline{p}_2(1 + \gamma) + \delta(V_2 - V_1)] [V_2 - \delta V_1] \geq \theta^2 \bar{p}_2 \underline{p}_2 \quad (32)$$

Combining (31) and (32) gives equation 14.

Prove $\bar{p}_2 = 1$: $\bar{p}_1 - \underline{p}_1 = c$ follows from definition. $\bar{p}_2 = 1$ follows from the deviation condition on $p_2 > \bar{p}_2$.

Consider $p_2 > \bar{p}_2$ where $p_2 - c$ is in the flat region of $F_1(p)$. Deviation condition is:

$$p_2 \theta [1 - F_1(p_2 - c)] + \delta V_1 \leq V_2$$

$$p_2 \theta [1 - k] \leq V_2 - \delta V_1$$

$$p_2 \theta \left[\frac{V_2 - \delta V_1}{\theta \bar{p}_2} \right] \leq V_2$$

$$p_2 [V_2 - \delta V_1] \leq V_2 \bar{p}_2$$

$$p_2 \leq \bar{p}_2 \quad ==> \text{Contradiction.}$$

Thus the only solution is when $\bar{p}_2 = 1$.

Region C: By construction $\bar{p}_1 - \underline{p}_1 < c$, $\bar{p}_2 - \underline{p}_2 < c$

In this region, c is large enough so that R_2 never loses the loyal customers. Below we define conditions on c such that this region exists. All of the mixing is competition for the switching customers. We show that equation 17-19 and the following characterize the equilibrium Region C:

$$\bar{p}_1 = \bar{p}_2 = 1 \quad (33)$$

$$\underline{p}_1 = \underline{p}_2 = \underline{p} \quad (34)$$

Eqn 33: Just as in Region B, deviation constraints for $p_1 > \bar{p}_1$ and $p_2 > \bar{p}_2$ are only satisfied if $\bar{p}_1 = 1$ and $\bar{p}_2 = 1$. Thus: $\bar{p}_1 = \bar{p}_2 = 1$

Eqn 34: Given $\bar{p}_i = 1$ and $\bar{p}_i - \underline{p}_i < c$, we know that R_1 never steals the loyal customers from R_2 . Given this, R_1 has no incentive to charge a lower price than R_2 . Thus: $\underline{p}_1 = \underline{p}_2 = \underline{p}$.

Eqns 17 and 18: $F_i(\underline{p}) = 0$ gives us:

$$F_1(\underline{p}) = \frac{\underline{p}(1 + \gamma + \theta) - V_2(1 - \delta)}{\underline{p}(1 + \gamma) + \delta(V_2 - V_1)} = 0$$

$$F_2(\underline{p}) = \frac{\underline{p}(1 + \gamma) + \delta V_2 - V_1}{\underline{p}(1 + \gamma) + \delta(V_2 - V_1)} = 0$$

$$V_2 = \underline{p}(1 + \gamma + \theta) / (1 - \delta)$$

$$V_1 = \underline{p}(1 + \gamma + \delta\theta) / (1 - \delta)$$

$$V_2 - V_1 = \underline{p}\theta \quad (35)$$

Eqn 19: Given that R_2 always gets the loyal customers, let's find the lower bound of his support. $p_2 = 1$ leads to profits $\theta + \delta V_1$. Thus $p_2 = 1$ dominates

all $p(1+\gamma+\theta) + \delta V_2 \leq \theta + \delta V_1$. Using 35, we get that all $p \leq \theta/[1+\gamma+\theta + \delta\theta]$ are dominated by $p_2=1$. Given this, R_1 will never charge less than $p_1 = \theta/[1+\gamma+\theta + \delta\theta]$. This defines the bottom of the support for both players. Thus,

$$\underline{p} = \theta/[1+\gamma+\theta + \delta\theta]$$

In regions D and E we show that after solving for the appropriate mixing distribution, the deviation conditions outside the support imply a contradiction. For regions F-I we show that there does not exist a mixing distribution $\tilde{F}_i(p)$ such that R_i is indifferent. Therefore, there are no equilibria in regions D-I.

Region D: Solve for $F_2(p)$ as in Lemma 2 (Table 1). Now consider deviations to by R_1 to $p_1^* > \bar{p}_1$. As in the proof for Region B, this constraint is not satisfied for any p_1^* . This implies the only possibility is $\bar{p}_1 = 1$. But, this cannot be true as this implies $\underline{p}_1 > \underline{p}_2$ which violates a constraint.

Region E: Same proof as in Region D.

Region F: R_1 can't mix to keep R_2 indifferent over his support. As in Lemma 2, we can use equations 1 and 2 and derive the mixing distributions, $\tilde{F}_i(p)$, that must exist for particular regions.

$$\tilde{F}_1(p) = 1 - \left[\frac{V_2 - \delta V_1}{\theta(p+c)} \right] \quad \underline{p}_1 \leq p < \underline{p}_2$$

$$\tilde{F}_1(p) = \frac{p(1+\gamma+\theta) - V_2(1-\delta)}{p(1+\gamma) + \delta(V_2 - V_1)} \quad \underline{p}_2 \leq p < \bar{p}_1$$

Consider $\underline{p}_2 + c \leq p_2 < \bar{p}_2$

$\pi(p_2) = p_2\theta[1 - \tilde{F}_1(p_2-c)] + \delta V_1 = V_2$ and $\underline{p}_2 \leq p_2-c < \bar{p}_1$. Thus:

$$p_2\theta \left[\frac{p\theta + V_2 - \delta V_1}{p(1+\gamma) + \delta(V_2 - V_1)} \right] + \delta V_1 = V_2 \text{ and } V_2 \text{ and } V_1 \text{ are not constant.}$$

Region G: R_2 can't mix to keep R_1 indifferent over his support.

$$\tilde{F}_2(p) = 1 - \frac{1}{\theta} \left[\frac{V_1 - \delta V_2}{p - c} - (1 + \gamma) \right] \quad \underline{p}_1 + c \leq p < \bar{p}_1$$

$$\tilde{F}_2(p) = \frac{\bar{p}_1(1 + \gamma) + \delta V_2 - V_1}{\bar{p}_1(1 + \gamma) + \delta(V_2 - V_1)} \quad \bar{p}_1 \leq p < \bar{p}_2$$

$$\tilde{F}_2(p) = \frac{p(1 + \gamma) + \delta V_2 - V_1}{p(1 + \gamma) + \delta(V_2 - V_1)} \quad \underline{p}_2 \leq p < \underline{p}_1 + c$$

Consider $\underline{p}_1 + c \leq p_1 < \bar{p}_1$.

$$\pi(p_1) = [p_1 + \delta(V_2 - V_1)][1 - \tilde{F}_2(p_1)] - \delta V_1 = V_1. \text{ Thus:}$$

$$[p_1 + \delta(V_2 - V_1)] \left[\frac{1}{\theta} \left[\frac{V_1 - \delta V_2}{p - c} - (1 + \gamma) \right] \right] = V_1(1 - \delta)$$

and V_2 and V_1 are not constant.

Region H: Same proof as in Region G.

Region I: Same proof as in Region G.

QED

Appendix 2: Proof of Propositions

Proof of Proposition 1: Sketch of Proof. When R_2 has loyal customers, he has incentive to charge a higher price. We know that at $\theta=0$, there are no loyals and $p_i = 0$. As θ increases, the firm with more loyal customers has an incentive to charge higher expected prices.

More formally, we show that $F_2(p) \leq F_1(p) \forall p$.

Lemma 4: $F_2(p) \leq F_1(p) \forall p$.

Proof of Lemma 4:

From Table 2, this is obvious in regions $p < \underline{p}_1$, $\underline{p}_1 \leq p < \bar{p}_2 - c$, $\bar{p}_2 - c \leq p < \underline{p}_2$, $\bar{p}_1 \leq p < \underline{p}_1 + c$, $\underline{p}_1 + c \leq p < \bar{p}_2$, $p \leq \bar{p}_2$. This leaves region $\underline{p}_2 \leq p < \bar{p}_1$. We need to show that:

$$\frac{p(1+\gamma+\theta) - V_2(1-\delta)}{p(1+\gamma) + \delta(V_2 - V_1)} \geq \frac{p(1+\gamma) + \delta V_2 - V_1}{p(1+\gamma) + \delta(V_2 - V_1)}$$

This reduces to:

$$p\theta \geq V_2 - V_1$$

Substituting:

$$\frac{p(1+\gamma) + \delta V_2 - V_1}{p(1+\gamma) + \delta(V_2 - V_1)}$$

To show this for all p , we show it for \underline{p}_2 . Substituting and rearranging:

$$\underline{p}_2 \left[\theta + \frac{(1+\gamma)}{(1+\delta)} \right] \geq \frac{\theta(\underline{p}_1 + c)}{(1+\delta)} \quad (38)$$

To show this, we recognize that $F_1(\underline{p}_2) > 0$. This implies:

$$\underline{p}_2(1+\gamma+\theta) > V_2(1-\delta)$$

Substituting for V_2 , and rearranging terms, one can show that inequality 38 holds. Therefore we have shown that $F_2(p) \leq F_1(p)$.

QED

Proposition 1 follows immediately having proved Lemma 4.

QED

Proof of Proposition 2:

Comparative statics in Region C are straight forward:

$$V_1 = \underline{p}(1+\gamma+\delta\theta)/(1-\delta) \quad (39)$$

$$V_2 = \underline{p}(1+\gamma+\theta)/(1-\delta) \quad (40)$$

$$\underline{p} = \theta/[1+\gamma+\theta + \delta\theta] \quad (41)$$

$$\frac{\partial \underline{p}}{\partial \theta} = \frac{(1+\gamma)}{(1+\gamma+\theta(1+\delta))^2} > 0, \quad \frac{\partial \underline{p}}{\partial c} = 0, \quad \frac{\partial \underline{p}}{\partial \gamma} = \frac{-\theta}{(1+\gamma+\theta(1+\delta))^2} < 0, \quad ,$$

$$\frac{\partial \underline{p}}{\partial \delta} = \frac{-\theta^2}{(1+\gamma+\theta(1+\delta))^2} < 0$$

$$\frac{\partial V_i}{\partial c} = 0$$

$$\frac{\partial V_1}{\partial \theta} = \frac{\partial \underline{p}}{\partial \theta} \frac{(1+\gamma+\delta\theta)}{(1-\delta)} + \frac{\delta \underline{p}}{(1-\delta)} > 0 \text{ since all terms are positive.}$$

$$\frac{\partial V_1}{\partial \gamma} = \frac{\partial \underline{p}}{\partial \gamma} \frac{(1+\gamma+\delta\theta)}{(1-\delta)} + \frac{\underline{p}}{(1-\delta)} = K\theta^2 > 0 \text{ where } K > 0 \text{ is a constant.}$$

$$\frac{\partial V_1}{\partial \delta} = \frac{\partial \underline{p}}{\partial \delta} \frac{(1+\gamma+\delta\theta)}{(1-\delta)} + \frac{\underline{p}(1+\gamma+\theta)}{(1-\delta)^2}$$

dropping the positive terms, this reduces to

$$\left(\frac{-\theta^2(1+\gamma+\delta\theta)}{(1+\gamma+\theta(1+\delta))} + \theta \frac{1+\gamma+\theta}{1-\delta} \right) \text{ we drop the } (1-\delta) \text{ term as it blows up the right term.}$$

Collecting terms, we get:

$$(-\theta^2(1+\gamma+\delta\theta) + \theta(1+\gamma+\theta)(1+\gamma+\theta+\delta\theta)) > (1+\gamma+\theta)[\theta(1+\gamma)] > 0.$$

$$\frac{\partial V_2}{\partial \theta} = \frac{\partial \underline{p}}{\partial \theta} \frac{(1+\gamma+\theta)}{(1-\delta)} + \frac{\delta \underline{p}}{(1-\delta)} > 0 \text{ since all terms are positive.}$$

$$\frac{\partial V_2}{\partial \gamma} = \frac{\partial \underline{p}}{\partial \gamma} \frac{(1+\gamma+\theta)}{(1-\delta)} + \frac{\underline{p}}{(1-\delta)} = K\delta\theta^2 > 0 \text{ where } K > 0 \text{ is a constant.}$$

$$\frac{\partial V_2}{\partial \delta} = (1 + \gamma + \theta) \left(\frac{\partial p}{\partial \delta} \frac{1}{(1 - \delta)} + \frac{p}{(1 - \delta)^2} \right)$$

dropping the positive terms, this reduces to

$$\left(\frac{-\theta^2}{(1 + \gamma + \theta(1 + \delta))} + \theta \right) > 0 \text{ since } \theta > \theta^2 \text{ } (\theta < 1) \text{ and } (1 + \gamma + \theta(1 + \delta)) > 1.$$

For regions A and B, one can use the implicit function theorem on the system of equations.

References

- Beggs, Alan and Paul Klemperer (1992), "Multi-Period Competition with Switching Costs," *Econometrica*, Vol. 60, No. 3, May, pp. 651-666.
- Carpenter, Gregory S. and Kent Nakamoto (1989), "Consumer Preference Formation and Pioneering Advantage," *Journal of Marketing Research*, August, pp. 295-298.
- Farrell, Joseph and Carl Shapiro (1988), "Dynamic Competition with Switching Costs," *RAND Journal of Economics*, Vol. 19, No. 1, pp. 123-137.
- Lal, Rajiv, (1990), "Manufacturer Trade Deals and Retail Price Promotions," *Journal of Marketing Research*, XXVII (November), pp. 428-444.
- Lal, Rajiv, John D.C. Little, and Miguel J. Villas-Boas, (1993), "A Theory of Forward Buying, Merchandising, and Trade Deals," Working Paper, August.
- Narasimhan, Chakravarthi (1988) "Competitive Promotional Strategies," *Journal of Business*, Vol. 61, no. 4, pp. 427-449.
- Raju, Jagmohan S., V. Srinivasan, and Rajiv Lal (1990), "Effectiveness of Brand Loyalty on Competitive Price Promotional Strategies," *Management Science*, 36 (March), pp. 274-304.
- Rao, Ram C. (1991), "Pricing and Promotions in Asymmetric Duopolies," *Marketing Science*, 10 (Spring), pp. 131-144.
- Treacy, Michael and Fred Wiersema, (1993), "Customer Intimacy and Other Value Disciplines," *Harvard Business Review*, Jan-Feb, pp. 8-93.
- Rosenthal, Robert W. (1982), "A Dynamic Model of Duopoly with Customer Loyalties," *Journal of Economic Theory*, 27, pp. 69-76.
- Shilony, Y., (1977), "Mixed pricing in Oligopoly," *Journal of Economic Theory*, April, 14, pp. 373-388.
- Simester, Duncan, (1994), "Analytic Essays in Marketing," Unpublished Ph.D. Dissertation.

Varian, H. (1980), "A Model of Sales," *American Economic Review*, September, pp. 651-659.

Villas-Boas, Miguel J. (1993), "Predicting Advertising Pulsing Policies in an Oligopoly: A Model and Empirical Test," *Marketing Science*, Winter, Vol 12, No. 1, pp. 88-102.

Wernerfelt, Birger, "Brand Loyalty and Market Equilibrium" (1991), *Marketing Science*, Summer, pp. 229-245.

Essay 2

Why Advertise Both Regular and Sale Prices?

Introduction

Browse through the advertisements and weekly store fliers in your Sunday paper and you will find numerous items on sale. Take a closer look, and you may also notice that stores typically advertise sales in different formats. Two particularly common selling formats are what we will call the sale price only format and the regular/sale price format. Typical examples of the sale price only format and regular/sale price format are:

Sale Price Only

Pert Shampoo, 15oz., \$3 ²⁹

Panasonic Microwave, \$149 ⁹⁷

Regular/Sale Price

Extra Strength Tylenol, 100 cnt, \$6 ⁶⁶

Our Regular \$8 ⁶⁷

Brother Sewing Machine, \$119 ⁹⁹

Reg. \$139 ⁹⁹

As you continue browsing through your Sunday paper, you will notice that different retailers use different selling formats, and sometimes the same retailer may employ different formats for different product categories. This motivates one to ask the following questions: Why would a retailer advertise both regular and sale prices? If the consumer only pays the sale price, how could the regular price provide any useful information? One plausible explanation in the behavioral literature on reference prices is that including a regular price may create transaction utility (Thaler (1985)) and is therefore beneficial. However, if advertising the regular price is beneficial, why don't retailers always include the regular price? The explanation we offer is that a retailer may choose to include a regular price in an advertisement to signal information to uninformed consumers. We argue that the unknown information which is signaled is the retailer's relative price. We highlight the intuition in two separate models. In one model, consumers buy today but are allowed to search across competing retailers. In this model, the "relative" price that is signaled is with respect to competing retailers. In the second model, consumers buy from one retailer, but are willing to wait for a low promotional price. In this model, the "relative" price that is signaled is with respect to future sale prices at that retailer.

Consider some simple examples to clarify the intuition. Suppose it is spring, and you have decided to learn the game of tennis. You need to buy a tennis racket soon, and you learn from an advertisement that a local store has a promotion on tennis rackets. As it is spring, you believe that many stores are currently having tennis racket sales. Since you want to find the lowest priced racket, the question you

may ask is: "Should I buy the racket from the local retailer or search for a lower priced racket?" If you are uninformed about competitive prices for tennis rackets, you may not be able to immediately answer this question. However, you know that if your local store has low regular prices on rackets, they also have low promotional prices on rackets. We show that by including a low regular price in the advertisement, the local retailer can provide information to you indicating that he has a lower sale price on rackets than the competition. Not all stores will advertise the regular price. If the local store has high regular prices, posting this information in the advertisement will influence you to search at competing retailers. This intuition will be highlighted in our first of two models we label the Search Model.

To motivate the second model, consider another example. Suppose you are in the market for a new TV. You know that you are going to buy the TV from your favorite local store, but you are in no hurry to buy the TV today. You know that periodically TV's go on sale and that you can get a good deal if you wait. When a sale comes along, you have to decide whether to buy at the sale price or to hold out for the next, hopefully better, promotion. Suppose you see an advertisement with just the promotional price of \$900. Since you haven't bought a TV in a while and don't know the regular price of a TV at your local store you are not sure how "low" this price is. If you learn that the regular price is \$950, you may decide this isn't a "good deal" and you may wait for a better deal (lower price). Knowing this, the retailer has incentive to just advertise the sale price. In contrast, if you learn that the regular price is \$1200, you may think the price is a great deal and buy today. So the retailer with the higher regular price has incentive to include this information in the advertisement. We highlight this intuition in our Wait Model.

A further point about the Wait Model. Many people's intuition is that regular prices inform consumers about how much they are "saving." This is consistent with Thaler's reference price theory in which a consumer gets transaction utility from the gap between the regular price and the sale price. In our model, this gap is important, but for a different reason. We demonstrate that the gap provides information about the expected probability of getting a lower price in the future. More consumers purchase when the gap is large because they believe that the probability of a lower sale price is small. Thus, our model offers an alternative rationalization of reference prices.

We began each example assuming that you were uninformed about the regular price at your local store. Obviously, if everyone knows the regular price (i.e. consumers are informed), advertising the regular price is pointless. In our model, we

assume that some consumers are informed while others are uninformed¹. Some examples help illustrate why consumers may be uninformed. If a consumer is purchasing a durable good with long interpurchase times (e.g. a new TV), this consumer may not be informed about current retail prices. Similarly, a customer that is new to the market (e.g. the novice tennis player), may have limited information about retail prices. Finally, conditional on a store's price image, a consumer may have residual uncertainty about prices. For example, a consumer may shop at mass merchandisers that have a low price image. Within this class of retailers, the consumer may have uncertainty about regular prices.

Our main results are that the incentive to signal information by including the regular price in an advertisement increases with the proportion of uninformed consumers. When consumers primarily engage in searching behavior, the firm with low regular prices has incentive to include this information in the advertisement. When consumers are loyal to a firm and are waiting for promotions, the firm with high regular prices has incentive to include this information in the advertisement.

Finally, any discussion of advertised regular prices immediately raises the issue of credibility. What prevents retailers from manipulating prices and deceiving consumers? In the United States, the Uniform Deceptive Trade Practices Act (1966) prohibits retailers from manipulating regular prices. The Act commits a retailer to a regular price as it requires them to demonstrate that a regular price was charged during the normal course of business for a reasonable period of time². Thus a regular price carries information to the extent a retailer is legally obligated to charge that price in future periods. We will assume that a retailer that behaves deceptively is punished to the extent that they are legally obligated to charge posted regular prices.³

In summary, this paper argues that advertising a regular price is a way for retailers to signal valuable information to deal prone consumers about the relative

¹Uninformed consumers lack information to the extent that they don't know whether a retailer has high or low prices for an item. However, given a regular price, an uninformed consumer will be able to infer whether that price is high or low. In the TV example, if the consumer sees a regular price of \$950 they know that this is a low regular price.

²Section 2.a.11 of the Uniform Deceptive Trade Practices Act states: "A person engages in a deceptive trade practice when, in the course of his business, vocation, or occupation, he makes false or misleading statements of fact concerning reasons for, existence of, or amounts of price reductions."

³In our model, we will rely on this legal restriction. However, there are other reasons outside our model as to why it may not be in a firm's best interest to behave deceptively. For example, a firm with a low price image may not want to deceptively announce a high regular price as this erodes the firm's price image. Similarly, if a firm has positioned itself as high price/quality, posting a deceptively low regular price would be costly to the firm's reputation.

value of today's sale price. Our main result in both the Search Model and the Wait Model is that, *ceteris paribus*, the likelihood of a retailer using the regular/sale price selling format increases with the proportion of deal prone consumers. In the Search Model, where deal prone consumers search across retailers but cannot postpone a purchase, a retailer has incentive to post a low regular price. Retailers will not advertise high regular prices as this encourages customers to search at competing retailers. In the Wait Model, where deal prone consumers don't search across stores but can postpone a purchase, a retailer has incentive to advertise a high regular price to signal that lower future prices are unlikely. Retailers will not advertise low regular prices as the small gap between the regular and sale price will influence customers to wait for a lower sale price in future periods. We find empirical support for the main result across several product categories at different retailers.

We organize the paper around the two models as follows: Section 1 reviews related literature. Section 2 develops the general framework of the models. Section 3 develops the Search Model and presents the main result of the paper, Section 4 develops the Wait Model and related results. Section 5 presents Empirical Evidence. Section 6 concludes the paper with a discussion and brief summary.

Section 1: Related Literature

This paper combines elements from multiple streams of research: the literature on selling formats, reference prices, and price promotions. Retail selling formats is a research stream in marketing that is receiving increasing interest and encompasses work that looks at a more fundamental level of firm decisions. For example, rather than looking at an issue such as size of a sales force, Anderson (1985) considers what type of sales force a firm uses. Anderson compares a firm's incentives to use either a direct sales force or manufacturer representatives. McGuire and Staelin (1983) address the issue of retail integration and suggest conditions under which a retailer prefers an integrated versus a non-integrated channel relationship. Recent work by Wernerfelt (1994a) extends our understanding of the comparative performance of different retail selling formats when consumers incur search costs. Wernerfelt rationalizes common retail formats such as price advertising, co-location, and bargaining. Finally, in order to rationalize many common selling formats, much research in marketing and economics has examined the practice of signaling unobserved information with prices, advertising spending, and money back guarantees⁴. Related work of particular interest is Simester (1994) who examines how firms can use advertised prices to signal price image of the store. While this literature has not explicitly considered the implications of alternative retailer selling formats, it is relevant to the research stream because implicitly some of these authors provide an explanation for the practice of advertising and branding. In general, the use of alternative selling formats is attracting more and more attention in marketing (Wernerfelt (1994b)). This paper contributes to that literature by evaluating the regular/sale price format.

The second stream of research this paper contributes to is the literature on reference prices. Following work on prospect theory by Kahneman and Tversky

⁴Beginning with Klein and Leffler (1981), researchers have shown that a price premium may serve as a credible signal of high product quality. (see also Shapiro (1983)) Motivated by Nelson (1974), researchers have also examined advertising spending as a signal of product quality. Papers in this area include Schmalensee (1978), Kihlstrom and Riordan (1984), and Milgrom and Roberts (1986). These papers show that when product quality is not perfectly observed by all consumers, advertising spending may carry information about product quality. Higher advertising spending is generally associated with higher product quality. Related work by Montgomery and Wernerfelt (1992) and Wernerfelt (1988) offers an alternative explanation of the price/quality relationship and demonstrates that umbrella branded (high quality) products may have lower average prices. Thus depending on the market, high quality may be signaled with either high or low prices. Moorthy and Srinivasan (1994) look at the practice of signaling information using money back guarantees.

(1979) and later work on mental accounting by Thaler (1985), reference prices have received increased attention in marketing. A premise of this work is that customers don't evaluate alternatives in isolation but incorporate a reference point. Thaler's work on mental accounting suggests that when a consumer purchases a product there is both consumption utility (from the product itself) and transaction utility (from the purchase itself). In the spirit of prospect theory, transaction utility is evaluated relative to a reference point. Winer (1986) empirically tests choice models which include reference points. Winer demonstrates the improved fit of reference point choice models over traditional models of choice. Experimental research by Urbany, Bearden, and Weilbaker (1988) looks at some of the implications of incorporating reference prices in advertisements. Their results suggest that consumers believe that the value of a product is increasing with the reference price and that consumers search less when a reference price is included. A more thorough description of important findings in this stream of research can be found in Monroe (1990). Our paper is complementary to the current stream of research on reference prices as we rationalize the use of a type of reference price by consumers.

The final stream of research this paper contributes to is the literature on price promotions. Much of the research on promotions focuses on firms who sell to heterogeneous consumers. The term "deal prone" has been broadly used to identify those consumers who are more sensitive to promotions. Research has focused on better understanding properties of deal prone consumers such as response characteristics (Webster (1965), Narasimhan (1984), Bawa and Shoemaker (1987)) and demographic characteristics (Blattberg, Buesing, Peacock, and Sen, (1978)). Similar research has considered the issue of segmentation (Blattberg and Sen (1974,1976)).⁵ Other research in this stream examines optimal promotional strategies for efficient selling to heterogeneous consumers. These papers typically include segments of consumers that vary in their brand loyalty, reservation price, price sensitivity, income level or some other dimension. Papers include Gerstner and Hess (1991), Lal (1990), Narasimhan (1988), Raju, Srinivasan, and Lal (1990), and Banks and Moorthy (1994). When analyzing price promotions, these papers take the sale price only format as given. Here we first ask if the seller can do better by using the regular/sale price format and next ask how heterogeneity impacts this question.

In summary, this paper contributes to the literature on comparative selling formats by evaluating the implications of bargain hunting (searching and waiting for

⁵An extensive review of this literature can be found in Blattberg and Neslin (1990).

lower prices) on the part of consumers. We also contribute to the literature on reference prices by offering a rational explanation for their use.

Section 2: General Model

To highlight each aspect of our intuition more clearly, we develop two separate but related models that we label the Search Model and the Wait Model. The purpose of this section is to lay out the general framework which is common to both models.

In each model, we consider a single product sold in multiple periods in a channel consisting of a single manufacturer, a retail monopolist⁶, and heterogeneous consumers. For reasons outside the model, the manufacturer provides incentive for the retailer to lower the retail price by offering trade deals in alternating periods⁷. The current period, Period 1, will have a trade deal and the next period, Period 2, will not have a trade deal. The timing of trade deals is known to both retailers and consumers. There is a single retailer who has constant marginal cost, c_i , which for exogenous reasons may be either low or high (type L or type H, where $c_L < c_H$). Consumers cannot directly observe c_i but have beliefs that each store type is equally likely.

The trade deal reduces the retailer's marginal cost of each item sold⁸ by an amount D . In a given period, each retailer in the market receives the same D . Across sale periods, the size of the trade deal varies. Retailers learn the size of D before setting retail price. Consumers do not directly observe D but have prior beliefs that D is drawn from a commonly known cumulative distribution function $K(D)$ which we assume is $U(0,1)$ ⁹.

In each model there are two segments of consumers in proportions $(1-\alpha)$ and α . The $(1-\alpha)$ segment of consumers are in the market every period and will buy if the price is below their reservation price. For ease of exposition, we refer to these consumers as a retailer's "loyal" consumers. The $(1-\alpha)$ mass of loyal consumers are indexed by their taste address $m \in [0,1]$. A consumer located at m has utility $U(p,m)$

⁶The purpose of this model is to explain the existence of a particular selling format in the simplest possible model. The Search Model has passive "competition" that is characterized as an outside option. In a later paper we hope to more directly address more strategic competition.

⁷Possible reasons to offer trade deals include heterogeneous customer reservation prices or price elasticities. One explanation for the variation in trade deal size may be manufacturer response to stochastic inventory fluctuations.

⁸We assume the retailer cannot inventory items so that items sold in a period equal the number of items purchased from the manufacturer in the same period.

⁹It is possible to simplify this to $K(D)$ binomial (large or small trade deal). Without further assumptions, the Wait Model then becomes a knife edge case. For consistency and robustness, we will consider $K(D)$ uniform.

$= a - bp - m$ where a and b are constants and p is the price. A loyal consumer will purchase one unit if $U(p,m)$ is greater than or equal to their outside option utility which is normalized to zero. If we assume that m is uniformly distributed then loyal consumers buy $(1-\alpha)y = (1-\alpha)(a-bp)$ units per period. The retailer also sells to α deal prone consumers that are more price sensitive and only buy when an item is on sale. Assumptions about deal prone consumer behavior will vary in the Search Model and the Wait Model. In both models, deal prone consumers are heterogeneous and aggregate demand is downward sloping.

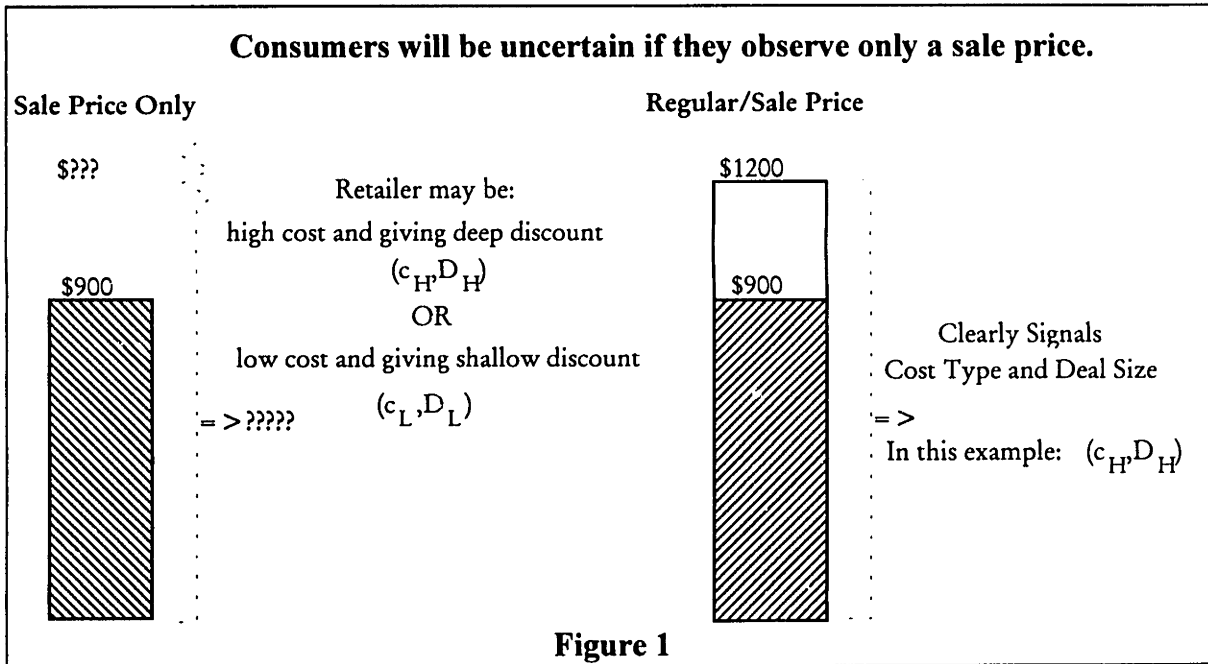
In both the Search Model and the Wait Model, retailers will take into consideration additional costs associated with posting a regular price such as loss of flexibility, execution costs, and opportunity costs. If there are unexpected shocks in demand, posting a regular price reduces a retailer's flexibility in setting the regular price when legal constraints lock a retailer in to the advertised regular price¹⁰. While not part of these models, in a practical setting flexibility may be lost in large chain stores because all retailers may not face the same demand conditions. Advertising a regular price may also create additional execution costs for the retailer. These range from additional legal constraints which must be met to the hassle of identifying the regular price. Finally, advertising a regular price takes up more physical space in an advertisement which creates an opportunity cost. We let F_i equal the additional cost of the regular/sale price format at retailer type i ¹¹. These additional costs will be of concern to the retailer and will impact the decision of whether to advertise in the regular/sale price format.

The unknown information to be signaled to uninformed consumers is the retailer's cost type and the size of the trade deal. A consumer that observes only the sale price is unable to distinguish a (high cost type, large trade deal) retailer from a (low cost type, small trade deal) retailer. (see Figure 1) Returning to our TV example, suppose you see a TV advertised on sale for \$900. You know that a low regular price

¹⁰In the United States, the Uniform Deceptive Trade Practices Act (1966) prohibits retailers from manipulating regular prices. The act commits a retailer to a regular price as it requires them to demonstrate that a regular price was charged during the normal course of business for a reasonable period of time. Section 2.a.11 of the Uniform Deceptive Trade Practices Act states: "A person engages in a deceptive trade practice when, in the course of his business, vocation, or occupation, he makes false or misleading statements of fact concerning reasons for, existence of, or amounts of price reductions."

¹¹ For the remainder of the paper, we follow the flexibility interpretation. That is, an unexpected occurrence may occur such that altering the regular price increases profits. We let $F_i = \rho(1-\alpha)^{\alpha}I_i$ where ρ is the probability of a stochastic shock, a is a constant, α is the proportion of deal prone consumers, and I_i is the incremental profit that could be earned by retailer i if the regular price could be adjusted. See the Appendix for further discussion.

for the item is \$950 and a high regular price is \$1200. After observing only the sale price of \$900, you are unable to tell whether you are at a store with a \$50 discount or a \$300 discount. In each model (Search and Wait), we will explain what information retailers have incentive to signal.



Both the Search Model and the Wait Model contain this same basic structure. The main difference between the models will be the outside option available to deal prone consumers (i.e., search or wait). As there are two unknown parameters (c, D), in a general model signaling can get a bit confusing. But in each model that we consider, either c or D is fixed. In the Search Model, all retailers get the same trade deal, D , so the focus is on signaling cost type. In the Wait Model, consumers only shop at one retailer, c is fixed, so the focus of signaling is on D . In the discussion, we note what happens in a general formulation of both searching and waiting. We now lay out the model specific assumptions and the main results of each model.

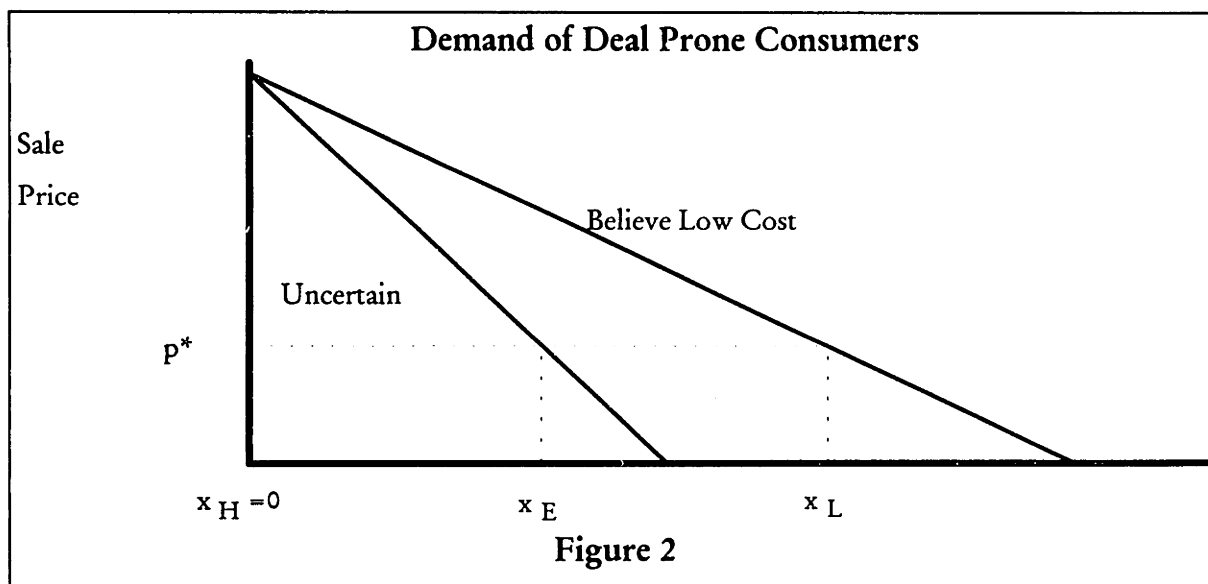
Section 3: Search Model

Model Assumptions: Search Model

The key characteristic of the Search Model is that the mass of α deal prone consumers is willing to search across stores to find the lowest price but cannot postpone a purchase until the next sale period. (e.g. novice tennis player example) We model the ability to search as a stochastic outside option that one can think of as representing passive competition in the market. The outside option is stochastic because the manufacturer trade deal is not observable by consumers. Further, uninformed consumers do not know the current price at the outside option (e.g. through advertising). For example, the outside option may be a retailer known to have moderate prices (i.e. moderate cost type). In the tennis racket example, the novice player must decide whether to buy a racket from the local store at a known price or to buy from the moderate price retailer.

In each sale period we assume¹² that aggregate deal prone consumer demand at a known low cost retailer with price p is $\alpha x_L = \alpha(v_L - w_L p)$ where v_L and w_L are constants. Demand is zero if a deal prone consumer believes a retailer has high marginal costs as he either exercises the outside option or does not buy. Recall that we assume that before observing the advertised price, deal prone consumers have prior beliefs that low and high type stores are equally likely. After observing a retailer's prices, deal prone consumers that are still not sure whether a retailer has high or low marginal costs have expected demand $\alpha x_E = \alpha(v_E - w_E p)$ where p is price, and v_E and w_E are constants such that $x_E(p) < x_L(p)$. (Figure 2)

¹² Similar to loyal demand, one can rationalize these assumptions by assuming that utility is linear in price, and consumers tastes are distributed $U(0,1)$. We formalize this in the appendix.



If both types of retailers follow the proposed equilibrium strategies, consumers posterior beliefs are clearly specified. However, if a retailer announces an out of equilibrium price, consumers have prior probability zero of observing such a price and posterior beliefs are not specified. To completely define an equilibrium we need to specify consumer posterior beliefs when they observe out of equilibrium prices. We apply the intuitive criterion¹³ (IC) (Cho and Kreps 1987) to pin down out of equilibrium beliefs. The intuitive criterion allows one to eliminate specific out of equilibrium strategies for certain types of retailers. To understand how the criterion applies to this model, first fix an equilibrium strategy p_i^* and equilibrium profit π_i^* for retailer i . Assume that consumers have posterior beliefs, θ , that a store is the low cost type ($\theta=1$) after a deviation p^{**} by that store. Given that $\theta=1$, if the profit to type i of the proposed deviation is less than the equilibrium profit ($\pi_i^* > \pi_i(p^{**} | \theta=1)$), then the criterion suggests that consumers conclude that type i could not have deviated to strategy p^{**} . If for some p^{**} only a single type is ruled out, posterior beliefs are either 1 or 0. If both types or neither type are ruled out, the IC puts no restrictions on out-of-equilibrium beliefs. We assume consumer beliefs revert to their priors. The equilibrium fails the IC if there exists a p^{**} such that the profit of type i , given posterior beliefs θ (that are consistent with IC), is greater than the equilibrium profit.

We now consider a retailer that discounts future profits with discount factor $\delta_R < 1$. As mentioned earlier, a manufacturer provides a periodic trade deal of size D that reduces a retailer's constant marginal cost, c_i . Under full information, demand

¹³For a summary, see Fudenberg and Tirole (1991), Chapter 11.

and profits per period are summarized in Table 1. We use the superscript notation to refer to periods and subscript notation to refer to types (e.g. π_i^1 = period 1 profit at retailer type i).

<u>Summary of Full Information Demand and Profit</u>	
<u>Period 1: Sale Period</u>	
Loyal consumer Demand	$(1-\alpha)y^1 = (1-\alpha)[a-bp]$
Deal prone consumer Demand ¹⁴	$\alpha x_i = \alpha[v_i - w_i p]$
Profit	$\pi_i^1 = (\alpha x_i + (1-\alpha)y^1)(p - c_i + D)$
<u>Period 2: Non Sale Period</u>	
Loyal consumer Demand	$(1-\alpha)y^2 = (1-\alpha)[a-bp]$
Profit	$\pi_i^2 = (1-\alpha)y^2(p - c_i)$

Table 1

The retailer's strategy consists of choosing a selling format and prices each period. We only consider pure strategies for the retailer. A type i retailer that chooses the regular/sale price format announces a price pair (s_i, r_i) in period 1 where s_i is the period 1 sale price and r_i is the period 2 regular price. In period 2, legal restrictions commit the retailer to the regular price. If a type i retailer chooses the sale price only format, he announces a sale price p_i^1 in period 1. In period 2 he announces a non-sale price p_i^2 . We summarize this strategy as (p_i^1, p_i^2) .

Total retail profit is a function of announced prices as well as posterior consumer beliefs, θ , about the probability a retailer is the low cost type. We use the following notation for total profit as a function of the price each period and posterior consumer beliefs:

Regular/Sale Price Format

$$\pi_i(p_i, r_i, \theta) = \pi_i^1(s_i, \theta) + \delta_R \pi_i^2(r_i) - F_i$$

Sale Price Only Format

$$\pi_i(p_i^1, p_i^2, \theta) = \pi_i^1(p_i^1, \theta) + \delta_R \pi_i^2(p_i^2)$$

¹⁴ $x_{iH} = 0$.

Note that θ only appears in the period 1 profit function because in period 2 all consumers that buy are loyal consumers and don't care about the stores cost type. If conditions were such that consumers had a priori full information and $F_i=0$, the selling formats are equivalent since prices carry no information. The low type full information profits are $\pi_L(s_L, r_L, 1) = \pi_L(p_L^1, p_L^2, 1)$ and the high type full information profits are $\pi_H(s_H, r_H, 0) = \pi_H(p_H^1, p_H^2, 0)$.

To summarize the game, we consider a retailer selling a product in two periods. In period 1, all consumers know that there is a sale (trade deal present) and that the trade deal is drawn from $K(D)$. In period 2, there is no sale. A mass of α deal prone consumers either purchase at the sale price, exercise an outside option, or don't buy at all. In both Period 1 and Period 2, a $(1-\alpha)$ mass of loyal consumers decides whether to purchase the item based only on the price. The retailer maximizes profits by choosing the optimal sale price, the optimal regular price, and the optimal selling format.

Results: Search Model

In the general formulation of this model, many equilibria are possible. We first consider a model with no future uncertainty, $F_i=0$, and describe equilibria that highlight the signaling and commitment properties of the regular/sale price format. The signaling proposition shows the benefit to the low cost type of advertising in the regular/sale price format. The commitment proposition shows the cost of deception if the high cost type advertises a low regular price. Next, we focus on a specific equilibrium and describe the relationship between the proportion of deal prone consumers, α , and the retailer's optimal selling format when $F_i > 0$.

As shown in Figure 1, a consumer that observes only a sale price cannot determine whether a retailer is a high cost type or low cost type retailer. Since $x_L > x_E > x_H$, uncertainty leads to incremental benefit to the high cost store and incremental loss to the low cost store. Therefore, a low cost type retailer prefers to signal his low sale price by including a low regular price. The high type does not want to advertise a high regular price as this will cause deal prone consumers to search at competing retailers for a lower price.

Proposition 1a: Signaling Property of Regular/Sale Price Format

Let $F_i=0$. The low cost type weakly prefers advertising both a regular price and a sale price.

Proof: See Appendix.

Proposition 1 highlights the signaling property of the regular/sale price format. Since $F_1=0$, it can never hurt to advertise the regular price. The benefit of including a regular price in the advertisement is that it may signal the retailer's low cost type.

The next question to address is how the low cost type retailer will distinguish himself from the high cost type retailer. Perhaps the retailer doesn't have to do anything other than advertise prices. When advertising the "normal" prices uniquely identifies the low cost store's type, we refer to this as a cheap separating equilibrium¹⁵. Alternatively, the low cost store may need to decrease prices to signal that he is the low cost type. When equilibrium prices are distorted from full information prices, we refer to this as a "costly" separating equilibrium.

It should be clear how a low cost retailer that advertises a low regular price can separate itself from the high cost retailer. For example, in a cheap separating equilibrium customers see a low regular price and a low sale price and then infer that the store is a low cost type. Can a retailer separate by posting just a sale price?

We have already shown in Figure 1 how posting a sale price will lead to consumer uncertainty. Suppose a retailer decides to lower his price even further to signal that he is a low cost type. Because D is unknown, this may not be an effective strategy. A consumer that sees a lower sale price may just infer that D is larger. Consumers are still unable to tell whether they are at a low cost type retailer with a slightly larger D or a high cost type retailer with a slightly larger D . In the Search Model, a clear signal is one where the consumer can infer that for all possible D , a store is definitely a low cost type retailer. Therefore, advertising only a sale price will generally be uninformative.

Despite this, there do exist parameters such that the low cost retailer can separate from the high cost store using only a sale price. For example, if the cost differences are large and the trade deal is large, a low cost store may be able to separate with only a sale price. We now want to consider such regions.

¹⁵ We use the terminology "cheap" because consumer uncertainty is costless to the retailer. The retailer's actions and profits are identical to those under full information.

Proposition 1b: Commitment Property of Regular/Sale Price Format

Let $F_i=0$. If signaling is possible in both formats, advertising a low regular price deters imitation by the high cost type retailer. In a costly separating equilibrium, the low cost type strictly prefers the regular/sale price format.

Proof: See Appendix.

Proposition 1a establishes that it is the low cost type retailer who wants to signal. In a cheap separating equilibrium, selling formats lead to equivalent prices and profits. In a costly separating equilibrium, the low cost store needs to lower prices to convince the consumer that he is truly a low cost type. Now the two selling formats are not equivalent. Because the regular/sale price format punishes the high cost type retailer via a commitment to a low future price there is less incentive for the high cost store to deviate. Therefore, there is less distortion in the low cost retailers prices and greater profits in the regular/sale price format relative to the sale price only format.

Propositions 1a and 1b show that when $F_i=0$, the low cost store has incentive to signal, and that the more effective format for signaling is the regular/sale price format. Given that a low cost type store is going to advertise both a regular and sale price, we now address the optimal way to signal in a costly separating equilibrium. The low cost retailer has two prices to manipulate which leads to three strategies: 1) lower the sale price; 2) lower the regular price; 3) lower both the regular and sale prices. Let (s_L^*, r_L) , (s_L, r_L^*) , and (s_L^*, r_L^*) be the respective prices in these strategies. This brings us to our second proposition:

Proposition 2: Efficient Separation in Regular/Sale Price Format

In the regular/sale price selling format, if the low cost store needs to lower prices to convince consumers of his true type the optimal strategy is to lower both prices.

Proof: See Appendix.

The intuition in the proof is as follows. When the low cost store is at the optimum full information price, s_L or r_L , small changes in price lead to no first order changes in profits. However, at $s_L^* \neq s_L$ or $r_L^* \neq r_L$ the low cost stores optimum price is distorted and a small change in price leads to a change in profits. To prevent imitation by the high cost store, the low cost store chooses some (s_L^*, r_L^*) which makes it undesirable for the high cost store to imitate the low cost store. If the low

cost type retailer were to discount only one price (sale or regular), the low cost type retailer could increase profits by slightly lowering the non-discounted price (with no change in profits) while slightly increasing the discounted price (increasing profits). Simester (1993) presents a similar finding. In his paper, firms are discounting multiple products to efficiently deliver surplus to consumers. Here firms are effectively discounting multiple products (today's product and tomorrow's product) to efficiently reduce the incentives to imitate.

Thus far we have outlined the signaling and commitment properties of the regular/sale price selling format under the assumption that there are no costs associated with the regular/sale price format (i.e., $F_i=0$). We have also demonstrated that low cost type retailers will distort both the regular and sale prices to prevent imitation by the high cost type store. If $F_i=0$, the model is less interesting because the regular/sale price format is a weakly dominant strategy for the low cost retailer for all α . When there are expected costs of using the regular/sale price format, the incremental benefit from signaling must be greater than these costs. We now consider $F_i>0$ and find the optimal selling format.

Proposition 3: Optimal Selling Format when $F_i>0$

a) Let the regular/sale price format have a positive expected cost, $F_i > 0$. For a given D , there exists $\underline{\alpha}$ and $\bar{\alpha}$ such that the sale price only format is optimal for $\alpha < \underline{\alpha}$ and the regular/sale price format is optimal for $\underline{\alpha} \leq \alpha < \bar{\alpha}$ for the low cost type retailer. The high cost type retailer uses the sale price only format.

b) In a pooling equilibrium, the optimal selling format for both retailers is the sale price only format.

Proof: See Appendix.

For low levels of α , signaling is less important because there are few deal prone consumers. In the limit, $\alpha=0$, all consumers are loyal and revenues are equal under both formats. For either retailer type, commitment to a future price makes the regular/sale price format less desirable because of future uncertainty. As the number of deal prone consumers increases, the incremental benefit of signaling the depth of discount may outweigh any expected costs. The low cost retailer may use the regular/sale price format when there are a large number of deal prone consumers. If the low cost type retailer can clearly signal ($\theta=1$) by posting a regular price, the high cost type retailer will only post the sale price. Therefore, as α increases, we should see more low cost type retailers advertising regular prices.

A pooling equilibrium may exist if deterring entry is too costly. If pooling is optimal, there is no information signaled with the regular price. Revenues under either format are equal and since $F_i > 0$, the optimal selling format is sale price only.

The reason the regular/sale price format may not be optimal for $\alpha > \bar{\alpha}$ is that in the limit $\alpha = 1$, no information is signaled by advertising the regular price. If there are no loyal consumers to purchase in Period 2, retailers can post any regular price at no commitment cost. Since no information is signaled, retailers will advertise in the sale price only format if $F_i > 0$. Even if we fix the number of loyal consumers, (e.g. mass of 1) and look at a positive mass of α consumers, we still have this problem in the limit as α approaches infinity.

We first note that this seems like a very technical problem of limited practical interest. Firms often go out of business if they have no loyal consumers. This implies that firms sell nothing in non-sale periods which seems unrealistic. We can handle this problem by relying on the flexibility interpretation of $F_i > 0$. That is, if there are stochastic shocks which create future uncertainty, F_i will be proportional to the size of the loyal consumers. If we assume that F_i converges to zero fast enough as α approaches 1, then $\bar{\alpha} = 1$ and the regular/sale price format is weakly dominant. We recognize this as a technical limitation of our model¹⁶. For practical considerations, we probably can consider a range of α bounded away from 1. In this paper we will focus on the equilibrium in Proposition 3 and $\bar{\alpha} = 1$ such that the regular/sale price format is optimal when α is large enough.

¹⁶ Another technical complication is proving that the difference in profits has a unique maximum. If advertising the optimal sale price only leads to beliefs $\theta = 1/2$, this is obvious by construction. To fully characterize the equilibrium, we must also consider parameters where the optimal sale price only leads to beliefs $\theta = 1$ and the only difference in profits is the commitment property (Proposition 2). This proof is in the appendix.

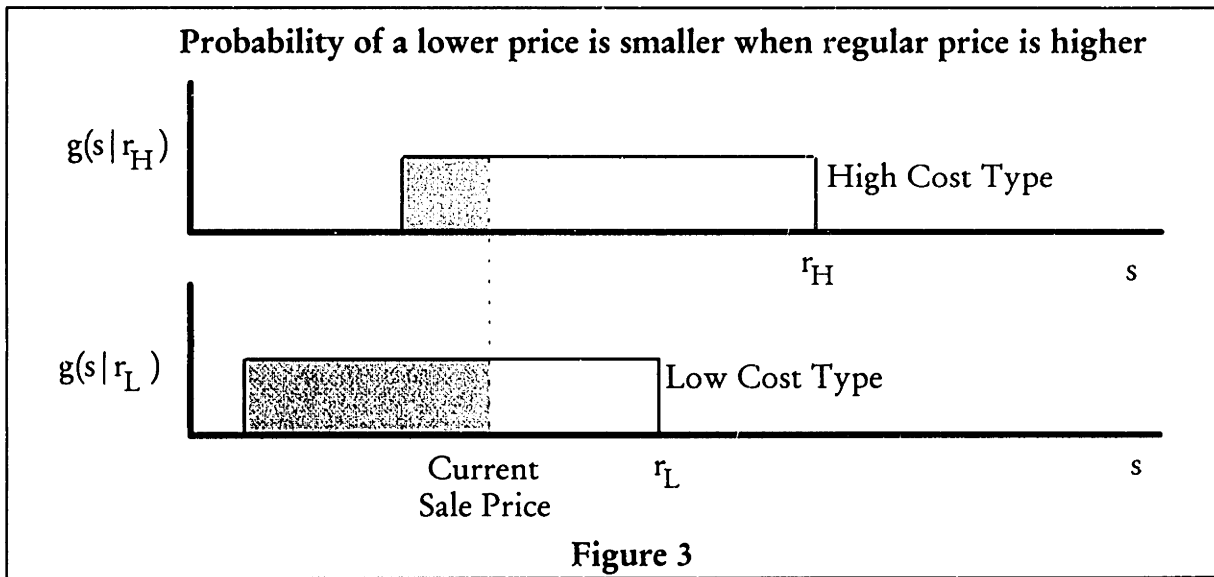
Section 4: Wait Model

Model Assumptions: Wait Model

In the Wait Model we assume deal prone consumers can time their purchase and wait for the optimal price cut but cannot search across retailers. We show that when deal prone consumers do not search across retailers for the lowest price there is incentive to advertise a high regular price to signal that the size of today's price cut is large. In our TV example, the retailer with a \$1200 regular price has incentive to advertise this information.

Before proceeding with mechanics, let us understand the intuition in another manner. Let $g(s)$ be the p.d.f. of sale prices and $g(s|r)$ the distribution of sale prices given the regular price. Fix the cost type, c_i , and consider two trade deal sizes, $D_H > D_L$. A deal prone consumer who infers that the current trade deal size is D_H is more likely to buy today. The motivation to buy today is due to the small probability of a lower future promotional price at that retailer. In Figure 3, $g(s|r_H)$ is small. A customer who is patient and infers that the current trade deal size is D_L is more likely wait until a future sale period. Waiting may be beneficial because there is a high probability of a lower promotional price. In Figure 3, $g(s|r_L)$ is large. Thus firms have incentive to signal that the trade deal size is larger and can do so by advertising a high regular price.

We now formalize the model in a manner consistent with profit maximization and utility maximization. As this gets somewhat complicated, the reader uninterested in the mechanics should focus on the intuition in Figure 3 and skip to the results section.



We consider the general model developed in Section 1. To accommodate deal prone consumers waiting until future sale periods we consider an infinitely repeated two period game. In odd periods there is a sale and in even periods there is no sale. As before, trade deals vary from sale period to sale period according to a commonly known distribution that we assume is $U(0,1)$.

In each period there are α deal prone consumers that are indexed by their taste address $n \in [0,1]$ ¹⁷. A deal prone consumer located at n has utility function $V^i(p,n)$ where $i \in \{L,H\}$ and p is the price. Deal prone consumers cannot search across retailers but are patient and can postpone a purchase until a future sale period. Each deal prone consumer will purchase exactly one unit of the product and must determine the optimal time to purchase. Deal prone consumers are willing to wait for a deep discount from the regular price, however, they prefer to consume sooner rather than later. When deal prone consumers observe a sale price, they trade off the probability of getting a lower price in the future versus the opportunity cost of postponing the purchase.

¹⁷To sustain this assumption we need to assume one of two things. Either bargain hunters leave as they buy and are replenished at exactly the right rate or they remain in the market and buy repeatedly but retain no memory of store cost types. This assumption makes it appear as if the mass of patient customers is identical from period to period. Thus the retailer and consumer problems are identical every sale period which simplifies calculations. In a more general model, a mass of very patient customers may accumulate over time. Some of these very patient customers may also learn the store's cost type. These factors will influence both the retailer's pricing decision and the customer's purchase decision. We recognize that these phenomenon exist but choose to eliminate them from the model. Future research may wish to address such issues.

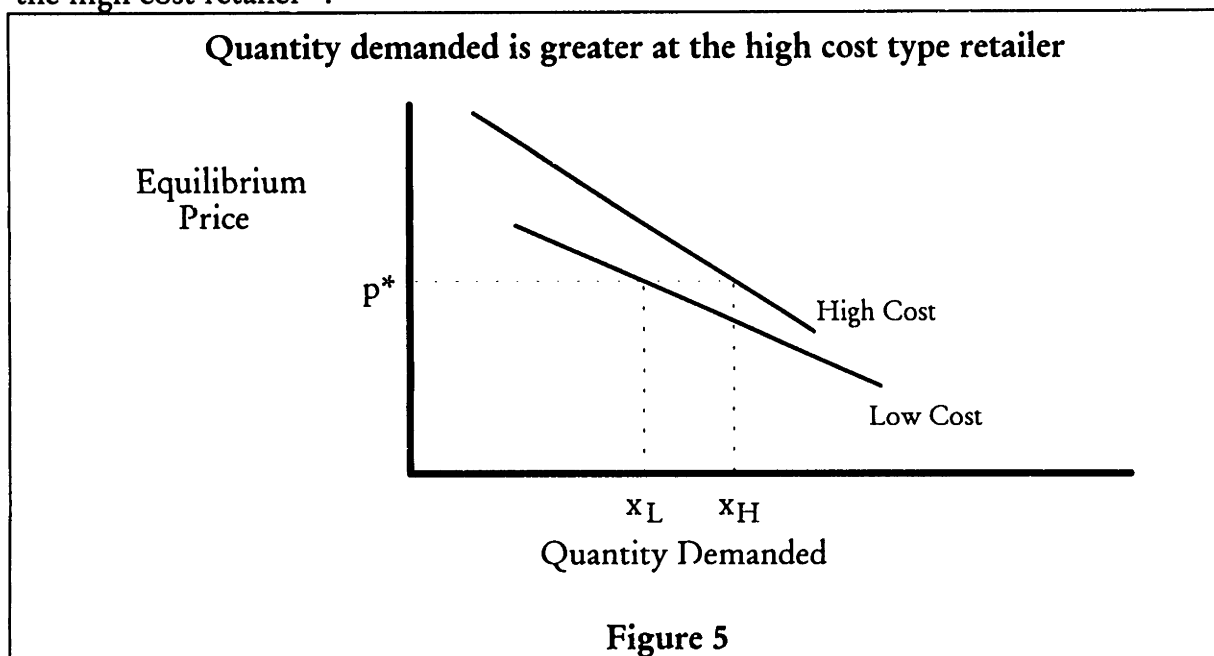
The desire to purchase earlier than later is captured in the deal prone consumer's discount factor, δ_n , assumed to be less than one. Deal prone consumers know how retailers set the optimal price based on the size of trade deals and costs $\phi(D|c_i) \equiv \phi_i(D)$ and they also know the distribution of trade deals, $K(D)$. When deal prone consumers observe a sale price, they can calculate the probability that a future sale price will be lower than the current price. This information allows each deal prone consumer to summarize the decision to buy today or wait for a better deal with a simple threshold price rule. A deal prone consumer will purchase today if the current price is lower than the threshold price otherwise he postpones the purchase. As deal prone consumers have heterogeneous tastes, each deal prone consumer will have a different threshold price rule. Furthermore, because each retailer type has different prices (under full information), deal prone consumers will have a different threshold price rule for each retailer type. We use the notation $p_i^t(n)$ for the threshold price of deal prone consumer n at retailer type i .

One may recognize that the basic trade off for deal prone consumers takes the form of a classical optimal stopping problem from dynamic programming. Bellman's equation captures this basic trade off between purchasing today and postponing a purchase. Assuming that the retail price function, $p = \phi_i(D)$, is linear and invertible, we can drop the (n) notation and let $D_i^t = \phi_i^{-1}(p_i^t)$ be the threshold deal size at retailer i . From Bellman's equation, the optimal threshold prices are defined by:

$$V^i(p_i^t, n) = \delta_n [K(D_i^t) V^i(p_i^t, n) + (1 - K(D_i^t)) E[V^i(p, n | p < p_i^t)]] \quad (1)$$

The term on the left hand side of equation 1 is the utility to deal prone consumer n from purchasing today from store i at price p_i^t . The term on the right hand side is the discounted expected utility of waiting until the next sale period. $K(D_i^t)$ is the probability that the next period sale price is greater than the threshold value. The term $(1 - K(D_i^t))$ is the probability that the next period sale price is less than the threshold value. $E[V^i(p, n | p < p_i^t)]$ is the expected utility to deal prone consumer n at store i given that the price is less than the threshold value. Deal prone consumers can solve equation 1 for p_i^t . For a given price at retailer i , deal prone consumers can infer the actual deal size using $D = h_i^{-1}(p)$. Some deal prone consumers will have a threshold deal size less than the actual deal size and will make a purchase. We can show there exists a deal prone consumer utility, $V^i(p, n)$, such that the optimal deal prone consumer demand at retailer i as a function of deal size is linear: $\alpha x_i =$

prefer the low cost retailer. At a given price, why is the quantity demanded greater at the high cost retailer²⁰?



To convince oneself that this makes intuitive sense, fix one price p^* for type L and type H and consider two quantities demanded $x_L < x_H$ with $D_L = h_L^{-1}(p^*)$ and $D_H = h_H^{-1}(p^*)$. Because $c_H > c_L$, if both retailer types charge the same price, the high cost retailer must have a larger trade deal $D_H > D_L$. At p^* , the quantity demanded is greater for the high type because the magnitude of the trade deal itself is larger $D_H > D_L$. Patient consumers know that a lower future sale price is more likely at retailer L, $[1-K(D_L)] > [1-K(D_H)]$, which creates an incentive to postpone their purchase. (Figure 3)

As we have not specified utility functions, it is possible for these derived demand functions to take many forms. We claim that for a given p^* , $x_L < x_H$ should be satisfied²¹. Let the set P equal the intersection of equilibrium prices at retailers L and H. For all $p \in P$, we require $x_H(p) \geq x_L(p)$. In words this means that the demand curves do not cross over the range of equilibrium prices. If $K(D)$ has infinite or very large support, this implies that the curves do not cross in the positive orthant of the (p,x) plane.

²⁰Be careful not to confuse this counter intuitive example with equilibrium demand. Under full information, $x_L(D) > x_H(D)$ as one would expect and as shown in Figure 2.

²¹If $E(s|r_H) > E(s|r_L)$ this must hold.

Thus far we have focused on deal prone consumers willingness to postpone a purchase. We have demonstrated that each of them will have a threshold price, p_i^t , that will trigger a purchase. Consumers observe advertised prices, but they observe neither the size of the trade deal nor the retailer's cost type. They must infer this information from advertised prices and past experience. Equilibrium prices, $p = \phi_i(D)$, are a function of the store's cost type, H or L, and trade deal size, D. A consumer that knows a store's cost type and observes the sale price can accurately infer the deal size using $\phi_i^{-1}(p)$. But deal prone consumers do not know a store's cost type and may not have sufficient information to accurately infer D. As in the Search Model, when the consumer observes only the sale price, it is not clear whether the store has low regular prices (type L) and a shallow discount (D_L) or high regular prices (type H) and a deep discount (D_H). (Figure 1)

One can similarly see in Figure 5 that advertising in the sale price only format does not clearly communicate the store's cost type to deal prone consumers. If the equilibrium prices do not fully reveal the store's type and consumers' posterior beliefs are that store types are equally likely, expected demand is: $\alpha x_E = \alpha[f_E + g_E D]$ where $f_L > f_E > f_H$ and $g_L \geq g_E \geq g_H$. In words this means that when consumers are ex-post uncertain, they optimize expected profits and use an "in between" threshold price rule.

We simplify the retailer's problem by assuming that they have a positive discount factor for profits in the next two periods (period 1 and period 2) and ignore all future profits. Thus, full information demand and profits per period are the same as in Search Model (Table 1) except that now deal prone demand is endogenous. Equilibrium prices for the Wait Model can be found in the Appendix. As in the Search Model, the retailer's strategy will take into consideration additional costs of the regular sale price format due to future uncertainty, execution costs, or opportunity costs, $F_i > 0$.

To summarize the game, we consider a retailer selling a product in multiple periods. In odd periods all consumers know that there is a sale (trade deal present) and that the trade deal, D, is drawn from a known uniform distribution. In period 2, there is no sale. A mass of α deal prone consumers choose to purchase at the sale price or wait until the next sale period. In both Period 1 and Period 2, a $(1-\alpha)$ mass of loyal consumers decides whether to purchase the item. The retailer maximizes profits by choosing the optimal sale price, the optimal regular price, and the optimal selling format.

Results: Wait Model

Results for the Wait Model are analogous to the Search Model except now retailers have incentive to imitate the high cost type. When consumers can postpone a purchase, a retailer with high costs is hurt by consumer uncertainty. When consumers are uncertain as to whether the discount is small or large, they infer that the discount is average. A high cost store with a deep discount loses deal prone consumers as some of the consumers incorrectly decide to wait for a deeper discount and a lower price.

We quickly summarize the main results and refrain from formally restating Propositions 1-3 with the low cost type replaced by the high cost type. Analogous to Proposition 1, the high cost store now has incentive to signal his cost type and depth of discount and the commitment property deters the low cost store from raising the regular price. Similar to Proposition 2, the high cost store will distort both the regular and sale price to signal that it has high costs and a deep discount. However, now the high cost store raises prices to signal his type. Finally, the main result of the paper, Proposition 3, is also analogous. The high cost store has incentive to post both the regular and sale prices, but only when there are enough deal prone consumers. That is, for α close to 0, profits from the deal prone consumers are small and expected costs are large. Once there are enough deal prone consumers, the high cost store has incentive to post both the regular and sale prices to signal his cost type and the correct size of the discount.

Comparing the results of the Search Model and the Wait Model, one observes that the main result of the paper is consistent across both models. We find that for any retailer to have incentive to post both the regular and sale price, there have to be enough deal prone consumers. The incentives to distort prices higher or lower depend on the extent to which consumers either search at competing retailers and/or wait for lower prices. We now empirically test this result.

Section 5: Empirical Test

The main result of both the Search Model and the Wait Model is that the regular/sale price format should be less prevalent in markets with a smaller proportion of deal prone consumers (Proposition 3). The parameter α combines two critical characteristics of the market: the level of information and the willingness of consumers to postpone a purchase. *Ceteris paribus*, the regular/sale price selling format should be more prevalent when price information level is low and there is a larger proportion of patient consumers who either search across retailers or wait for better deals.

To test this result we collected print advertisements from six regional Sunday papers on August 21, 1994. Weekly store fliers and print advertisements were included in the sample. The criterion for inclusion in the sample was mention of a specific product being sold at a specific price. We did not collect travel or automobile advertisements. In the six markets, there were over 200 different retailers who advertised in either the regular/sale price format or sale price only format in the Sunday paper. Because the number of retailers and the number of products advertised was quite large, we focus on a single market which contains 37 different retailers. The data set consists of a random sample of products in each advertisement for a total of 363 product advertisements.

Since neither the level of information, the willingness to search, nor the ability to postpone a purchase are observable, we need to identify variables which are correlated with these constructs. We collected the following variables: Price of the Product, Number of Products in the Advertisement, Purchase Infrequency, Search or Experience Good, Durable Good, Dissimilarity of (Product, Retailer) pair to others, and measures of Category and Retailer Diversification. We now explain the reason for collecting each of these variables and their hypothesized correlation with the use of the regular/sale price format.

While the patience of consumers is unobservable, researchers have found that the magnitude of the price is correlated with the implied discount rate. (Ben-Zion et al 1989, Thaler 1981) That is, as products become more expensive people tend to become more patient. In the Wait Model, the key property of deal prone consumers is that they are willing to wait until the next sale period. If discount factors are heterogeneous across the population, there will be a threshold discount factor and consumers with discount factors above this threshold will postpone a purchase. The empirical evidence suggests that for more expensive products, the mean of the

distribution of discount factors increases. Thus price may serve as a rough proxy for an individual's willingness to postpone a purchase²² and should be positively correlated with use of the regular/sale price format.

Another variable which was convenient to collect was the number of products in the retailer's advertisement. While not directly part of the theoretical model, we hypothesize that a retailer advertising more products has more to gain by signaling that he has low costs. Thus we hypothesize that the number of products in the advertisement will be positively correlated with the use of the regular/sale price format.

While we were unable to directly measure consumer knowledge of a product category, it seems likely that purchase infrequency should lead to less product and retailer knowledge. To obtain an objective measure of purchase frequency, we collected the data used by Tellis and Wernerfelt (1987). This data set coded hundreds of products on a integer scale of 1-4 where 1 is high purchase frequency and 4 is low purchase frequency. We then mapped purchase infrequency for product types from this data set to product types in our data set. When a product in our data set did not exist in the Tellis and Wernerfelt data set, the most similar type of product was used to estimate the purchase infrequency. We hypothesize that purchase infrequency should be positively correlated with use of regular/sale price selling format.

In the Search Model, consumer's ability to search at other retailers creates an incentive to post a low regular price. An accepted technique for classifying products into search or experience goods is Nelson (1970). Often search goods are used as a proxy for lack of product quality information. In our model we expect search to be positively correlated with use of the regular/sale price format.

The desire to postpone a purchase is also not observable. However, whether or not a product is a durable good seems to be a reasonable proxy for the ability to postpone a purchase. In fact, some researchers have advocated that one definition of a durable good is the ability to postpone a purchase. We use the simple definition of

²²We will primarily focus on the price of an item because of its ease of implementation. However other factors may influence the decision to postpone a purchase. If the product satisfies a critical need consumers tend to behave myopically. For example, a decision to upgrade one's television is much different than a decision to buy more milk. You can probably live with the old television, but cereal is very dry without milk. In addition, the ability to inventory a product for future consumption influences a consumer's ability to time a purchase. Bulky items and items that perish are more difficult to inventory and thus we expect consumers to behave myopically. While not in our test, we expect both of these factors to influence the decision to postpone a purchase.

whether a product was consumed during use to classify durable goods. We expect that durable goods will be more likely to be advertised in the regular/sale price format.

In addition, we have collected measures of information by product category and retailer from the 1987 Census of Retailers. The first variable we collect is the dissimilarity of the (product category, retailer) pair. We define the similarity of a (product category, retailer) pair as the proportion of a product category sold at a specific retailer type. For example, the proportion of all tires sold at department stores. We hypothesize that consumers will have more price information if this variable is larger. In some sense, this variable captures the similarity of a retailer to other retailers in the product category. A low percentage indicates that the retailer's store is not the "normal" place to buy such a product. Hence, one may expect less price knowledge and more use of the regular/sale price format. We work with dissimilarity by transforming this variable to one minus the proportion of products sold at the (product, retailer) pair. We expect this variable to be positively correlated to the regular/sale price format.

Using the same retailer census, we were also able to construct a product category and retailer measure of diversification. For each product category, we collected the percentage of products sold at each retailer type. For each retailer type, we collected the percentage of products sold in each product category. We then used the procedure in Jacquemin and Berry (1979)²³ to construct an index of diversification. The index indicates the extent of concentration in the product category and retailer. An index closer to zero indicates a more concentrated product category or retailer. For the product category, the diversification measure is an indicator of product category knowledge. Consumers should have more price knowledge in more concentrated product categories. For retailers, the diversification measure is an indicator of the incentive to sell other products on a store visit. Thus more concentrated retailers may have less incentive to include a regular price. We expect that both of these variables will be positively correlated with the regular/sale price format.

The correlation matrix of the variables can be found in Table 2. Note that all of the variables are correlated with the regular/sale price variable in the hypothesized direction. We use a binomial logit model to estimate model coefficients. The dependent variable is whether the advertisement includes a regular price. We take the log of both the price and number of products in the ad due to their large variance.

²³The entropy measure is: $E_j = \sum_j s_j \cdot \log(1/s_j)$ where s_j = share.

Correlation Matrix of Dependent and Independent Variables

Regular/Sale Price Format	1.00							
log(Price)	0.09	1.00						
log(Number of Products in Ad)	0.25	0.11	1.00					
Infrequency of Purchase	0.12	0.55	0.09	1.00				
Search Good	0.31	0.11	0.18	0.25	1.00			
Durable Good	0.34	0.58	0.11	0.51	0.46	1.00		
Dissimilarity of(Product, Retailer)	0.37	0.10	0.38	0.18	0.26	0.09	1.00	
Product Category Diversification	0.25	0.22	0.24	0.32	0.19	0.28	0.24	1.00
Retailer Diversification	0.25	0.21	0.59	0.03	0.22	0.20	0.36	0.16

Table 2

Results of the logit model can be found in Table 3. The results from the model tend to support the main result of the paper. The most significant variables are durable good, similarity of (product category, retailer) and product category diversification. While the price and purchase infrequency of the item were positively correlated with the regular/sale price format (consistent with hypothesis), the coefficients in the full logit were negative (inconsistent with hypothesis). The search good variable had weak significance in the full logit. Examining the correlation matrix in Table 2 this tends to make some sense. The durable good variable is correlated with the price, infrequency of purchase, and search good variables ($\rho=0.58, 0.51, 0.46$). This may explain why only one of these variable is significant and has the proper sign in the full logit. The model also seems to predict fairly well with respect to the null model. Note that the percent correctly identified with a dummy variable is 53% and the percent correctly identified by the model is 79%.

Results from Binomial Logit

Variable	Coefficient	Standard Error	
log(Price)	-0.17	0.09	
log(Number of Products in Ad)	0.20	0.16	
Infrequency of Purchase	-0.24	0.20	
Search Good	0.30	0.20	
Durable Good	2.55	0.51	
Dissimilarity of (Product, Retailer)	3.65	0.68	
Product Category Diversification	1.37	0.59	
Retailer Diversification	-0.01	0.49	
Constant	-1.72	1.04	
Log likelihood value	-190	U-sq adjusted	0.21
Null Log likelihood (equal prob.)	-252	Mean Probability of Correct Choice	0.65
Percentage of Regular/Sale Price Ads	47%	Percentage of Correct Choice	79%
Dependent Variable: Regular/Sale Price Format			

Table 3

It is also interesting that the retailer diversification measure is insignificant. One possibility for this is that the measure is confounded with consumer price information at the retailer. That is, consumers have greater price knowledge at retailers that concentrate in selling one product category. One reason for not finding support for the purchase infrequency variable in the full logit is that the scale may not be broad enough for the products we collected. For packaged goods, the IRI Marketing Factbook has more precise measures of purchase infrequency. However, a comparable source for non-packaged goods was not available.

Why Advertise Both Regular and Sale Prices?

In general, the results support the main hypothesis. The regular/sale price tends to be used more often when consumers have less information and are willing to search or postpone a purchase.

Section 6: Discussion

The main result of both models is that there have to be enough deal prone consumers for any retailer to have incentive to advertise both the regular and sale prices. The purpose of approaching the problem with two distinct models was to provide rational signaling explanations as to why a retailer may have incentive to advertise either a low or high regular price.

As we consider the models in isolation, it makes some sense to consider what happens when we merge the two models and allow consumers to both search and wait. While we do not present any formal analysis, it is clear that the search incentives outweigh any waiting incentives. Consider a high cost type retailer that posts a high regular price. Deal prone consumers know that whether they are going to purchase today or in the future, they are better off searching at other retailers. Therefore, all retailers have incentive to signal that they are the low cost type.

If searching behavior dominates, why is it interesting to model consumer waiting behavior? First, we believe that in practical situations there are many customers that are price sensitive, patient, and store loyal. Second, we argue that retailers exert considerable effort to make comparison across stores (search) difficult. A classic example of this is the use of different model numbers for the same product at different retailers. Given the difficulty of comparison across retailers, the Wait Model offers an explanation as to why retailers may advertise a high regular price. Finally, in our empirical evidence we find retailers that advertise high regular prices (e.g. Bloomingdale's) and retailers that advertise low regular prices (e.g. Bob's Stores). If the search model was the only explanation, we would only observe low regular prices.

The Wait Model demonstrates the existence of a signaling equilibrium where retailers have incentive to advertise high regular prices. We note that simple modifications of our model can provide signaling explanations of high reference prices. Suppose a consumer believes that a retailer's regular price is correlated with product quality. Now observing a sale price leads to confusion over the size of the current discount and product quality. A retailer with a large discount will have incentive to include a regular price to signal high product quality. As another example, suppose consumers are interested in signaling to others and that higher regular prices have higher signaling value. For example, you are buying skis at a year end sale. The signaling value you receive when other people see your skis increases with the regular price. By posting a high regular price, retailers inform consumers that the discount is large and that there is a large signaling value.

These simple extensions may also provide signaling explanations for alternative types of reference prices. While the regular price was the most frequently used reference price, we found sixteen other types of reference prices other types of reference prices. For example, "Compare to", "<Named Competitor's> Price", or "Their Price" reference price may be used to leverage the reputation of another retailer. If consumers believe that product quality and regular prices are correlated at an established retailer, then a competing retailer can use this as a signal. In addition to signaling quality, advertising competitive prices may also discourage search at other retailers. Another common example are "Was" prices which have a similar flavor to regular prices in the Wait model. A consumer may know that prices fall over time and that the optimal time to buy from this retailer is when the price reaches 40% discount. Just as in the Wait Model, posting the regular price signals to the consumer what the current percent discount. Thus our model provides a signaling explanation for many types of reference prices.

A further contribution of this paper is that it offers a rational explanation for why consumers use a type of reference price. There is a tremendous amount of literature in marketing, psychology, and economics on reference prices. There is well documented evidence suggesting that individuals may evaluate alternatives with respect to a reference price. However, in the classical economic framework, there is not a rational explanation as to why consumers should use reference prices. Further, the behavioral explanation by Thaler suggests that reference prices should always be as high as possible. Our signaling model demonstrates that competition (Search Model) provides incentive for retailers to post low regular prices. In the absence of competition, we demonstrate that retailers have incentive to advertise high regular prices. We have also discussed how our model can be modified to rationalize many other types of observed reference prices.

The Search Model seems particularly complementary to the reference price literature. Using the transaction utility model, this literature has suggested that raising the reference price should reduce search. Our Search Model suggests that what may also matter is consumers' beliefs about alternative retailers, in this case an outside option (passive competition). If the reference price truly carries information, raising the reference price should signal that other retailers may have lower prices. The Search Model is consistent with retailers such as Bob's Stores who sometimes include the phrase "Our Low Regular Price" in their advertisements. These retailers are trying to send a clear message that their regular price is lower than other retailers' regular prices and therefore their sale price should also be lower.

An additional result of the Wait Model is that it rationalizes why more consumers may purchase when the discount (percentage off) is larger but the price is the same. One would arrive at the same conclusion using the transaction utility model. But what drives the result in the Wait Model is that consumer demand is determined by expectations over future prices. More consumers buy when the percent off is larger because there is less chance of a lower price in the future. While the reasoning is different, we arrive at the same conclusion as the transaction utility theory model. The consistency of this result with two different theories raises an interesting question. From a managerial perspective it would be useful to understand the degree to which consumer price expectations and psychophysical phenomenon (prospect theory) influence which products should be promoted. In future work we hope to disentangle some of these issues with experimental evidence.

Retailers often select a particular selling format to signal information to their consumers. In this model, a retailer signals his cost type and depth of discount with advertised prices. Another possibility is that the prices signal information other than the retail cost type and depth of discount. For example, a result of the Search Model is that when deal prone consumers search, high cost retailers never use the regular/sale price format²⁴. In practice, one may observe high cost retailers using the regular/sale price selling format. One reason why this may make sense is if the high cost store also offers higher service quality than the low cost store. In this case, the selling format is being used to both signal high quality/high regular price as well as a deep discount. Future work may wish to address implications of signaling a quality image in addition to a price image.

Both models contain a number of simplifying assumptions which may warrant future work. In the Wait Model we assume that the proportion of patient, uninformed consumers is constant from period to period. It may be more realistic to assume that when prices are relatively high, more and more patient consumers postpone their purchase. As this mass of deal prone consumers increases, there is incentive to offer a deep discount to attract this large segment of consumers. This suggests a pattern where the time since the last deep discount is positively correlated

²⁴It should be noted that with simple extensions to the model, we can construct an equilibrium where the high cost store has incentive to reveal his type when customers search. From the customers perspective, they only care about getting a low price and don't care whether the store has high or low costs. Adding additional states of uncertainty is one way to generate an equilibrium strategy where the high cost store creates an unusually large gap between the regular and sale price to signal his strategy of matching the low cost store's price.

with the probability of a deep discount today. In practice, we tend to observe deep discounts, such as One Day Sales, separated from each other. Examining the impact of patient consumers accumulating over time may provide insight into the optimal depth of discount. Future work may explore a model that endogenizes the deal size and the size of the deal prone consumer segment each period. Finally, the focus of this model is on explaining the comparative effectiveness of the regular/sale price format versus the sale price only format. To do this, we focus on a retail monopolist with passive competition. It may be useful to consider the influence of competition on the optimal retailer strategies.

While these suggestions are in the direction of further complexity, we may also want to consider simplification. We currently assume that the distribution of trade deals is uniform which, as one can see in the appendix, leads to complicated analysis. As noted in the paper, it is possible to construct a similar model with two deal sizes (high and low) rather than continuous deal sizes. This would simplify analysis, but the current model demonstrates the robustness of the intuition. In particular, in our model there does not need to be purposeful deception by retailers for consumers to be uncertain. The fact that the range of full information equilibrium prices for the high and low type retailer intersect may generate consumer uncertainty.

The current paper has offered an explanation for the existence of the regular/sale price format and argues that it is one means of effectively communicating information about the relative value of a price. A curious part of the data collection exercise was discovering alternative selling formats. As mentioned, there were at least sixteen other types of reference prices. In addition, terminology to indicate the presence of a sale varied by retailer. At one retailer, "Value Price" was a non-discounted everyday low price while "Extra Value Price" was a discounted price. Two other common occurrences in the sample were the use of competitive price guarantees and low price financing. Both the length (e.g., 7 day, 30 day, etc.) and size (e.g., 100%, 110% or 150% "price protection") of these offers varied by retailer. While some of these are just curious findings, it is suggestive of alternative selling formats to consider in future research.

Conclusion

We have offered an explanation as to why retailers may prefer the regular/sale price selling format over a sale price only format. Our model has highlighted the importance of deal prone consumers who may seek lower prices by either searching or waiting. An important contribution of this paper is that it offers a rational explanation for the existence of sale prices. In contrast to the existing literature on reference prices, we find that increasing the reference price may not always be beneficial, particularly if it induces customers to search at other retailers. Consistent with the reference price literature, we provide a rational explanation for why more consumers may purchase when the price is the same but the discount is larger. We also provide empirical evidence to support our main result.

Multi-product retailers use a wide variety of selling formats to signal many different types of information that they consider valuable such as a low price image, a reputation for high service quality, outstanding availability etc. This paper contributes to the literature on comparative selling formats by evaluating the implications of bargain hunting (searching and waiting for lower prices) on the part of consumers. In general, much work remains to be done in understanding the comparative properties of alternative selling formats.

Why Advertise Both Regular and Sale Prices?

Appendix 1: Rationalization of Linear Demand

Search Model

We justify the linear demand assumption: $x_i = v_i - w_i p$

Consumer's Utility function:

$$U(n,p) = v - wp - n$$

Consumer's Outside Option:

$$\bar{U}^o(n,p) = E(U^o(n,p)) = 0.5((U^o(n,p|D_H)) + (U^o(n,p|D_L))) - s(n)$$

where $s(n)$ is search cost of consumer with taste address n and let $E^o(p)$ be the expected price at the outside option.

Consumer's Problem:

- Buy from retailer i
- Buy from outside option
- Don't Buy

Full Information: Consumer knows retailer's cost type and deal size

(i) Low Cost Type Demand

$$v - wp - n > U^o(n,p|D_L) \text{ implies}$$

$$x_L = v - wp$$

(ii) High Cost type Demand

$$v - wp - n < U^o(n,p|D_H) \text{ implies}$$

$$x_H = 0$$

Partial Information: Consumer does not know retailer's cost type

(i) Demand at retailer

$$U(\text{Buy from retailer}) > U(\text{Buy from outside option}) > 0$$

$$v - wp > \bar{U}^o(n,p)$$

(ii) Demand at outside option

$$U(\text{Buy from outside option}) > U(\text{Buy from retailer}) > 0$$

$$v-wp > \bar{U}^0(n,p)$$

Assume that $s(n)$ is randomly distributed as \bar{s} and \underline{s} among deal prone consumers. We assume:

$$\bar{s} > w(p-E^0(p)) > \underline{s}$$

This implies that consumers with high search cost, \bar{s} , will buy from retailer i and consumers with low search cost will buy from the outside option (assuming that utility is positive: $v-wp > 0$).

(iii) Conditions such that consumers do search retailer after searching the outside option.

Assume a consumer searches at the outside option. Let \bar{p} be a bad realization at the outside option. We assume:

$$\underline{s} > w(\bar{p}-p)$$

such that the consumer will not search at the high cost retailer.

These assumptions justify our linear demand. One can think of the outside option as a below average cost retailer. As a simple example, suppose $w=1$, $p=11$, $\bar{s}=3$, and $\underline{s}=1.5$. Assume that $\bar{p}=12$ and $\underline{p}=6$ so expected price at the outside option is 9. Since $(11-9-3) < 0$ for all n , high search cost consumers buy from the retailer or don't buy at all. Since $(11-9-1.5) > 0$ for all n , low search cost consumers buy from the outside option or don't buy at all. Even if these consumers learn that the price is high ($p=12$) they do not return to the retailer as $(12-11-1.5) < 0$.

Wait Model: Justification of linear demandCheap Separating Equilibrium Demand

We describe the system of equations that define an equilibrium and show that there exist linear FOC and a linear stopping rule when the constraints of the retailer profit maximization are not binding (as in a cheap separating or pooling equilibrium). This implies that there exists an endogenous linear demand function $x_i = v_i - w_i p$.

Claim 1:

In a cheap separating equilibrium, linear Demand: $x_i = f_i + g_i D$ and Linear Equilibrium Prices: $p = \phi_i(D)$ are consistent with a utility function which is quadratic in n .

Proof:Retailer Profit Maximization

Assume that $\phi_i(D)$ is linear and invertible. Rewrite as: $\phi_i^{-1}(p) = D = mp + K_i$. The outline of the proof is as follows. We first maximize retailer profits and solve for m and K_i . Next we maximize expected utility using Bellman's equation and show that it is satisfied by a linear demand function $x_i(D)$.

Because we are focusing on the unconstrained problem we only consider period 1 profits. From Table 1, Retailer i Profit in Period 1 is:

$$\text{Max}_p \pi_i^1 = [\alpha x_i + (1-\alpha)y^1](p + D_t - c_i)$$

where: $D_t = \text{ACTUAL trade deal}$

p = price

c_i = marginal cost

We add the subscript to D_t to distinguish it from the inferred deal size D .

Substituting Demand and equilibrium prices:

$$\text{Max}_p \pi_i^1 = [\alpha(f_i + g_i D) + (1-\alpha)(a-bp)](p + D_t - c_i)$$

$$\text{Max}_p \pi_i^1 = [\alpha(f_i + g_i \phi_i^{-1}(p)) + (1-\alpha)(a-bp)](p + D_t - c_i)$$

FOC:

$$[\alpha(f_i + g_i \phi_i^{-1}(p)) + (1-\alpha)(a-bp)] + (p + D_t - c_i)[\alpha g_i \phi_i'^{-1}(p) - (1-\alpha)b] = 0$$

$$[\alpha(f_i + g_i \phi_i^{-1}(p)) + (1-\alpha)(a-bp)] + (p + D_t - c_i)[\alpha g_i m - (1-\alpha)b] = 0$$

Setting: $D_t = \phi_i^{-1}(p)$ and recalling that $D = mp + K_i$

$$\phi_i^{-1}(p) = D$$

$$= \frac{p[2(1-\alpha)b - \alpha g_i m] - [\alpha f_i + (1-\alpha)a] - c_i[-\alpha g_i m + (1-\alpha)b]}{[\alpha g_i(1+m) - b(1-\alpha)]}$$

$$\frac{d\phi_i^{-1}(p)}{dp} = m = \frac{[2b(1-\alpha) - \alpha g_i m]}{[\alpha g_i(1+m) - (1-\alpha)b]}$$

A1.1

Solving A1.1 for m we get $m = -2$ as one solution.

Using $m = -2$, we get:

$$\begin{aligned} K_i &= -\frac{[\alpha f_i + (1-\alpha)a]}{[\alpha g_i(1+m) - (1-\alpha)b]} - c_i \frac{[-\alpha g_i m + (1-\alpha)b]}{[\alpha g_i(1+m) - (1-\alpha)b]} \\ &= \frac{[\alpha f_i + (1-\alpha)a]}{[\alpha g_i + (1-\alpha)b]} + c_i \frac{[2\alpha g_i + (1-\alpha)b]}{[\alpha g_i + (1-\alpha)b]} \end{aligned}$$

Rewriting:

$$D = -2p + K_i$$

$$p = K_i/2 - D/2$$

$$p = \frac{[\alpha f_i + (1-\alpha)a]}{2[\alpha g_i + (1-\alpha)b]} + \frac{c_i}{2} \frac{[2\alpha g_i + (1-\alpha)b]}{[\alpha g_i + (1-\alpha)b]} - \frac{D}{2}$$

Deal Prone Consumer Utility Maximization:

We now show that a utility function that is quadratic in n leads to linear demand $x_i = f_i + g_i D$.

Let:

$$V_i(p) = A_i - B_i p - n + L_i n^2$$

where: A_i, B_i, L_i are constants and i is the retailer type.

Using $p = \phi_i(D)$, we can substitute the equilibrium price into the utility function.

$$D = mp + K_i$$

$$p = (D - K_i)/m$$

Let D_i^t equal the trade deal size such that a customer is indifferent between buying today and waiting until tomorrow. \underline{D}_i^t is the expected trade deal size given that $D > D_i^t$. We assume that $D \sim U(0,1)$. Therefore:

$$K(D) = D_i^t$$

$$\underline{D}_i^t = E[D | D > D_i^t] = (D_i^t + 1)/2$$

Utility from Buy Today:

$$V_i(p_i^t, n) = A_i - n + L_i n^2 - B_i p_i^t$$

$$V_i(D_i^t, n) = A_i - n + L_i n^2 - (B_i/m)(D_i^t - K_i)$$

Expected Utility if Wait:

$$E[V^i(p,n | p < p_i^t)] = V^i(\underline{p}_i^t, n) = A_i - n + L_i n^2 - B_i \underline{p}_i^t$$

$$E[V^i(p,n | D > D_i^t)] = V^i(\underline{D}_i^t, n) = A_i - n + L_i n^2 - (B_i/m)((D_i^t + 1)/2 - K_i)$$

Bellman's Equation leads to:

$$V^i(D_i^t, n) = \delta_n [K(D_i^t) V^i(D_i^t, n) + (1 - K(D_i^t)) V^i(\underline{D}_i^t, n)]$$

$$[A_i - n + L_i n^2 - (B_i/m)(D_i^t - K_i)](1 - \delta_n D_i^t) =$$

$$[A_i - n + L_i n^2 - (B_i/m)((D_i^t + 1)/2 - K_i)](\delta_n - \delta_n D_i^t)$$

with a little algebra we get:

$$[A_i - n + L_i n^2](1 - \delta_n) + B_i/2[-K_i - \delta_n/2 + \delta_n K_i + D_i^t - \delta_n(D_i^t)^2/2] = 0$$

We have a quadratic in n and a quadratic in D_i^t . Appropriate choice of parameters will lead to a linear solution. As we assume that $n \sim U(0,1)$, demand is linear: $x_i = f_i + g_i D_i$.

Appropriate parameters are:

$$4L_i(1 - \delta_n) = B_i \delta_n$$

$$B_i/[2(1 - \delta_n)] = g_i$$

$$f_i = A_i + (B_i/2)[-K_i - \delta_n/2 + \delta_n K_i]/(1 - \delta_n)$$

As $V^i(p,n)$ is quadratic in n , we need to rule out regions ($n > n^*$) where marginal utility is increasing in n : $\frac{dV^i(p,n)}{dn} > 0 \quad \forall n > n^*$. Since $n \sim U(0,1)$, one could assume

$$\text{that } \left. \frac{dV^i(p,n)}{dn} \right|_{n=1} < 0.$$

This completes the proof. We have found a utility function and profit function that are consistent with linear demand and linear equilibrium prices.

QED

Pooling and Costly Separating Equilibrium Demand

When customers are uncertain about the retailer's cost type, $\theta=1/2$, they maximize expected utility. We continue to assume that this demand is linear. This is an additional constraint that must be considered in selecting the parameters f_i, g_i, A_i, B_i, N_i . There are sufficient degrees of freedom that this is not a problem.

In a costly separating equilibrium, retailer type i solves a constrained maximization problem and the consumer again solves Bellman's equation. While we have not shown it, we assume that there exists a sufficiently general utility function such that demand continues to be linear in this region.

Retailer Profit Maximization:

$$\begin{aligned} \text{Max}_{s_L^*, r_L^*} \quad & \pi_L^1(s_L^*, \theta) + \delta_R \pi_L^2(r_L^*) - F_L \\ \text{s.t.} \quad & \pi_H^1(s_L^*, 1) + \delta_R \pi_H^2(r_L^*) - F_H = \pi_H(p_H^1, p_H^2, 0) \end{aligned}$$

Let $p = \tilde{\phi}(D|c_i) = \tilde{\phi}_i(D)$ equal the equilibrium sale price (Solution to the FOC) We assume that it is monotonic and invertible. Therefore, the consumer can observe p and invert the function to solve for a unique D .

Consumer Utility Maximization

Bellman's equation is:

$$V^i(p_i^t, n) = \delta_n [K(D_i^t) V^i(p_i^t, n) + (1-K(D_i^t)) V^i(\underline{p}_i^t, n)]$$

This leads to a system of two equations and two unknown functions. The two equations are the FOC from $R(s_L, r_L)$ and Bellman's equation and the unknown functions are $\tilde{\phi}_i(D)$ (or $\tilde{\phi}_i^{-1}(p)$) and $p_i^t(n)$. Solving for these

Why Advertise Both Regular and Sale Prices?

functions we can get the equilibrium prices and profit: (s_L^*, r_L^*) and $\pi_L(s_L^*, r_L^*, 1)$.

Appendix 2: Cheap Separating Equilibrium Prices and Profits

General Conditions for Cheap Separating Equilibrium:

A cheap separating equilibrium will exist as long as the high cost store has no incentive to imitate the low cost store. Let π^*_H equal the maximum profits to the high cost store when his type is revealed: $\pi^*_H = \text{Max}[\pi_H(p_H, r_H, 0), \pi_H(p_{1H}, p_{2H}, 0)]$. The high cost store will have no incentive to deviate to each format if either of the following conditions hold.

$$\text{Regular/Sale Price Format:} \quad \pi_H(s_L, r_L, 1 | D=1) \leq \pi^*_H \quad \text{C1}$$

$$\text{Sale Price Only Format:} \quad \pi_H(p^1_L, p^2_H, 1 | D=1) \leq \pi^*_H \quad \text{C2}$$

Search Model:

Prices and Profits

In the search model, the cheap separating equilibrium is equivalent to the full information equilibrium. Signaling is not relevant and retailers maximize current period profits.

Retailer Maximizes:

$$\pi_i = \pi_i^1 + \delta_R \pi_i^2$$

$$\text{Max}_p \pi_i^1 = [\alpha x_i + (1-\alpha)y^1](p+D-c_i)$$

$$\text{Max}_r \pi_i^2 = [(1-\alpha)y^1](p+D-c_i)$$

FOC lead to:

$$p^*_i = \frac{\alpha v_i + (1-\alpha)a}{2[\alpha w_i + (1-\alpha)b]} + \frac{c_i - D}{2} \text{ for } i = L, H$$

$$r^*_i = \frac{a}{2b} + \frac{c_i}{2} \text{ for } i = L, H$$

$$\pi_i(p_i^*, r_i^*) = (p_i^* - c_i + D)^2[\alpha w_i + (1-\alpha)b] + \delta_R(r_i^* - c_i)^2(1-\alpha)b$$

where: $v_L = v$, $w_L = w$, and $v_H = w_H = 0$.

Therefore, C1 and C2 can be expressed as:

$$(p_{L-C_H}^* + D)^2[\alpha w_L + (1-\alpha)b] + \delta_R(r_{L-C_H}^*)^2(1-\alpha)b < (p_{H-C_H}^* + D)^2(1-\alpha)b + \delta_R(r_{H-C_H}^*)^2(1-\alpha)b \quad C1$$

$$(p_{L-C_H}^* + D)^2[\alpha w_L + (1-\alpha)b] < (p_{H-C_H}^* + D)^2(1-\alpha)b \quad C2$$

There exist parameters such that these conditions are satisfied.

Wait Model:

In the Wait Model, due to endogenous demand, the cheap separating equilibrium and full information equilibrium are not equivalent. One can solve for the full information demand, prices and profits using the utility function in Appendix 1. We only consider the cheap separating equilibrium.

Cheap Separating:

As demand is endogenous, we have solved for equilibrium prices in Appendix 1.

$$p_i^* = \frac{[\alpha f_i + (1-\alpha)a]}{2[\alpha g_i + (1-\alpha)b]} + \frac{c_i [2\alpha g_i + (1-\alpha)b]}{2 [\alpha g_i + (1-\alpha)b]} - \frac{D}{2}$$

$$r_i^* = \frac{a}{2b} + \frac{c_i}{2}$$

$$\pi_i(p_i^*, r_i^*) = (p_i^* - c_i + D)^2[2\alpha g_i + (1-\alpha)b] + \delta_R(r_i^* - c_i)^2(1-\alpha)b$$

As in the Search Model, conditions C1 and C2 can be written in terms of the parameters $(\alpha, c_i, D, a, b, v_i, w_i, f_i, g_i)$. We leave this to the reader.

Appendix 3: Costly Separating Equilibrium

General Conditions for a Costly Separating Equilibrium

The general conditions for a costly separating equilibrium in both the Search Model and the wait model are:

No Incentive for Type H to Imitate:

$$\pi_H(s^*_L, r^*_L, 1 | D=1) \leq \pi^*_H \quad C3$$

$$\pi_H(p^{1*}_L, p^{2*}_H, 1 | D=1) \leq \pi^*_H \quad C4$$

Pooling is not Preferred by type L:

$$\pi_L(s^*_L, r^*_L, 1) \geq \text{Max}[\pi_L(s_{P_0}, r_{P_0}, 1/2), \pi_L(p^1_{P_0}, p^2_{P_0}, 1/2)] \quad C5$$

$$\pi_L(p^{1*}_L, p^2_{P_0}, 1) \geq \text{Max}[\pi_L(s_{P_0}, r_{P_0}, 1/2), \pi_L(p^1_{P_0}, p^2_{P_0}, 1/2)] \quad C6$$

Where the subscript (p_0) represents the optimal pooling equilibrium prices. (i.e. when deal prone demand is x_E)

Search Model and Wait Model:

We assume the same demand for both models and solve for the optimal prices. In a costly separating equilibrium, the retailer solves a constrained maximization where constraint C3 or C4 holds. We solve for the optimal prices and profits in the regular sale price format.

$$\text{Max}_{s,r} \pi_L(s, r, \theta = 1)$$

$$\text{s.t. } \pi_H(s, r, \theta = 1 | D=1) \leq \pi^*_H$$

FOC:

$$\frac{\partial \pi_L}{\partial s} - \lambda \frac{\partial \pi_H}{\partial s} = 0$$

Why Advertise Both Regular and Sale Prices?

$$\frac{\partial \pi_L}{\partial r} - \lambda \frac{\partial \pi_H}{\partial r} = 0$$

$$\pi_H(s, r, \theta = 1 | D = 1) - \pi_H^* = 0$$

Eliminating λ :

$$\frac{\partial \pi_L}{\partial s} \frac{\partial \pi_H}{\partial r} = \frac{\partial \pi_L}{\partial r} \frac{\partial \pi_H}{\partial s} = 0$$

Solving for s, r :

$$s = s^o + r - a/2b - K_1 \tag{A3.1}$$

where:

$$s^o = \frac{\alpha v_i + (1 - \alpha)a}{2[\alpha w_i + (1 - \alpha)b]}$$

$$K_1 = \frac{\delta(1 - \alpha)(a - br)(1 - D) - \delta(1 - \alpha)b(c_H - c_L D)}{(c_H - c_L)}$$

$$r = \frac{a}{2b} + \frac{c_H}{2} - \frac{\sqrt{(a - bc_H)^2 + 4bK_2(s)}}{2b} \tag{A3.2}$$

where $K_2(s) = \frac{\pi_H^1(s, \theta = 1 | D = 1) - \pi_H^*}{\delta_R(1 - \alpha)}$

$K_2(s)$ is a quadratic in s . Substituting A3.2 into A3.1 and squaring both sides, we end up with a quadratic in s . We can solve this for s and then substitute into A3.2 to solve for r .

Similarly, one can solve for the optimal prices in the sale price only format.

$$\text{Max}_p \pi_L(p, \theta = 1)$$

$$\text{s.t. } \pi_H(p, \theta = 1 | D = 1) \leq \pi_H^*$$

FOC:

$$\frac{\partial \pi_L}{\partial p} - \lambda \frac{\partial \pi_H}{\partial p} = 0$$

$$\pi_H(s, \theta = 1 | D = 1) - \pi_H^* = 0$$

Solving for p:

$$p = \frac{C_1 + C_2 C_3}{2C_2} - \frac{\sqrt{(C_1 - C_2 C_3)^2 - 4\pi_H C_2}}{2C_2}$$

where:

$$C_1 = [\alpha v_L + (1-\alpha)a]$$

$$C_2 = [\alpha w_L + (1-\alpha)b]$$

$$C_3 = [c_H - 1]$$

We can now express conditions C3 and C4 in terms of the parameters.

There exist parameters such that these conditions are satisfied.

Appendix 4: Pooling Equilibrium

General Conditions for a Costly Separating Equilibrium

If either C5 is not satisfied at the optimal (s_L, r_L) or C6 is not satisfied at the optimal (p_L^1, p_L^2) separation is too costly. Pooling is optimal if profits are greater than the retailer's outside option, $\bar{\pi}_i$ which we assume is always the case:

$$\pi_L(s_{P_0}, r_{P_0}, 1/2) \geq \bar{\pi}_L \quad C7$$

$$\pi_L(p_{P_0}^1, r_L, 1/2) \geq \bar{\pi}_L \quad C8$$

$$\pi_H(s_{P_0}, r_{P_0}, 1/2) \geq \bar{\pi}_H \quad C9$$

$$\pi_H(p_{P_0}^1, r_H, 1/2) \geq \bar{\pi}_H \quad C10$$

In a pooling equilibrium, the retailer solves an unconstrained profit maximization.

Retailer Maximizes:

$$\pi_i(p, r, 1/2) = \pi_i^1 + \delta_R \pi_i^2$$

$$\text{Max}_p \pi_L^1(p, 1/2) = [\alpha x_E + (1-\alpha)y^1](p + D - c_L)$$

$$\text{Max}_r \pi_i^2 = [(1-\alpha)y^1](p + D - c_i)$$

Since the only change is that $x_i = x_E$ for both low and high types, prices can be found by substituting (v_E, w_E) for (v_L, w_L) for the low cost type cheap separating equilibrium. Both firms price as if they were the low cost type firm in period 1. In period 2, both firms price at the optimal regular price. Both firms advertise in the sale price only format. Conditions C7-C10, can be written in terms of the parameters. There exist parameters such that these conditions hold.

Appendix 5: Proof of Propositions

Proof of Proposition 1a:

Full information deal prone demand is x_L for the low type and x_H for the high type. Pooling deal prone demand is $x_H < x_E < x_L$. Uncertainty decreases the low cost type retailer's demand and profits. If advertising a regular price reveals the low cost type, the low cost retailer demand is non-decreasing (i.e. either stays at x_L or increases from x_E to x_L). When $F_i=0$, there is no cost to advertising a regular price. Since the low cost store can only benefit ($x_E < x_L$), advertising the regular price is a weakly dominant strategy. By the same argument, advertising in the sale price only format is a weakly dominant strategy for the high cost retailer.

QED

Proposition 1b:

Show: $\pi_L(s^*_L, r^*_L) > \pi_L(p_L^{1*}, p_L^{2*})$ in a costly separating equilibrium.

Proof: In the regular/sale price format, a retailer can always imitate the sale price only prices. We show that imitation is strictly worse. Suppose $s^*_L = p_L^{1*}$ and $r_L = r^*_L = p_L^{2*}$ were optimal where r_L is the full information regular price. Profits would be

equivalent in both formats. Consider a slight decrease in r_L . As $\left. \frac{\partial \pi_L}{\partial r} \right|_{r_L} = 0$, small changes in r_L do not change profits for the low cost type. However, $\left. \frac{\partial \pi_H}{\partial r} \right|_{r_L} > 0$ in a

costly separating equilibrium and commitment to a marginally lower regular price punishes the high cost type, decreasing his profits. This relaxes the constraint, allowing the low cost type to slightly raise sale prices which leads to an increase in profits. Therefore, profits must be strictly greater in a costly separating equilibrium.

QED

Proposition 2:

To determine the optimal (s_L^*, r_L^*) the low cost store solves the following constrained maximization.

$$\begin{aligned} & \text{Max}_{s_L^*, r_L^*} \pi_L^1(s_L^*, 1) + \delta_R \pi_L^2(r_L^*) - F_L \\ \text{s.t. } & \pi_H(s_L^*, 1) + \delta_R \pi_H(r_L^*) - F_H \leq \text{Max}[\pi_H(s_H, r_H, 0), \pi_H(p_H^1, p_H^1, 0)] \\ & \pi_L(s_L^*, 1) + \delta_R \pi_L(r_L^*) - F_H \geq \text{Max}[\pi_L(s_{Po}, r_{Po}, 1/2), \pi_L(p_{Po}^1, r_L, 1/2)] \end{aligned}$$

Assume that customer beliefs and parameter values are such that pooling is not desirable and thus ignore the last constraint. Also assume that the sale price only format is optimal for type H. $\text{Max}[\pi_H(s_H, r_H, 0), \pi_H(p_H^1, p_H^2, 0)] = \pi_H(p_H^1, p_H^2, 0) = \pi_H(p_H^1, 0) + \delta \pi_H(p_H^2)$.

The FOC are:

$$\begin{aligned} s_L^*: & \frac{\partial \pi_L(s_L^*, 1)}{\partial s_L^*} + \lambda \frac{\partial \pi_H(s_L^*, 1)}{\partial s_L^*} = 0 \\ r_L^*: & \delta_R \frac{\partial \pi_L(r_L^*)}{\partial r_L^*} + \lambda \delta_R \frac{\partial \pi_H(r_L^*)}{\partial r_L^*} = 0 \\ \lambda: & \pi_H(s_L^*, 1) + \delta_R \pi_H(r_L^*) - F_i = \pi_H(p_H^1) + \delta_R \pi_H(p_H^2) \end{aligned}$$

Let (s_L, r_L) be the solution to the unconstrained maximization. Suppose that $s_L^* = s_L$. From the FOC w.r.t. s_L^* , this implies that $\lambda = 0$. From the FOC w.r.t. r_L^* , $\lambda = 0$ implies that $r_L^* = r_L$. But at (p_L, r_L) the constraint is violated in a costly separating equilibrium, contradicting our assumption that $s_L^* = s_L$ was optimal. The only solution must be for both s_L^* and r_L^* to be distorted. ($s_L^* < s_L$ and $r_L^* < r_L$)

QED

Proposition 3:

Show:

a) When the regular/sale price format has a positive expected cost due to future uncertainty, $F_i > 0$, there exists $\underline{\alpha}$ and $\bar{\alpha}$ such that the sale price only format is optimal for $\alpha < \underline{\alpha}$ and the regular sale price format is optimal for $\underline{\alpha} \leq \alpha < \bar{\alpha}$.

b) In a pooling equilibrium, the optimal selling format is the sale price only format.

Since we are interested in the optimal selling format, we need to compare optimal profits under each format. In particular, we are interested in how the optimal format changes as a function of α . There are two cases to consider:

Case 1:

Optimal Regular/Sale Price is revealing $\theta = 1$.

Optimal Sale Price Only is not revealing $\theta = 1/2$.

Since $x_L > x_E$, demand is greater in the regular sale price format for all α . If costs are zero, $F_i = 0$, the regular sale price format is optimal. When $F_i > 0$, there now exists an α such that the increased demand outweighs the increased cost and the low cost retailer prefers the regular/sale price format.

Case 2:

Optimal Regular/Sale Price is revealing $\theta = 1$.

Optimal Sale Price Only is revealing $\theta = 1$.

Commitment to a low regular price helps the low cost type because it is not always necessary to distort below the equilibrium range of $p(D)$ to signal. But for large D , all that separates the profit curves is the cost of commitment by the high cost type to the regular price. Since this is the "worst case" scenario, we will address only the case of

large D . The next claim sketches that the difference between the profit curves will have a unique maximum. (This is formally proven in the final appendix for those interested.)

Claim 2:

When $F_i = 0$, the difference in profits $\pi_L(s_L^*, r_L^*, 1) - \pi_L(p_L^{1*}, p_L^2, 1)$ has a unique maximum (quasi-concave) when the optimal price of selling to only the loyals is close to the optimal price of selling to only the deal prone consumers. ($s_L|_{\alpha=0} - s_L|_{\alpha=1} = \varepsilon \sim 0$)

Sketch of Proof:

We calculate the SOC under π_L^{RS} and π_L^{SO} and show that $\left. \frac{\partial^2 \pi_L^{SO}}{\partial \alpha^2} \right|_{\alpha_0} > \left. \frac{\partial^2 \pi_L^{RS}}{\partial \alpha^2} \right|_{\alpha_0}$

where α_0 satisfies $\frac{\partial \pi_L^{SO}}{\partial \alpha} = \frac{\partial \pi_L^{RS}}{\partial \alpha}$. We then show that

$$\left. \frac{\partial \pi_L^{SO}}{\partial \alpha} \right|_{\alpha^*(s_L, r_L)} < \left. \frac{\partial \pi_L^{RS}}{\partial \alpha} \right|_{\alpha^*(s_L, r_L)} \quad \text{and} \quad \lim_{\alpha \rightarrow \infty} \frac{\partial \pi_L^{SO}}{\partial \alpha} > \frac{\partial \pi_L^{RS}}{\partial \alpha}.$$

We then use the Poincaré-Hopf theorem (Arrow 1981) which demonstrates that the function $\pi_L^{RS} - \pi_L^{SO}$ has a unique maximum (is quasi-concave) when the above conditions are satisfied.

QED

Now consider $F_i > 0$. There are some technical problems as α approaches 1 because advertising the regular price becomes "free" when there is no one to buy in non-sale periods. In other words, at $\alpha = 1$, there is nothing supporting the regular price.

Fixed Cost Interpretation

If we assume that F_i is a fixed cost, then $\bar{\alpha} < 1$. That is, profits in either format converge to the same value as α approaches 1. If we add a fixed cost to the regular/sale price format, then the sale price only format is the optimal format for $\alpha > \bar{\alpha}$.

Flexibility Interpretation

We interpret the additional cost of the regular/sale price format as a flexibility cost. Incremental profits of $I_i(1-\alpha)^a$ can be earned if there is a stochastic shock, ρ . Thus $F_i = \rho I_i(1-\alpha)^a$. If F_i converges quickly enough to zero as α approaches 1, then $\bar{\alpha}$ approaches 1.

We will use the flexibility interpretation of F_i .

Proof:

a) At $\alpha = 0$ and $F_i = 0$, $\pi_L(p_L, r_L) = \pi_L(p_{1L}, p_{2L})$. Thus it is obvious that for $F_i > 0$, the sale price only format is optimal. The lower bound, $\underline{\alpha}$, is the smallest α satisfying $\pi_L^1(s_L^*, 1) + \delta\pi_L^2(r_L^*) - F_i = \pi_L^1(p_L^{1*}, 1) + \delta\pi_L^2(p_L^2)$. For $\alpha < \underline{\alpha}$ it is clear that the sale price only format is optimal. The upper bound, $\bar{\alpha}$, is the largest α satisfying $\pi_L^1(s_L^*, 1) + \delta\pi_L^2(r_L^*) - F_i > \pi_L^1(p_L^{1*}, 1) + \delta\pi_L^2(p_L^2)$. From Claim 2, the difference in profits is quasi-concave and has a unique maximum. Thus the profits functions cross exactly once for $\alpha < \bar{\alpha}$. There exists a constant, a , such that F_i converges to zero fast enough as α approaches 1 and $\bar{\alpha} = 1$.

b) If a retailer is in a pooling equilibrium and $F_i = 0$, profits are equal under both formats since there is no value to signaling. For $F_i > 0$, the sale price only format must be optimal.

Appendix 6: Proof of Claim 2

We prove that the difference in profit functions between the regular/sale price format and sale price only format for $\theta=1$ has a unique maximum.

Claim 2:

Let α_0 be such that $\frac{\partial \pi_L^{RS}}{\partial \alpha} = \frac{\partial \pi_L^{SO}}{\partial \alpha}$.

The function $\pi_L^{RS} - \pi_L^{SO}$ has a unique maximum at α_0 (is quasi-concave) if the equilibrium prices of selling to each segment alone are not too different. That is, $(v_b - wa) = \varepsilon \sim 0$.

Proof:

We use the following notation:

Let: $\pi_i^j(k)$ be the incremental profit from segment j to retailer i at a price k .

where: $j \in \{x, y\}$, $i \in \{L, H\}$, $k \in \{p, s, r\}$

$\pi_H^x(s_H, 0)$ be the incremental profit from segment x to retailer H at a price s_H

that is known to be type H . By definition, $s_H = p_H$.

Loyal Profit $\pi_i^y(k) = (a - bk)(k + D - c_i)$

From the definition and derivation of deal prone consumer demand we have:

$$x_L = f_L + g_L D$$

$$x_H = f_H + g_H D$$

We assumed that the inferred deal size, D , is linear in equilibrium price and therefore demand is linear. Let deal prone demand be:

$$x_L = v - wp$$

$$x_H = v_H - w_{HP}$$

Deal prone Profit when customers believe retailer is low type:

$$\pi_1^x(k) = (v - wk)(k + D - c_i)$$

Deal prone Profit when customers believe retailer is high type:

$$\pi_H^x(s_H, 0) = (v_H - w_H s_H)(s_H + D - c_H)$$

WOLG let $c_L = 0$ and $\delta_R = 1$.

Outline of Proof:

1) Derive SOC in the regular/sale price format: $\frac{\partial^2 \pi_L^{RS}}{\partial \alpha^2}$

2) Derive SOC in the sale price only format: $\frac{\partial^2 \pi_L^{SO}}{\partial \alpha^2}$

3) Show that $\left. \frac{\partial^2 \pi_L^{SO}}{\partial \alpha^2} \right|_{\alpha_0} > \left. \frac{\partial^2 \pi_L^{RS}}{\partial \alpha^2} \right|_{\alpha_0}$

4) Show that $\left. \frac{\partial \pi_L^{SO}}{\partial \alpha} \right|_{\alpha^*(s_L, r_L)} < \left. \frac{\partial \pi_L^{RS}}{\partial \alpha} \right|_{\alpha^*(s_L, r_L)}$

5) Show that $\lim_{\alpha \rightarrow 1} \frac{\partial \pi_L^{SO}}{\partial \alpha} > \frac{\partial \pi_L^{RS}}{\partial \alpha}$

1) Derive SOC for Regular/Sale Price Format

In a costly separating equilibrium for the Regular/Sale (RS) price format the retailer maximizes:

$$\text{Max}_{s,r} \pi_L^{\text{RS}} = \alpha\pi_L^x(s) + (1-\alpha)\pi_L^y(s) + (1-\alpha)\pi_L^y(r)$$

$$\text{s.t.} \quad \alpha\pi_H^x(s) + (1-\alpha)\pi_H^y(s) + (1-\alpha)\pi_H^y(r) = \alpha\pi_H^x(s_H, 0) + (1-\alpha)\pi_H^y(s_H) + (1-\alpha)\pi_H^y(r_H)$$

The FOC are:

$$\text{G:} \quad \left(\alpha \frac{\partial \pi_L^x(s)}{\partial s} + (1-\alpha) \frac{\partial \pi_L^y(s)}{\partial s} \right) - \lambda \left(\alpha \frac{\partial \pi_H^x(s)}{\partial s} + (1-\alpha) \frac{\partial \pi_H^y(s)}{\partial s} \right) = 0$$

$$\text{F:} \quad \left(\frac{\partial \pi_L^y(r)}{\partial r} \right) - \lambda \left(\frac{\partial \pi_H^y(r)}{\partial r} \right) = 0$$

$$\text{H:} \quad \left(\alpha\pi_H^x(s) + (1-\alpha)\pi_H^y(s) + (1-\alpha)\pi_H^y(r) \right) - \left(\alpha\pi_H^x(s_H, 0) + (1-\alpha)\pi_H^y(s_H) + (1-\alpha)\pi_H^y(r_H) \right) = 0$$

We apply the Implicit Function Theorem:

$$\begin{bmatrix} \frac{\partial s}{\partial \alpha} \\ \frac{\partial r}{\partial \alpha} \\ \frac{\partial \lambda}{\partial \alpha} \end{bmatrix} = \begin{bmatrix} G_s & 0 & -H_s \\ 0 & F_r & -H_r \\ H_s & H_r & 0 \end{bmatrix}^{-1} \begin{bmatrix} -G_\alpha \\ 0 \\ -H_\alpha \end{bmatrix} = D \begin{bmatrix} H_r^2 & -H_r H_s & H_s F_r \\ -H_r H_s & H_s^2 & G_s H_r \\ -H_s F_r & -G_s H_r & G_s F_r \end{bmatrix} * \begin{bmatrix} -G_\alpha \\ 0 \\ -H_\alpha \end{bmatrix}$$

$$D = (G_s H_r^2 + F_r H_s^2)^{-1}$$

$$\frac{\partial s}{\partial \alpha} = [-H_r^2 G_\alpha - H_s F_r H_\alpha] D$$

$$\frac{\partial r}{\partial \alpha} = [H_r H_s G_\alpha - H_r G_s H_\alpha] D$$

$$\frac{\partial \lambda}{\partial \alpha} = [F_r H_s G_\alpha - F_r G_s H_\alpha] D$$

Apply the Envelope Theorem for Constrained Maximization to π_L^{RS}

$$\frac{d\pi_L^{RS}}{d\alpha} = \pi_L^x(s) - \pi_L^y(s) - \pi_L^y(r) - \lambda \left\{ \left(\pi_H^x(s) - \pi_H^y(s) - \pi_H^y(r) \right) - \left(\pi_H^x(s_H, 0) - \pi_H^y(s_H) - \pi_H^y(r_H) \right) \right\}$$

Taking the Second Derivative:

$$\frac{d^2 \pi_L^{RS}}{d\alpha^2} = \frac{\partial s}{\partial \alpha} \left[\frac{\partial \pi_L^x(s)}{\partial s} - \frac{\partial \pi_L^y(s)}{\partial s} \right] - \lambda \left\{ \frac{\partial s}{\partial \alpha} \left[\frac{\partial \pi_H^x(s)}{\partial s} - \frac{\partial \pi_H^y(s)}{\partial s} \right] - \frac{\partial s_H}{\partial \alpha} \left[\frac{\partial \pi_H^x(s_H, 0)}{\partial s_H} - \frac{\partial \pi_H^y(s_H)}{\partial s_H} \right] \right\}$$

$$- \frac{\partial \lambda}{\partial \alpha} \left\{ \pi_H^x(s) - \pi_H^y(s) - \pi_H^y(r) - \pi_H^x(s_H, 0) + \pi_H^y(s_H) + \pi_H^y(r_H) \right\}$$

Rearranging Terms we get:

$$\frac{\partial^2 \pi_L^{RS}}{\partial \alpha^2} = \frac{\partial s}{\partial \alpha} G_\alpha - \frac{\partial \lambda}{\partial \alpha} H_\alpha + \lambda \left\{ \frac{\partial s_H}{\partial \alpha} \left[\frac{\partial \pi_H^x(s_H, 0)}{\partial s_H} - \frac{\partial \pi_H^y(s_H)}{\partial s_H} \right] \right\}$$

$$\frac{\partial^2 \pi_L^{RS}}{\partial \alpha^2} = D[-H_r^2 G_\alpha^2 - 2G_\alpha H_\alpha H_s F_r + H_\alpha^2 G_s F_r] + \lambda \left\{ \frac{\partial s_H}{\partial \alpha} \left[\frac{\partial \pi_H^x(s_H, 0)}{\partial s_H} - \frac{\partial \pi_H^y(s_H)}{\partial s_H} \right] \right\}$$

Why Advertise Both Regular and Sale Prices?

$$H_r > 0 \quad H_r = [a-b(2r-c_H)]$$

$$F_r < 0 \quad -2b(1-\lambda)$$

$$G_s < 0 \quad -2[\alpha w + (1-\alpha)b](1-\lambda)$$

$$H_s > 0 \quad \text{FOC w.r.t } s \text{ which are positive since } s < s_H$$

$$\left(\alpha \frac{\partial \pi_H^x(s)}{\partial s} + (1-\alpha) \frac{\partial \pi_H^y(s)}{\partial s} \right) = (\alpha v + (1-\alpha)a) - (\alpha w + (1-\alpha)b)(2s + D - c_H)$$

G_α and H_α :

$$H_\alpha > 0 \quad H_\alpha = \left[\pi_H^x(s) - \pi_H^y(s) - \pi_H^x(r) - \pi_H^x(s_H, 0) + \pi_H^y(s_H) + \pi_H^y(r_H) \right]$$

$$= \frac{-1}{\alpha} \left[\pi_H^y(s) - \pi_H^y(s_H) + \pi_H^y(r) - \pi_H^y(r_H) \right]$$

$$G_\alpha \quad G_\alpha = \left[\frac{\partial \pi_L^x(s)}{\partial s} - \frac{\partial \pi_L^y(s)}{\partial s} - \lambda \left(\frac{\partial \pi_H^x(s)}{\partial s} - \frac{\partial \pi_H^y(s)}{\partial s} \right) \right]$$

$$= \frac{-1}{\alpha} \left[\frac{\partial \pi_L^y(s)}{\partial s} - \lambda \frac{\partial \pi_H^y(s)}{\partial s} \right]$$

$D < 0$ by $G_s < 0$, and $F_r < 0$

We can simplify G_α :

Let:

$$\tilde{f} = (\alpha v + (1-\alpha)a), \tilde{g} = (\alpha w + (1-\alpha)b)$$

Rewriting G_α we get:

$$\begin{aligned}
 -\alpha G_\alpha &= (a - b(2s + D)) - \frac{(\tilde{f} - \tilde{g}(2s + D))}{(\tilde{f} - \tilde{g}(2s + D - c_H))} (a - b(2s + D - c_H)) \\
 &= \frac{(a - b(2s + D))(\tilde{f} - \tilde{g}(2s + D - c_H)) - (\tilde{f} - \tilde{g}(2s + D))(a - b(2s + D - c_H))}{H_s} \\
 &= \frac{(a - bs_L)(\tilde{f} - \tilde{g}s_H) - (\tilde{f} - \tilde{g}s_L)(a - bs_H)}{H_s} \\
 &= \frac{(a\tilde{f} - a\tilde{g}s_H - b\tilde{f}s_L + b\tilde{g}s_Hs_L) - (\tilde{f}a - \tilde{f}bs_H - a\tilde{g}s_L + b\tilde{g}s_Hs_L)}{H_s} \\
 &= \frac{(-a\tilde{g}(s_H - s_L) - b\tilde{f}(s_L - s_H))}{H_s} = \frac{(s_H - s_L)(b\tilde{f} - a\tilde{g})}{H_s} = \frac{(c_H)(-b\tilde{f} + a\tilde{g})}{H_s} \\
 &= \frac{(c_H)(-b(\alpha v + (1 - \alpha)a) + a(\alpha w + (1 - \alpha)b))}{H_s} = \frac{(c_H)\alpha(aw - bv)}{H_s}
 \end{aligned}$$

therefore:

$$G_\alpha = \frac{c_H(bv - aw)}{H_s}$$

2) Derive SOC in Sale Price Only Format:

In a costly separating equilibrium for the Sale Price Only (SO) price format the retailer maximizes:

$$\begin{aligned}
 \text{Max } \pi_L^{\text{SO}} &= \alpha\pi_L^x(p) + (1 - \alpha)\pi_L^y(p) + (1 - \alpha)\pi_L^y(r) \\
 \text{s.t. } \alpha\pi_H^x(p) + (1 - \alpha)\pi_H^y(p) &= \alpha\pi_H^x(p_H, 0) + (1 - \alpha)\pi_H^y(p_H)
 \end{aligned}$$

The FOC are:

Why Advertise Both Regular and Sale Prices?

$$J: \left(\alpha \frac{\partial \pi_L^x(p)}{\partial p} + (1-\alpha) \frac{\partial \pi_L^y(p)}{\partial p} \right) - \mu \left(\alpha \frac{\partial \pi_H^x(p)}{\partial p} + (1-\alpha) \frac{\partial \pi_H^y(p)}{\partial p} \right) = 0$$

$$K: \left(\alpha \pi_H^x(p) + (1-\alpha) \pi_H^y(p) \right) - \left(\alpha \pi_H^x(p_H, 0) + (1-\alpha) \pi_H^y(p_H) \right) = 0$$

$$\text{and } \frac{\partial \pi_L^y(r)}{\partial r} = 0$$

We apply the Implicit Function Theorem:

$$\begin{bmatrix} \frac{\partial p}{\partial \alpha} \\ \frac{\partial \mu}{\partial \alpha} \end{bmatrix} = \begin{bmatrix} J_p & -K_p \\ K_p & 0 \end{bmatrix}^{-1} * \begin{bmatrix} -J_\alpha \\ -K_\alpha \end{bmatrix} = D \begin{bmatrix} 0 & K_p \\ -K_p & J_p \end{bmatrix} * \begin{bmatrix} -J_\alpha \\ -K_\alpha \end{bmatrix}$$

$$D = (K_p^2)^{-1}$$

$$\frac{\partial p}{\partial \alpha} = [-K_p \ K_\alpha] D = -K_\alpha / K_p$$

$$\frac{\partial \mu}{\partial \alpha} = [K_p \ J_\alpha - J_p \ K_\alpha] D = \frac{J_\alpha}{K_p} - \frac{J_p K_\alpha}{K_p^2}$$

Apply the Envelope Theorem for Constrained Maximization to π_L^{SO}

$$\frac{d\pi_L^{SO}}{d\alpha} = \pi_L^x(p) - \pi_L^y(p) - \pi_L^y(r_L) - \mu \left\{ \left(\pi_H^x(p) - \pi_H^y(p) \right) - \left(\pi_H^x(p_H, 0) - \pi_H^y(p_H) \right) \right\}$$

Taking the Second Derivative:

$$\frac{d^2\pi_L^{SO}}{d\alpha^2} = \frac{\partial p}{\partial \alpha} \left[\frac{\partial \pi_L^x(p)}{\partial p} - \frac{\partial \pi_L^y(p)}{\partial p} \right] - \mu \left\{ \frac{\partial p}{\partial \alpha} \left[\frac{\partial \pi_H^x(p)}{\partial p} - \frac{\partial \pi_H^y(p)}{\partial p} \right] - \frac{\partial p_H}{\partial \alpha} \left[\frac{\partial \pi_H^x(p_H, 0)}{\partial p_H} - \frac{\partial \pi_H^y(p_H)}{\partial p_H} \right] \right\} - \frac{\partial \mu}{\partial \alpha} \left\{ \pi_H^x(p) - \pi_H^y(p) - \pi_H^x(p_H, 0) + \pi_H^y(p_H) \right\}$$

Rearranging Terms we get:

$$\frac{\partial^2\pi_L^{SO}}{\partial \alpha^2} = \frac{\partial p}{\partial \alpha} J_\alpha - \frac{\partial \mu}{\partial \alpha} K_\alpha - \mu \left\{ - \frac{\partial p_H}{\partial \alpha} \left[\frac{\partial \pi_H^x(p_H, 0)}{\partial p_H} - \frac{\partial \pi_H^y(p_H)}{\partial p_H} \right] \right\}$$

$$\frac{\partial^2\pi_L^{SO}}{\partial \alpha^2} = -2 \frac{J_\alpha K_\alpha}{K_p} + J_p \left(\frac{K_\alpha}{K_p} \right)^2 + \mu \left\{ \frac{\partial p_H}{\partial \alpha} \left[\frac{\partial \pi_H^x(p_H, 0)}{\partial p_H} - \frac{\partial \pi_H^y(p_H)}{\partial p_H} \right] \right\}$$

$$J_p < 0: \quad J_p = -2[\alpha w + (1-\alpha)b](1-\mu)$$

$K_p > 0$: K_p = Full Information FOC w.r.t p for type H, if $p < p_H$ this is positive.

$$\left(\alpha \frac{\partial \pi_H^x(p)}{\partial p} - (1-\alpha) \frac{\partial \pi_H^y(p)}{\partial p} \right) = (\alpha v + (1-\alpha)a) - (\alpha w + (1-\alpha)b)(2p + D - c_H)$$

K_α and J_α are monotonic and don't change sign.

$$K_\alpha > 0 \quad K_\alpha = \left[\pi_H^x(p) - \pi_H^y(p) - \pi_H^x(p_H, 0) + \pi_H^y(p_H) \right] = \frac{-1}{\alpha} \left[\pi_H^y(p) - \pi_H^y(p_H) \right]$$

Why Advertise Both Regular and Sale Prices?

$$\begin{aligned}
 J_{\alpha} \quad J_{\alpha} &= \left[\left(\frac{\partial \pi_L^x(p)}{\partial p} - \frac{\partial \pi_L^y(p)}{\partial p} \right) - \mu \left(\frac{\partial \pi_H^x(p)}{\partial p} - \frac{\partial \pi_H^y(p)}{\partial p} \right) \right] \\
 &= \frac{-1}{\alpha} \left[\frac{\partial \pi_L^y(p)}{\partial p} - \mu \frac{\partial \pi_H^y(p)}{\partial p} \right]
 \end{aligned}$$

Similar to G_{α} , we can show that $J_{\alpha} = \frac{c_H(bv - aw)}{K_p}$.

Identities and Relationships:

$$\lambda = \frac{\frac{\partial \pi_L^Y(r)}{\partial r}}{\frac{\partial \pi_H^Y(r)}{\partial r}} = \frac{\alpha \frac{\partial \pi_L^X(s)}{\partial s} + (1-\alpha) \frac{\partial \pi_L^Y(s)}{\partial s}}{\alpha \frac{\partial \pi_H^X(s)}{\partial s} + (1-\alpha) \frac{\partial \pi_H^Y(s)}{\partial s}} = \frac{H_r - bc_H}{H_r} = \frac{H_s - (\alpha w + (1-\alpha)b)c_H}{H_s}$$

$$\mu = \frac{\alpha \frac{\partial \pi_L^X(p)}{\partial p} + (1-\alpha) \frac{\partial \pi_L^Y(p)}{\partial p}}{\alpha \frac{\partial \pi_H^X(p)}{\partial p} + (1-\alpha) \frac{\partial \pi_H^Y(p)}{\partial p}} \frac{K_p - (\alpha w + (1-\alpha)b)c_H}{K_p}$$

(1) $(1-\lambda)H_s = (1-\mu)K_p$

(2) $H_s b = H_r(\alpha w + (1-\alpha)b)$

(3) $(1-\lambda)J_p = (1-\mu)G_s$

(4) $H_r G_s = F_r H_s$

(5) $G_s H_s = J_p K_p$

(6) $J_\alpha K_p = G_\alpha H_s$

We can also show that:

(7) $H_r + H_s > K_p$

(8) $K_p - H_s = 2(\alpha w + (1-\alpha)b)(s-p) > 0$

(9) $K_p = 2(\alpha w + (1-\alpha)b)(s_H - p)$

(10) $H_s = 2(\alpha w + (1-\alpha)b)(s_H - s)$

(11) $H_r = 2b(r_H - r)$

(12) $s_H - s = r_H - r$

$$\text{And at } \frac{\partial \pi_L^{RS}}{\partial \alpha} = \frac{\partial \pi_L^{SO}}{\partial \alpha}$$

$$\begin{aligned} & \pi_L^x(s) - \pi_L^y(s) - \pi_L^y(r) - \lambda \left\{ \pi_H^x(s) - \pi_H^y(s) - \pi_H^y(r) - \pi_H^x(s_H, 0) + \pi_H^y(s_H) + \pi_H^y(r_H) \right\} \\ & = \pi_L^x(p) - \pi_L^y(p) - \pi_L^y(r_L) - \mu \left\{ \pi_H^x(p) - \pi_H^y(p) - \pi_H^x(p_H, 0) + \pi_H^y(p_H) \right\} \\ & \pi_L^x(s) - \pi_L^y(s) - \pi_L^y(r) - \pi_L^x(p) + \pi_L^y(p) + \pi_L^y(r_L) - \lambda H_\alpha = -\mu K_\alpha \\ \text{and } & \pi_L^x(s) = \pi_H^x(s) + c_H(v - ws) \quad \pi_L^x(p) = \pi_H^x(p) + c_H(v - wp) \\ & \pi_L^x(s) - \pi_L^x(p) = \pi_H^x(s) - \pi_H^x(p) - wc_H(s - p) \\ & \pi_L^y(s) - \pi_L^y(p) = \pi_H^y(s) - \pi_H^y(p) - bc_H(s - p) \\ & H_\alpha(1 - \lambda) = K_\alpha(1 - \mu) + c_H \left[(w - b)(s - p) + (r_H - r)b \right] \end{aligned}$$

Claim 3: $0 < \lambda < 1, 0 < \mu < 1, \lambda < \mu$

Proof:

$$\lambda = \frac{(\alpha v + (1 - \alpha)a) - (\alpha w + (1 - \alpha)b)(2s + D)}{(\alpha v + (1 - \alpha)a) - (\alpha w + (1 - \alpha)b)(2s + D - c_H)} < 1$$

$$\mu = \frac{(\alpha v + (1 - \alpha)a) - (\alpha w + (1 - \alpha)b)(2p + D)}{(\alpha v + (1 - \alpha)a) - (\alpha w + (1 - \alpha)b)(2p + D - c_H)} < 1$$

$$(\mu - \lambda) = \frac{2c_H(s - p)(\alpha w + (1 - \alpha)b)^2}{H_s K_p} > 0 \text{ since } s > p, H_s > 0 \text{ and } K_p > 0.$$

Claim 4: Profits are equal at $\alpha = 1$ and $\alpha = \alpha^*_{(sL, rL)}$

Proof: Obvious.

Claim 5: $|J_p| < |G_s|$

Proof: Obvious given the definition and that $\mu > \lambda$.

Claim 6: $K_p > H_s$.

Proof: Given that $s_L > s > p$, the result is obvious.

Claim 7:

$$\frac{\alpha H_\alpha}{H_s} = \frac{a - b(D - c_H) - b(s_H + s) + a + bc_H - b(r_H + r)}{2(\alpha w + (1 - \alpha)b)}$$

$$\frac{\alpha K_\alpha}{K_p} = \frac{a - b(D - c_H) - b(s_H + p)}{2(\alpha w + (1 - \alpha)b)}$$

$$\frac{\alpha H_\alpha}{H_s} - \frac{\alpha K_\alpha}{K_p} = \frac{-b(s - p) + a + bc_H - b(r_H + r)}{2(\alpha w + (1 - \alpha)b)}$$

Proof: Use definition of H_α and K_α and identities.

Claim 8:

$$\alpha \frac{H_\alpha}{H_s} - \alpha \frac{K_\alpha}{K_p} = \frac{2\alpha[(w - b)(s - p) + (r_H - r)b]}{2(\alpha w + (1 - \alpha)b)} \text{ when } \frac{\partial \pi_L^{RS}}{\partial \alpha} = \frac{\partial \pi_L^{SO}}{\partial \alpha}.$$

Proof: Rearrange FOC using identities.

Claim 9:

$$2\alpha[(w - b)(s - p) + (r_H - r)b] = -b(s - p) + b(r_H - r) \text{ when } \frac{\partial \pi_L^{RS}}{\partial \alpha} = \frac{\partial \pi_L^{SO}}{\partial \alpha}.$$

Proof: Use claims 7 and 8.

By definition, $s_H = p_H$.

$$p_H = \frac{\alpha v_H + (1 - \alpha)a}{2(\alpha g_H + (1 - \alpha)b)} + \frac{c_H}{2} - \frac{D}{2}$$

$$\frac{\partial p_H}{\partial \alpha} = \frac{bv_H - aw_H}{2(\alpha g_H + (1 - \alpha)b)^2}$$

$$\frac{\partial \pi_H^x(p_H, 0)}{\partial p_H} = \frac{(bv_H - aw_H)}{(\alpha g_H + (1 - \alpha)b)}$$

We let:

$$\begin{aligned} Z &= (\mu - \lambda) \left\{ \frac{\partial p_H}{\partial \alpha} \left[\frac{\partial \pi_H^x(p_H, 0)}{\partial p_H} \right] \right\} = (\mu - \lambda) \frac{(bv_H - aw_H)^2}{2(\alpha w_H + (1 - \alpha)b)^3} \\ &= \frac{2c_H(s - p)(\alpha w + (1 - \alpha)b)^2}{H_s K_p} \frac{(bv_H - aw_H)^2}{2(\alpha w_H + (1 - \alpha)b)^3} \end{aligned}$$

3) Show that SOC for sale only > SOC reg/sale

$$\begin{aligned} \left. \frac{\partial^2 \pi_L^{SO}}{\partial \alpha^2} \right|_{\alpha_0} &> \left. \frac{\partial^2 \pi_L^{RS}}{\partial \alpha^2} \right|_{\alpha_0} \\ -2 \frac{J_\alpha K_\alpha}{K_p} + J_p \left(\frac{K_\alpha}{K_p} \right)^2 + Z &> D[-H_r^2 G_\alpha^2 - 2G_\alpha H_\alpha H_{sr} F + H_\alpha^2 G_{sr} F] \end{aligned}$$

Using the definition of J_α and G_α and noting that $bv - fb = \varepsilon \sim 0$ by assumption, the above inequality reduces to:

$$J_p \left(\frac{K_\alpha}{K_p} \right)^2 + Z > D H_\alpha^2 G_{sr} F$$

Using identity (4), we get:

$$J_p \left(\frac{K_\alpha}{K_p} \right)^2 + Z > \frac{H_\alpha^2 G_{sr} F}{G_s H_r^2 + F_r H_s^2} = \frac{H_\alpha^2 G_s}{H_s (H_s + H_r)}$$

And as $Z > 0$, we ignore it.

$$0 > J_p \left(\frac{K_\alpha}{K_p} \right)^2 > \frac{H_\alpha^2 G_s}{H_s (H_s + H_r)}$$

$$\frac{(H_s + H_r) J_p K_\alpha^2}{K_p} > \frac{J_p (1-\lambda)^2 H_\alpha^2}{K_p (1-\mu)^2}$$

From the FOC being equal, a local condition is:

$$\frac{(1-\lambda)}{(1-\mu)} H_\alpha = K_\alpha + \frac{c_H \left[(w-b)(s-p) + (r_H - r)b \right]}{(1-\mu)} = K_\alpha + L$$

Substituting:

$$\frac{(H_s + H_r) J_p K_\alpha^2}{K_p} > \frac{J_p}{K_p} (K_\alpha + L)^2$$

$$\frac{(H_s + H_r)}{K_p} K_\alpha^2 < (K_\alpha + L)^2$$

$$\sqrt{\frac{(H_s + H_r)}{K_p}} K_\alpha < (K_\alpha + L)$$

Using the identities as well as $vb-wa = \varepsilon \sim 0$ we can rewrite:

$$L = \frac{c_H \left[(w-b)(s-p) + b(r_H - r) \right]}{(1-\mu)} = \frac{\left[(w-b)(s-p) + b(r_H - r) \right] K_p}{(\alpha w + (1-\alpha)b)}, \alpha \frac{K_\alpha}{K_p} = \frac{b(s_H - p)}{2(\alpha w + (1-\alpha)b)}$$

$$\frac{L}{K_\alpha} = \frac{2\alpha \left[(w-b)(s-p) + b(r_H - r) \right]}{b(s_H - p)}$$

Using the identities we can simplify this to:

Why Advertise Both Regular and Sale Prices?

$$\frac{L}{K_\alpha} = \frac{2\alpha[(w-b)(s-p) + b(r_H - r)]}{b(s_H - p)} = \frac{[(r_H - r) - (s-p)]}{(s_H - p)}$$

$$\frac{L}{K_\alpha} + 1 = \frac{[(s_H - s) - (s-p)] + (s_H - p)}{(s_H - p)} = \frac{2(s_H - s)}{(s_H - p)}$$

Substituting:

$$\sqrt{\frac{(H_s + H_r)}{K_p}} < \frac{L}{K_\alpha} + 1$$

$$\sqrt{\frac{(H_s + H_r)}{K_p}} < \frac{2(s_H - s)}{(s_H - p)} = \frac{2H_s}{K_p}$$

Which is true given that $H_s > H_r$ and $H_s + H_r > K_p$.

This demonstrates that

$$\left. \frac{\partial^2 \pi_L^{SO}}{\partial \alpha^2} \right|_{\alpha_0} > \left. \frac{\partial^2 \pi_L^{RS}}{\partial \alpha^2} \right|_{\alpha_0}$$

To complete the proof we need to show that:

Claim 10:

$$\left. \frac{\partial \pi_L^{SO}}{\partial \alpha} \right|_{\alpha^*(s_L, r_L)} < \left. \frac{\partial \pi_L^{RS}}{\partial \alpha} \right|_{\alpha^*(s_L, r_L)}$$

Proof:

Consider the point at which deviation by the high type just occurs in the regular/sale price format. At this point, the low type can lower the regular price and sale price by small amounts, ϵ , in the regular/sale price format and incur no change in profits.

From proposition 1, the high type is already deviating in the sale price only format. Small changes in price to meet this constraint lead to a loss in profits.

Claim 11:

$$\lim_{\alpha \rightarrow 1} \frac{\partial \pi_L^{SO}}{\partial \alpha} > \frac{\partial \pi_L^{RS}}{\partial \alpha}$$

Proof:

Using the Poincaré-Hopf theorem it is clear that the difference in the functions has a unique maximum. If $F_i = \rho I_i (1-\alpha)^a$ is strictly decreasing in α , we can select the decay constant, a , such that the cost of commitment decreases rapidly and the regular/sale price format is preferred $\underline{\alpha} < \alpha < \bar{\alpha} \cong 1$.

QED

References

- Anderson, Erin (1985), "The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis," *Marketing Science*, 4, 234-254.
- Banks, Jeffrey and Sridhar Moorthy (1994), "A Model of Price Promotions with Consumer Search," Unpublished Working Paper.
- Bawa, Kapil, and Robert W. Shoemaker, (1987) "The Coupon-Prone Consumer: Some Findings Based on Purchase Behavior Across Product Classes," *Journal of Marketing*, vol. 51, no. 4, October, pp. 99-110.
- Ben-Zion, U. Rapoport, A., and Yagil, J., (1989), "Discount Rates Inferred From Decisions: An Experimental Study," *Management Science*, 35, pp. 270-284.
- Blattberg, Robert C. and Scott S. Neslin, (1990), *Sales Promotion: Concepts Methods and Strategies*, Prentice Hall.
- Blattberg, Robert C., Thomas Buesing, Peter Peacock, and Subrata K. Sen, (1978) "Identifying the Deal prone Segment," *Journal of Marketing Research*, vol. 15, no. 3, August, pp. 369-377.
- Blattberg, Robert C., and Subrata K. Sen, (1976) "Market Segments and Stochastic Brand Choice Models," *Journal of Marketing Research*, vol. 13, no. 1, February, pp. 34-45.
- Blattberg, Robert C., and Subrata K. Sen, (1974) "Market Segmentation Using Models of Multidimensional Purchasing Behavior," *Journal of Marketing*, vol., 38, no. 4, October, pp. 17-28.
- Cho, I.K., and D. M. Kreps (1987), "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, pp. 179-221.
- Cox, Anthony D. and Dena Cox (1990), "Competing on Price: The Role of Retail Price Advertisements in Shaping Store-Price Image," *Journal of Retailing*, Vol. 66, No. 4, Winter, pp. 428-445.
- Fudenberg, Drew and Jean Tirole (1991), *Game Theory*, MIT Press, Cambridge Massachusetts, Chapter 11.
- Gerstner, Eitan and James D. Hess, (1991) "A Theory of Channel Price Promotions," *The American Economic Review*, vol. 81, no. 4, September, pp. 872-886.

- Jacquemin, Alexis P. and Charles H. Berry (1979), "Entropy Measure of Diversification and Corporate Growth," *Journal of Industrial Economics*, June, Vol. 17, No. 4, pp. 359-369.
- Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, XLVII, pp. 263-91.
- Klein, B. and K. Leffler, (1981), "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 81, pp. 615-641.
- Kihlstrom, R. and M. Riordan (1984), "Advertising as a Signal," *Journal of Political Economy*, 92, pp. 427-450.
- Lal, Rajiv (1990), "Manufacturer Trade Deals and Retail Price Promotions," *Journal of Marketing Research*, vol. 27, November, pp. 428-444.
- Narasimhan, Chakravarthi, (1988) "Competitive Promotional Strategies," *Journal of Business*, Vol. 61, no. 4, October, pp. 128-146.
- Narasimhan, Chakravarthi, (1984) "A Price Discrimination Theory of Coupons," *Marketing Science*, vol. 3, no. 2, Spring, pp. 128-146.
- McGuire and Staelin (1983), an Industry Equilibrium Analysis of Downstream Vertical Integration," *Marketing Science*, vol. 2, no. 2, Spring, pp. 161-191.
- Milgrom, P. and J. Roberts (1986), "Prices and Advertising Signals of Product Quality," *Journal of Political Economy*, 94, pp. 796-821.
- Monroe, Kent (1990), *Pricing: Making Profitable Decisions*, McGraw Hill, New York.
- Montgomery, Cynthia and Birger Wernerfelt,(1992), "Risk Reduction and Umbrella Branding," *Journal of Business*, vol. 65, no 1.
- Moorthy, Sridhar, and Kannan Srinivasan (1994), "Signaling quality with a money-back guarantee: the role of transaction costs," Unpublished Working Paper.
- Nelson, Phillip (1970), "Information and Consumer Behavior," *Journal of Political Economy*, pp. 311-329.
- Nelson, Phillip (1974), "Advertising as Information," *Journal of Political Economy*, 78, pp. 311-329.

Why Advertise Both Regular and Sale Prices?

- Raju, Jagmohan S., V. Srinivasan, and Rajiv Lal, "The Effects of Brand Loyalty on Competitive Price Promotional Strategies," *Management Science*, vol. 36, no. 3, March, pp. 276-304.
- Schmalensee, Richard, (1978), "A Model of Advertising and Product Quality," *Journal of Political Economy*, 86, pp. 485-503.
- Shapiro, Carl (1983) "Premiums for High Quality Products as Rents to Reputation," *Quarterly Journal of Economics*, 98, pp. 659-680.
- Simester, Duncan (1994), "Signaling Price Image using Advertised Prices," Forthcoming in *Marketing Science*.
- Simester, Duncan (1993), "Analytic Essays in Marketing," Unpublished Ph.D. Dissertation, MIT Sloan School of Management.
- Tellis, Gerald and Birger Wernerfelt (1987), "Competitive Price and Quality Under Asymmetric Information," *Marketing Science*, Vol. 6, No. 3, Summer, pp. 240-253.
- Thaler, R. (1981a), "Some Empirical Evidence on Dynamic Inconsistency," *Economic Letters*, 8, pp. 201-207.
- Thaler, R., and Shefrin, H.M., (1981b), "An Economic Theory of Self Control," *Journal of Political Economy*, 89, pp. 392-406.
- Thaler, R., (1985) "Mental Accounting and Consumer Choice," *Marketing Science*, Vol. 4, No. 3, Summer, pp. 199-214.
- Urbany, Joel E., William O. Bearden, and Dan C. Weilbaker, "The Effect of Plausible and Exaggerated Reference Prices on Consumer Perceptions and Price Search," *Journal of Consumer Research*, Vol. 15, June, pp. 95-110.
- Varian, Hal R., (1981) "Dynamical Systems with Applications to Economics," in Handbook of Mathematical Economics, Volume 1, Chapter 3, North-Holland Publishing Company.
- Webster, Frederick E., Jr., (1965), "The Deal Prone Consumer," *Journal of Marketing Research*, Vol. 2, no. 2, May, pp. 186-189.
- Wernerfelt, Birger (1994a) "Efficiency Criterion for Marketing Design," *Journal of Marketing Research*, November.

Wernerfelt, Birger (1994b) "Selling Formats for Search Goods," forthcoming *Marketing Science*.

Wernerfelt, Birger (1988) "Umbrella Branding as a Signal of New Product Quality: An Example of Signaling by Posting a Bond," *Rand Journal of Economics*, 19, pp. 458-66.

Winer, Russell S. (1986), "A Reference Price Model of Brand Choice for Frequently Purchased Products," *Journal of Consumer Research*, Vol. 13, September, pp. 250-256.

Why Advertise Both Regular and Sale Prices?

Essay 3

Long-Term Effects of Trade Promotion: Brand Advertising or Brand Erosion?

Introduction

Packaged goods manufacturers spent roughly \$60 billion dollars on trade promotions in 1993. (Spethmann, 1995) Recently, there has been considerable discussion in the industry about the distribution of promotional dollars between trade promotion and advertising. The evidence shows that the proportion of dollars spent on trade promotions continues to increase relative to advertising and consumer promotion. Over the last five years, trade promotion budgets increased 70% relative to a 37% increase in advertising budgets and a 35% increase in consumer promotion budgets. (Spethmann, 1995) Trade promotions are often seen as a short-term strategy which generates an immediate sale. If there is any long-term benefit, it is only to the extent that a customer makes a repeat purchase. In contrast, advertising is often viewed as a long-term strategy and an investment in brand equity. Of course one reason firms spend more on trade promotions is that there is an immediate impact on market share. It has been much more challenging to measure the effects of advertising and incorporate them into the “bottom line.” But many manufacturers see the increased reliance on trade promotion as myopic behavior which is detrimental to the long-term health of a brand. Proctor and Gamble (P&G), for example has spent considerable effort streamlining their brands and pulling back from heavy trade promotion spending¹. In addition, P&G has shifted some of its trade promotion budget to retailer specific advertising. While the above story oversimplifies some of the broader issues, the basic story for many manufacturers is that trade promotions are a necessary evil. Many manufacturers believe that a shift in promotional spending from advertising toward trade promotion focuses a brand on short-term objectives and erodes brand equity.

A main objective of this paper is to explore whether trade promotions have a long-term benefit which to this point has been largely unrecognized. We hypothesize that repeated brief exposure to the product via store displays and store features increases the probability of repeat purchase. In other words, multiple exposures to previous store displays and features are simple forms of advertising which may have a long-term positive impact. Thus we have two competing hypotheses. The advertising theory suggests that trade promotions have a positive effect which leads to future sales. The brand erosion theory suggests that these same exposures damage brand

¹ Inventory management and utilization of manufacturing facilities are also critical factors in P&G's decision to reduce trade promotion spending.

on deal. The result is that a customer will make fewer purchases at the regular price which has a negative effect on the brand.

The theory of a negative promotional carry-over effect has been challenged by other authors. For example, Bawa and Shoemaker (1987) find no carryover effect. Neslin and Shoemaker (1989) take a closer look at the findings of Dodson et al. (1978) and show that looking at the aggregate brand repurchase probability may be misleading. The authors demonstrate that household heterogeneity (i.e. separating brand-loyal and switching customers) may result in misleading aggregate repeat purchase probabilities. By looking at individual households, the negative carry-over effect is diminished. Davis, Inman, and McAlister (1992) use an experimental approach to examining this issue which aims at better understanding consumer's brand evaluation. The authors measure brand evaluation among a random sample of students before and after a three month promotional period at the student store. The authors reject the hypothesis that the change in brand evaluation is negative. Unfortunately, the authors do not collect panel data and are therefore unable to separate household specific effects. In sum, current research suggests that there is a non-positive (no-effect or negative) carry-over effect of past promotions.

A natural question to ask is why would one believe that displays and features have a long-term positive effect? We find two recent pieces of evidence to suggest that there may be a positive long-term effect. First, recent work on coupons and FSIs (Free Standing Inserts) suggests that they may have advertising effects. (Srinivasan, Leone, and Mulhern (1994), Leclerc and Little (1995)). FSIs have traditionally been viewed as executional devices for delivering coupons³. However, recent work by Leclerc and Little suggests that FSIs may also have an advertising effect. The authors use an experimental setting to test alternative FSI designs and find differences in recall, recognition, and brand attitude. Srinivasan et al (1994) find evidence of advertising effects for FSIs using scanner data. To estimate the advertising effect, they examine coupons delivered in FSIs and compare coupon redeemers to non-redeemers. They find evidence of an advertising effect in two of six categories. Existence of an advertising effect for FSIs and coupons suggests that there may be an advertising-like effect in other types of promotions.

³ FSIs with no coupon attached sometimes appear and so may be considered purely advertising. However, it has been suggested these coupon-less FSIs are due to lack of funds in the promotion budget for a coupon drop where the FSI had already been purchased.

image which hurts future sales. We will attempt to shed light on these competing hypotheses using scanner panel data.

An immediate point to be addressed are the scope of our study and our definition of trade promotion. There are many different types of trade deals such as off-invoice, bill-back, cooperative advertising allowances, display allowances, slotting allowances, street money, etc. (Blattberg and Neslin, 1991) We are primarily interested in focusing on what consumers observe in the store: price cuts, store displays, and store features (weekly store fliers and newspapers).

The bundled nature of trade deals makes it difficult to estimate how much manufacturers spend to obtain price points, displays, and features. For example, a manufacturer may pay a retailer a lump sum for display space, feature space, and specific price points during a given week.² However, it is estimated that in 1993, approximately 47% of trade promotion expenditure was for off-invoice allowances. (Spethmann, 1995) A conservative estimate indicates that tens of billions of dollars are spent annually by manufacturers to obtain store displays and features.

The prevailing hypotheses in the academic literature are that excessive trade promotions have a negative or neutral impact. One of the first papers to suggest that trade promotions have a negative impact was by Dodson, Tybout, and Sternthal (1978). The authors use self-perception theory to explain the negative effect of past promotions. This theory suggests that consumers will examine the conditions under which a purchase occurs in evaluating their attitude toward the purchase. When a consumer previously bought on deal, attitude toward the brand is clouded by a liking of the price reduction. In contrast, when a consumer previously bought at the regular price, there is less interference in attitude toward the brand. The authors conclude that previous purchases on deal lead to less positive attitude toward the brand. An alternative argument for a negative effect is that excessive trade promotions lower a customer's price expectation. (i.e. reference price) In a prospect theory framework, decreased price expectations places the customer in a "loss" framework when facing a non-promotional purchase decision. Thus, a customer is less likely to buy at the regular price. In addition, frequent dealing may increase a customer's expectation of the probability of a future deal and influence a customer to postpone a purchase until the next deal cycle. This is often referred to as customers becoming "trained" to buy

² Estimates on the retailer side indicate 1993 promotional revenues of \$19.3 billion for POP displays and \$6.2 billion for specialty advertising. (Spethmann, 1995) There was no indication as to how retailers were able to break down revenues into these categories. As always, caveat emptor.

The second piece of evidence that suggests there may be a long-term effect is the nature of household shopping behavior. Scanner data studies (Uncles (1991)⁴, Bucklin and Gupta (1992)⁵) find that households visit supermarkets frequently and an average household⁶ (consumer) may visit a supermarket more than twice per week. On each store visit, consumers may be exposed to brands on display and feature. In addition, consumers purchase most products (even those on promotion) much less frequently than they visit stores. Thus we have an environment where consumers have brief, frequent exposures to a product that they are not necessarily going to purchase. We argue that repeated exposure to a brand size via store displays and features increases the probability of purchase. There are multiple explanations consistent with a positive effect. One possibility is that multiple exposures generate product awareness and/or increase accessibility of the product in memory. On future purchase occasions, consumers are more likely to recognize and consider the product. The mere exposure effect suggests that these repeated exposures may enhance attitude toward the brand. (Zajonc and Markus (1982)). In low involvement decisions, such as purchasing a packaged good, affective response towards the brand is an important determinant of choice (Baker and Lutz (1987), MacInnis and Jaworski (1989)). The fact that the mere exposure effect has been demonstrated in the absence of recall or recognition seems particularly relevant. Thus, brief multiple exposures to brands may have a positive effect by either increasing brand accessibility in memory and brand attitude.

Thus we have two competing hypotheses. Prevailing theory suggests that past exposure to displays and features may erode brand equity and have a negative or non-positive impact. Recent work on advertising effects in FSIs and the frequency of exposure to brands on non-purchase occasions suggest that the effect will be positive. The rest of the paper is organized as follows. Section 1, provides a theoretical framework which accommodates both a positive and negative effect of previous trade promotions. Sections 2 and 3 describe the data and models. Section 4 develops the hypotheses and in Section 5 we discuss and interpret the results. In Section 6, we discuss findings and finally, we conclude with some brief remarks.

⁴ Uncles' data from Grand Junction, CO and Midland, TX show that patrons of two stores (City Market, and Store F) visited any store in their respective market 4.1 and 3.4 times per week.

⁵ In Bucklin and Gupta's dataset, 300 households made 30,966 shopping trips over 52 weeks. (roughly 2 shopping trips per week)

⁶ I will use the terms "household" and "consumer" interchangeably although strictly speaking they are not identical.

Section 1: Theoretical Framework

As discussed in the introduction, there is evidence to suggest that multiple exposures to previous store displays and features may have both positive and negative effects. The purpose of this section is to sketch an information based theory which accommodates both a positive and negative consumer response to previous trade promotions. While the self-perception theory used by Dodson et al (1978) is compelling, it relies on attitude formation on purchase occasions. In this paper, we want to consider the previous effect of trade promotions on both purchase and non-purchase occasions. Therefore our theory must encompass exposures on all types of shopping trips.

Our theory is based on the information content of store displays and features. Therefore, it seems worthwhile to describe a typical display and feature. Both displays and features generally provide a customer with at least two pieces of information: 1) the presence⁷ of the product at a particular store (chain) and 2) the promotional price of the product. Most in-store displays are assembled by store employees and primarily present the product to the customer. The display will often contain a sign which indicates that the product is available at a reduced price (e.g., "Save", "Special"). Specialty displays or manufacturer prepared displays may be more elaborate and contain pictures and/or copy related to an on-going campaign. (e.g. a Super Bowl display) However, most displays tend to present only the product and price with no special copy. Store features are very simple print advertisements. The larger features will usually contain a picture of the product and some features include a store coupon. Features almost always include copy which indicates the presence of a special price. There is normally not any extended copy indicating product benefits that one would see in typical print advertisements. Extended copy is usually related to a store promotion (e.g. 4th of July Sale) and not specific product. Thus both display and feature primarily present the promoted product, a price, and simple copy.

We argue that the product specific information (i.e. presence of product) leads to a positive effect while the price information leads to a negative effect. First, consider information about product presence. One can think of this information as a very simple form of advertising. If the consumer is unaware of the product, the display or feature may generate awareness. Among aware consumers, the display or feature is a reminder about the product and may increase product recognition on future purchase occasions. The exposure may also increase the probability that a

⁷ Assuming the product is in-stock.

consumer considers the brand on future purchase occasions. Increased brand awareness, recognition, or consideration will result in a positive advertising-like effect.

The second piece of information a consumer sees in a store display or feature is price. Repeated exposure to prices increases consumer's knowledge of relative prices in the category. We claim that consumers with more price information will be more price sensitive. Specifically, consumers that have multiple exposures to promoted items learn about the distribution of promotional prices in the category. As an extreme example, suppose that some consumers are very patient and become perfectly informed. The demand curve for this segment of consumers is perfectly flat as these consumers only purchase at the lowest promotional price. A flat demand curve is representative of consumers who are extremely price sensitive. In general, we expect consumers with more price information to be more price sensitive.

Moreover, consumers do not need to recall the magnitude of the promotional price to become more price sensitive. Studies have shown that consumers are more responsive to frequency cues than to magnitude cues. (Alba, Broniarczyk, Shimp, and Urbany (1994)) The promotion itself (display or feature) provides a simple frequency cue that the product is offered at a discount. This information will change a consumer's beliefs about the frequency of promotion for that product. Similar to the above arguments, this will increase a consumer's price sensitivity.

We think of these two effects in terms their impact on brand size demand. We expect the positive advertising-like effect to shift the demand function outward (i.e. intercept effect). In our model, we will look for a positive main effect of lagged displays and feature. We expect the negative effect to rotate the demand curve. (i.e. slope effect) As in our example, we expect the demand curve to become flatter. In our model, we will look for a negative interaction between lagged display or feature and price.

There are obviously other theories consistent with a negative and positive response to previous promotions. For example, the mere exposure effect would argue that consumer affect towards a product increases after multiple exposure to displays or features. This theory is appealing because it does not rely on any elaborative processing of information. We do not plan to test alternative theories such as this against our theory. Rather, our information based framework provides one simple explanation as to how past exposures to store displays and features may have both a positive and negative effect.

Section 2: Data

The data in our analysis is from multiple IRI BehaviorScan markets. The data was originally collected as part of a TV advertising analysis in the detergent category. The data is rich in that we have household level TV advertising exposure data for three major brands. Unfortunately, the data is also reduced in size from the “normal” panel because only households with television converter boxes in their homes are included in the data set. In all, we have access to detergent data in ten⁸ BehaviorScan markets. The amount of data across the ten markets is enormous which suggests that we initially narrow our focus to a single market, Pittsfield, MA. I select a sample of 150 households that made 927 detergent purchases in an eighteen month time period, or an average of one detergent purchase every 11 to 12 weeks. On each household purchase occasion, the household ID, store, day, week, UPC, price, and coupon usage are recorded.

What separates our analysis from previous work on this topic is the inclusion of household store visits. The previous scanner data studies by Bawa and Shoemaker (1987) and Shoemaker and Neslin (1989) limit their analysis to purchase occasions. The 150 households in our sample combined for 21,466 store visits. Note that the ratio of store trips to detergent purchases, 23:1, is extremely large. Therefore, consumers visit stores very frequently relative to the number of detergent purchase occasions. Even a heavy detergent user will have a high ratio of visits to trips. In the analysis, we treat each store visit as an opportunity to be exposed to the brand.

In the detergent category there are several hundred UPCs and over one-hundred brand sizes. Within a brand size there may also be differences in form (liquid or powder), scent (regular or scented), and whether or not bleach is added. We limit the dataset to the union of the twenty largest brand sizes by unit share and the twenty largest brand sizes by volume share⁹. The final dataset contains 29 brand sizes containing a total of 78 UPCs and 14 brands. The sampling criteria insure that our sample will contain both small and large volume brand sizes. One drawback of the methodology is that we eliminate small market share brand sizes. Thus, the leading national brand, Tide, comprises 8 of the 29 brand sizes. In the final sample, there were both liquid and powder brand sizes, all brand sizes were regular scented, and two

⁸ There are currently 8 BehaviorScan markets.

⁹ Volume is washloads per unit .

brand sizes of Tide contained bleach¹⁰. See Table 2, at the end of this section, for details.

The store environment data contains information on display, feature, and price by UPC. For a purchased items, there are two possible sources of price information: store data and purchase data. We use the store data price information because it is the only data which contains competitive information. Next is the question of aggregation. Data is recorded at the UPC level, but the customer makes choices at a brand size level. IRI groups UPCs together that have similar attributes (brand, size, scent, form etc.) and we also used this grouping. Once the mapping from UPC to brand size is defined, we aggregate display, feature, and price to the level of brand size. The aggregation of UPCs to brand sizes is intelligent so that most of the time there is only information for one UPC that comprises a brand size. However, the process is not flawless. Conflicts may occur, for example, when a UPC code is changed. In these cases, the old and new UPC may co-exist, with possibly different merchandising, until the old UPC is sold off. Thus, the choice of aggregation rules may create a small amount of noise in the data¹¹.

As mentioned earlier, the entire dataset was created for a TV advertising study for three major detergents. In most experiments, there are a set of test and control households. Test households receive brand advertising while control households receive a public service announcement (PSA) during the advertisement. A household must have a converter box to be part of the dataset. For test households, the converter box is needed to measure exposure (i.e., whether or not the set was turned on to a specific channel); for control households, the converter box is needed to send the PSA. In this particular study, however, the advertising study was created ex-post and there was no test/control group. That is, household exposures were inferred using the TV advertising schedule of the three brands and a household's viewing history as recorded by the converter box. For each household with a converter box, the data set contains seconds of exposure to the specific detergent advertising.

The execution of the TV advertisements is in flights over roughly 14 weeks. Data sets generated from this and other split-panel studies have the following properties: 1) short time duration; 2) limited number of exposures; 3) limited

¹⁰ We did not treat the brand sizes with bleach differently from other brand sizes.

¹¹ We use the following aggregation rules: 1) If any UPCs that comprise a brand size are on display or feature, the brand size is on display or feature. 2) The price of a brand size is the average of the prices for the UPCs. 3) If multiple feature types are present for a brand size, we use the largest feature type present. 4) If multiple displays are present for a brand size, we use the most frequent type of display.

number of brands. In this data set there were 191 household exposures to three brands over 14 weeks. From personal experience, it can sometimes be difficult to find advertising effects using this type of data. This is partly due to the limited number of exposures in the data.

It is interesting to contrast a typical TV exposure to a display/feature exposure. TV exposures are relatively low frequency, high duration exposures. While a typical consumer may see only a few TV commercials, the length of each commercial is 15-60 seconds. In contrast, display and features are high frequency, low duration exposures. For packaged goods, purchase decisions occur quite quickly and therefore the display/feature exposure is also brief. However, because consumers visit stores frequently, there are a large number of exposures. Thus there is a distinct difference between TV advertising exposures and display/feature exposures.

The display and feature exposure data is further enriched by the many types of displays and features. Analogous to advertising, one could potentially think of each type of display or feature as a different type of advertising. To create a meaningful aggregate measure of display and feature activity, one can ask the question: "What is the probability that a randomly selected individual (in our sample) would observe the brand size they are purchasing with a particular type of display or feature?" Table 1 displays these probabilities for the types of displays and features in our dataset¹². The aggregate measure should give the reader a feel for the relative frequency of displays and features. For example, from Table 1 one can see that Promotional Aisle Displays occur half as frequently as Back-End-of-Aisle Displays. In all there are data for ten different types of displays and features in our dataset. The alphabetic description of features (C, B, A) is rank ordered by increasing physical space of the feature. In addition, Super A features are defined as store coupons. The display coding is not ranked in any particular order. We will drop the specialty displays and shipper displays from our analysis but include them below for completeness.

¹² The measure used is the weighted total number of displays/features by brand and store over a 78 week period. The weights are brand sales and store sales. To get a probability, we divide the weighted measure by 78.

Distribution of Displays and Features

Display	Probability of Display*	Feature	Probability of Feature*
No Display		No Feature	
Lobby Display	6.0	C Feature	0.0
Front end-aisle Display	1.8	B Feature	1.0
Mid-aisle Display	3.7	A Feature	3.7
Back end-aisle Display	4.7	Super A Feature	2.2
Specialty All/Other	0.0	Any Feature	6.9
Shipper Display	0.0		
Promotional Aisle	1.9		
Any Display	18.1		

Table 1

* Multiplied times 100.

It is also interesting to look at the extent of feature and display activity by brand size. Table 2 shows the percent of weeks that a brand had any type of feature and display. One can see that certain brands, such as Tide, are more heavily promoted than others. Smaller brands may have no display activity in the dataset. Note that there is also variation across brands and within brands.

In sum, this is a very rich dataset. There is a large sample of households and purchases, TV advertising data for three brands, and thousands of exposures by household to several different types of displays and features. I now turn to the model.

Description and Summary Statistics of Brand Sizes

BRAND	SIZE	FORM	Market Share	% Of Weeks on Display	% of Weeks on Feature	Average Regular Price	Average Promotion Price
AIAX	128 OZ	LIQUID	3.5	2.9	0.4	5.00	4.84
ALL	64 OZ	LIQUID	2.8	0.0	0.0	2.98	2.92
ALL	128 OZ	LIQUID	4.4	3.2	0.4	5.55	5.26
ARM & HAMMER	128 OZ	LIQUID	4.4	9.7	1.1	5.18	5.03
ARM & HAMMER	224-227 OZ	POWDER	2.2	1.1	0.0	6.16	5.85
ARM & HAMMER	400 OZ	POWDER	1.8	1.1	0.4	10.34	9.79
BREAK THRU	64 OZ	LIQUID	5.3	0.0	0.0	2.43	2.40
BREAK THRU	128 OZ	LIQUID	6.4	3.6	1.1	4.65	4.51
DYNAMO 2	64 OZ	LIQUID	1.7	0.0	0.0	3.82	3.52
EARTH'S CHOICE	48 OZ	LIQUID	2.4	0.0	0.0	3.28	2.81
ERA	128 OZ	LIQUID	2.2	2.5	0.4	8.61	7.57
OXYDOL	67-84 OZ	POWDER	2.6	0.0	0.0	4.51	4.06
OXYDOL	136-171 OZ	POWDER	1.6	0.0	0.0	8.97	8.46
PUREX	136-147 OZ	POWDER	2.3	1.8	0.4	2.90	2.67
PUREX	64 OZ	LIQUID	2.1	1.4	0.7	5.00	4.86
PUREX	128 OZ	LIQUID	1.8	4.3	0.4	5.54	5.29
SOLO	128 OZ	LIQUID	3.6	1.1	0.0	6.99	6.02
SURF	64 OZ	LIQUID	7.3	0.4	0.4	4.35	3.73
TIDE	39-49 OZ	POWDER	2.2	0.0	0.0	2.84	2.82
TIDE*	67-72 OZ	POWDER	3.7	0.0	0.0	4.47	4.30
TIDE	67-84 OZ	POWDER	3.0	0.0	0.0	4.39	4.14
TIDE*	136-147 OZ	POWDER	4.4	1.4	0.0	8.79	8.52
TIDE	136-171 OZ	POWDER	2.3	1.8	0.0	9.10	8.51
TIDE	64 OZ	LIQUID	6.2	1.4	0.4	4.84	4.23
TIDE	96 OZ	LIQUID	5.9	5.0	0.4	6.63	5.89
TIDE	128 OZ	LIQUID	6.5	6.5	0.7	8.85	7.83
WISK	64 OZ	LIQUID	1.9	1.8	0.0	4.27	3.96
WISK	128 OZ	LIQUID	2.2	1.1	0.7	8.75	7.58
YES	64 OZ	LIQUID	3.4	0.0	0.0	2.83	2.73

Table 2

% Display and % Feature are percentage of weeks with any display or feature. Market Share over entire sample of households. * indicates bleach.

Section 3: Model

There are obviously many ways to model the effect of previous displays and features on current brand size choice. Our perspective is to choose a very simple base model from which we create variations that will allow us to test alternative hypotheses. We only consider a multinomial logit specification because the computing time necessary to estimate these models in a less restrictive framework, such as multinomial probit, is excessive. My approach in this section is to lay out the basic model and then describe extensions for testing alternative hypotheses.

Base Model

In our data set, there are J brand sizes ($j=1\dots J$), t_h purchase observations per household, T total purchase observations, t_h^* shopping trip observations per household, and T^* shopping trip observations. Purchase observations are indexed by $t \in T$ and shopping trip observations are indexed by $t^* \in T$. (Note that $T \in T^*$.) Consider a single household, h , and let:

$$U_{j,t} = 1\alpha_j + X_{j,t}\beta + \text{Loyalty}_{j,t}\theta + \text{Stock}_{j,t}\phi + \varepsilon_{j,t} \quad (1)$$

where:

$U_{j,t}$ is the utility of brand size j at purchase time t for household h .

α_j , β ($k \times 1$), θ , and ϕ ($z \times 1$) are parameters ($\alpha_1 = 0$)

$X_{j,t}$ is a $1 \times k$ vector of marketing mix variables for brand size j at purchase time t

$\text{Stock}_{j,t}$ is a $1 \times z$ vector of stock variables for brand size j at purchase time t

In our model, $X_{j,t}$ is the price, display, and feature of brand size j at purchase time t . The variables in each model are defined more precisely in Section 5. $\text{Loyalty}_{j,t}$ is an exponentially decaying function of past purchases.¹³

$$\text{Loyalty}_{j,t} = e^{-\lambda} \text{Loyalty}_{j,t-1} + r_{j,t-1} \quad (2)$$

where:

$r_{j,t-1} = 1$ if brand i was purchased at time $(t-1)$ and 0 otherwise.

¹³ Given the proper initial conditions, this is algebraically equivalent to the Guadagni Little variable where $(1-\gamma)$ replaces $e^{-\lambda}$ and lagged purchases are weighted by γ instead of 1. See the Appendix for details.

The Loyalty variable is initialized using a six month pre-period using each household's brand size share over this period. Household's that did not purchase any brands in the pre-period are initialized at the population brand size share during the pre-period.

Assuming that the error term¹⁴ is also Gumbel, the overall probability of brand size purchase can be expressed as:

$$\text{Prob(Purchase)}_{i,t} = \frac{e^{U_{i,t}}}{\sum_J e^{U_{j,t}}} \quad (3)$$

Stock Model

A simple structure lets us focus on the variables of interest which are the $\text{Stock}_{j,t}$ variables. Let:

$$\overline{\text{Stock}}_{j,t^*} = \text{Exposure}_{j,t^*} + e^{-\lambda(t^*-s)} \overline{\text{Stock}}_{j,s} \quad (4)$$

$$\text{Stock}_{j,t^*} = \overline{\text{Stock}}_{j,t^*} - \text{Exposure}_{j,t^*} \quad (5)$$

where:

t^* is the current store trip observation, s is the previous store trip

$\text{Exposure}_{j,t^*}$ is the exposure to brand size j at time t^* .

$\text{Stock}_{j,0} = 0 \quad \forall j$

This is a very general formulation and $\text{Exposure}_{j,t^*}$ can be any type of exposure to a brand size. In our model, we will consider exposures to TV advertising, displays, and features. In other words, we take exposure to display and feature on a store visit and treat it as an exponentially decaying advertising variable. At any point in time, a household has a "stock" of exposure to past TV advertising, displays, or features. (Abe 1991) A new exposure increases the stock of exposures; with no new exposures, the stock will exponentially decay. We will refer to this as TV advertising stock, display stock and feature stock. The focus of our paper will be looking at various formulations of stock variables.

Because the current exposure to display and feature is not of primary interest, we subtract the current exposure from the stock variable. (Equation 5) Therefore display and feature stock only contain historical information about what a household

¹⁴ $\epsilon_{j,t}$

has seen on previous shopping trips. The stock variable will include exposures on shopping trips where a purchase occurred.

A final point about our methodology and model specification. Our specification includes lagged purchases (endogenous variables) via the loyalty variable. Previous purchases influence current choice via purchase feedback. (see Figure 1) The model also contains lagged marketing mix variables (exogenous variables) in the display and feature stocks. Since previous marketing influences previous choice (Indirect Effect), there are two ways for previous display and feature to influence current choice. We are primarily interested in modeling what we label as the direct effect in Figure 1.

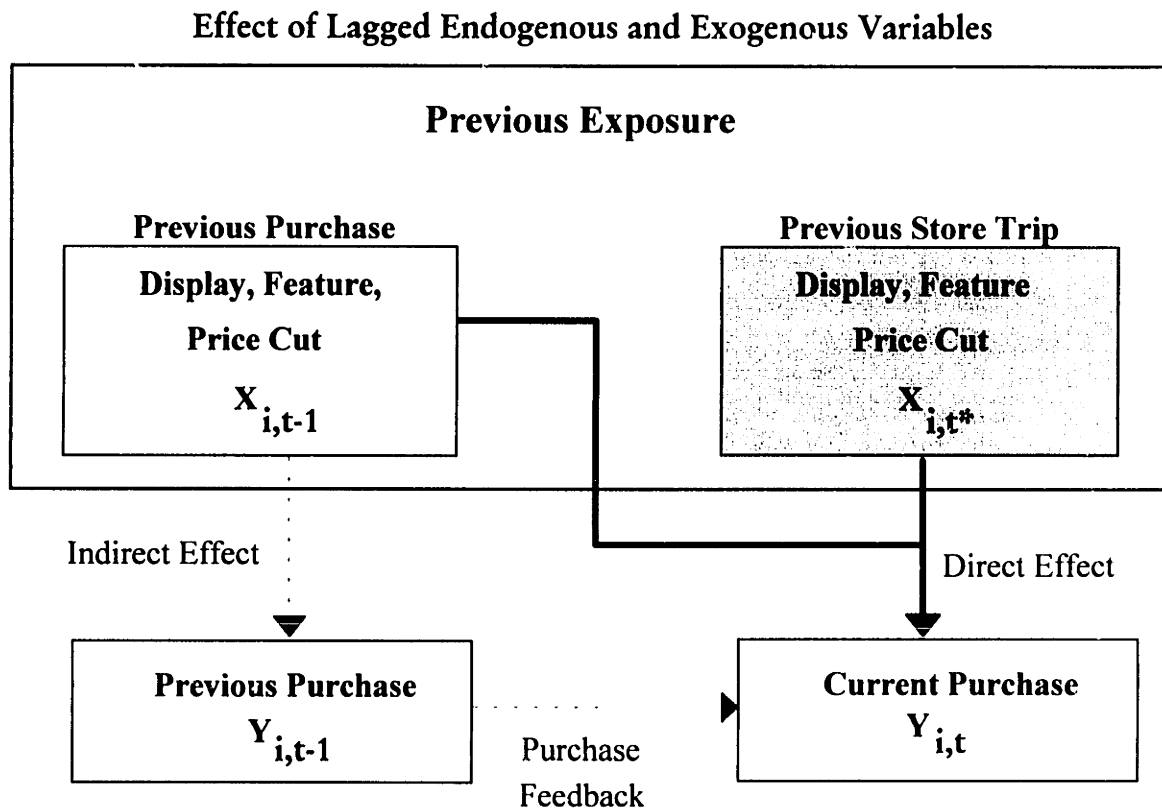


Figure 1

As there are two ways for previous marketing to influence current choice (direct and indirect), it raises the issue of model specification. We first recognize that assuming the model is correctly specified, the feature stock and display stock coefficients are unbiased and will only reflect the advertising/brand erosion effect. (Direct Effect) What if the purchase feedback is incorrectly specified?¹⁵ In particular,

¹⁵ As with any type of specification error, all the coefficients in the model are generally biased.

one may be worried about picking up effects in the stock variables which are related to misspecification of purchase feedback. We recognize that this is a potential problem, but will assume that the purchase feedback is correctly specified. If not, any problem arising is swamped out by the number of household shopping trips. That is, endogeneity only arises with those shopping trips where a purchase occurs. Since the ratio of shopping trips to purchases is large, the extent of any bias should be small.

Section 4: Hypotheses

Our theoretical framework suggests that we consider both a main effect and an interaction effect. The positive advertising-like effect of previous displays and features suggests that previous promotions have a positive main effect. Our theory of price information suggests that previous exposure to displays and features have a negative interaction effect with price.

H1 Main Effect: Exposure to either brand size displays or features on previous store visits increases the current probability of purchasing the brand size.

H2 Interaction Effect with Price: Exposure to either brand size displays or features on previous store visits increases price sensitivity.

H1 and H2 are very broad hypotheses. Given the richness of our data, we would like to control for other factors which may influence the effectiveness of different types of past displays and features. We begin by considering the different types of displays and features. A reasonable hypothesis is that lagged effects differ by type of display and feature. Since features have a convenient rank ordering, one would expect larger features to have a greater impact than smaller features. Unfortunately, displays do not have any natural a-priori ordering. However, the size of the current effect seems like a reasonable proxy for the effectiveness of the lagged effect. We hypothesize that displays with larger current effects should also have greater lagged effects.

We expect that the main effect and interaction effect will continue to hold for feature and display types. However, the absolute value of the effects should increase for more effective stock variables. To simplify verbiage, we define "effectiveness" as an increase in the absolute value of the stock main effect and stock*price interaction effect coefficients.

H3 Feature and Display Types: Larger feature stock variables (e.g. Super A) are more effective than smaller feature stock variables. Effectiveness of display type stock variables is increasing in the size of the current effect.

None of these hypotheses address the heterogeneity across household's. Given the work of Neslin and Shoemaker (1989), it seems logical that one should control for differences among households. The first issue to address is whether or not a

household is exposed to the display or feature. The dataset contains information on the store environment when a customer is in a store. However, it is uncertain whether a household saw the displayed or featured brand size. We hypothesize that consumers who regularly purchase any brand size on display or feature are more likely to be exposed to any type of display or feature.

We define a household's feature (display) propensity as the percent of purchases where a consumer purchased any item on feature (display). It makes intuitive sense that customers that purchase more of any brand that is featured (displayed) are more likely to be impacted by past features (displays). One can think of feature propensity as a way of sorting the store feature readers from the non-feature readers. Similarly, the display propensity variable sorts the display responsive consumers from the display insensitive consumers. For each household, the propensity variable is a number between 0 and 1 indicating the household's propensity to buy featured or displayed items.

H4 Propensity: Effectiveness of feature (display) stock variables should be greater for feature (display) prone households.

H1 - H4 summarize our hypotheses. We will look for simple main effects and interaction effects at an aggregate level. We will also control for differences in types of display and feature as well as different household response to displays or features.

Section 5: Results

We organize our results as follows. First, we describe baseline models which illustrate the heterogeneous response among households to different types of displays and features. This segmentation is a critical piece of our analysis. Using this segmented baseline model, we then present our main results.

Household Segmentation and the Current Marketing Mix

To begin our analysis we create a simple baseline model, M3a in Table 3, that includes the effect of current marketing. This simple model includes alternative specific constants, loyalty, and current marketing variables. Since both the physical size and form vary across brand sizes, brand size prices are normalized by washloads per container. A nominal value for the price variable is roughly 20 cents per washload. Current display and feature are binary variables (0/1) indicating the presence of any feature or display for that brand size. To get a better understanding of the marginal effects of the variables, it is useful to express the marginal utility from a display or feature in terms of price changes. The presence of a display creates the equivalent marginal utility as a \$0.11 price cut per washload. Similarly, the presence of a feature creates the equivalent marginal utility as an \$0.08 cent price cut per washload. Thus the current effects for both features and displays are significant and large in magnitude.

As stated in the hypotheses, our prior beliefs are that consumers should be more responsive to larger features. Similarly, we believe there should be different response to display types. Model M3b separates feature and display types into six categories: super A features, A or B features, lobby displays, front or back aisle displays, mid-aisle displays, and promotional aisle displays. We cluster the A and B features as well as the front and back aisle displays based on the coefficients and standard errors of a model which is not shown. We chose to keep the mid-aisle displays and promotion aisle displays distinct variables as we believe they capture different phenomena. As one would expect, the coefficient for Super A features is much larger than the AB feature coefficient. This captures the presence of a coupon. The magnitude and rank order of the display coefficients is somewhat surprising. A priori, one would not have expected lobby displays to be radically different from front or back end aisle displays. In addition, one would not have expected mid-aisle displays to have such a large coefficient. This model illustrates the heterogeneous

response to different types of features and displays. The log-likelihood is -1534 which is a significant improvement versus the prior base model.

Effect of Current Displays and Features is Significant

Variable	M3a		Variable	M3b	
	Coef	T		Coef	T
Smoothing (Loyalty)	0.395	(12.6)	Smoothing (Loyalty)	0.394	(12.7)
Loyalty	2.181	(23.7)	Loyalty	2.213	(23.9)
prices	-15.193	(-6.7)	prices	-17.510	(-8.0)
display	1.640	(13.5)	Lobby Display	0.685	(2.9)
feature	1.275	(8.6)	Front/Back Display	1.623	(9.6)
			Mid Aisle Display	2.055	(10.4)
			Promo Aisle Display	2.033	(7.5)
			Super A Feature	2.239	(9.7)
			A or B Feature	0.704	(3.3)
Log Likelihood	-1555.6		Log Likelihood	-1534.8	
Adjusted p ²	0.492		Adjusted p ²	0.497	

Table 3

Next we want to consider household heterogeneity. We base our household segmentation on the prior belief that households differ in their responsiveness to displays and features. Some households are more likely to read the weekly store features; some households pay close attention to store displays on a given shopping trip. We label our household segmentation variables as feature propensity and display propensity. We define feature (display) propensity variable as the percentage of household purchases where the purchased product was on feature. This static propensity variable is therefore a number between (0,1) indicating a household's sensitivity to feature (display).

We model heterogeneous household response to display and feature by including a main effect of each type of feature or display as well as an interaction effect with feature or display propensity. Households that respond to features and displays may also have different price sensitivity. A priori, one would expect these households to be more price sensitive. We model this by including price*propensity interaction variables.

Results of our household segmentation suggest that response to features, displays, and price is quite heterogeneous. (Table 4) As we expected, our propensity

variable is quite good at segmenting display/feature sensitive households. Other than mid-aisle displays, the main effects by display/feature type are insignificant and small in magnitude. In contrast, we find strong interaction effects for all types of display/feature. The interaction coefficients are also quite similar in magnitude. (roughly 4) The one exception is mid-aisle displays where we pick up a strong main effect. One possible explanation for this is that for all other types of displays, the physical location of the product changes. Mid-aisle displays, often referred to as "shelf talkers", are usually attached to the shelf at a product's normal location. Therefore it is more likely that all consumers will be influence by these type of displays.

A major surprise in the results are the price interaction terms. Consistent with our prior beliefs, we find that display prone households are significantly more price sensitive. However, we find that feature prone households are more price insensitive. A little thought about the data reveals some insight into this result. In our dataset, we use the store data price as this is the only data set which contains competitive price information. The price recorded in this data set is the shelf price. However, by definition, super A features include a store coupon. While the existence of the coupon is recorded, we have not included the out-the-door price in our model. Therefore, if a household purchases a product with a coupon, we do not record the actual price paid in our data. The model is now observing feature prone customers purchasing the product at relatively high prices. Hence, the decreased coefficient for price sensitivity.

Table 4 also contains the TV advertising stock variable. Given the low number of exposures in our data set, we were unable to estimate a decay parameter for this variable. We fixed the value at 0.10 which is consistent with previous TV advertising studies (Abe, 1991). In this model, and in other formulations, we generally find positive but insignificant effects for the TV advertising variable.

The important findings in our household level segmentation model are that display and feature prone consumers differ in their price sensitivity and are more responsive to displays and features. The dramatic change in log-likelihood indicates the significance of this segmentation.

Household Segmentation: Feature and Display Propensity

	Variable	Coef	T	
	Smoothing (Loyalty)	0.409	(12.0)	
	Loyalty	2.204	(22.8)	
	Price	-15.854	(-6.1)	
	(Current Price)*feature propensity	14.791	(2.3)	
	(Current Price)*display propensity	-14.469	(-2.9)	
	Super A Feature	0.439	(1.0)	
	A or B Feature	-0.254	(-0.7)	
	Lobby Display	-0.650	(-1.4)	
	Front or Back Aisle Display	-0.084	(-0.2)	
Current Marketing	Mid Aisle Display	1.170	(3.0)	
	Promotional Aisle Display	-0.141	(-0.2)	
	(Current Super A feature)*feature propensity	4.255	(4.1)	
	(Current AB feature)*display propensity	4.102	(4.5)	
	(Current Lobby)*display propensity	3.520	(3.9)	
	(Current Front_Back)*display propensity	4.279	(5.8)	
	(Current Mid)*display propensity	2.454	(3.0)	
	(Current Promo)*display propensity	4.593	(3.3)	
		Smoothing (TV Advertising)	.10	–
		TV Advertising	1.127	(1.1)
	Log Likelihood	-1454.8		
	Adjusted p ²	0.520		

Table 4*Effect of Previous Displays and Features*

Building on our segmentation model, we now want to explore the effect of lagged display and feature variables. Based on our theory, we expect to find a positive main effect and a negative interaction effect with price. Looking at Table 1, one notices that the most frequent type of displays are end-aisle displays. (i.e. front and back end-aisle displays) In addition, the current effect of end-aisle displays is larger than either lobby displays or mid-aisle displays. Given the high frequency and strong current effect, we would expect the lagged effect of this variable to be more significant.

In Table 5 we build on our baseline model by adding previous display variables. Much like the household segmentation baseline, we find significant effects

for display sensitive households. The main effect of front/back display stock for all households is positive and insignificant. The interaction effect, (1.03, $t=1.6$), is positive and marginally significant. Evidence of increased price sensitivity among display prone households is seen by a significant negative interaction effect with price (-8.45, $t=-2.5$). Therefore, in support of our theory, we find a positive main effect and a negative price interaction effect among display prone households.

Front and Back Aisle Display Stock

	Variable	Coef	T	
Current Marketing	Smoothing (Loyalty)	0.405	(12.0)	
	Loyalty	2.209	(22.8)	
	Price	-15.808	(-6.1)	
	(Current Price)*feature propensity	14.333	(2.3)	
	(Current Price)*display propensity	-11.759	(-2.3)	
	Super A Feature	0.492	(1.1)	
	A or B Feature	-0.223	(-0.6)	
	Lobby Display	-0.660	(-1.4)	
	Front or Back Aisle Display	-0.211	(-0.5)	
	Mid Aisle Display	1.169	(3.0)	
	Promotional Aisle Display	-0.226	(-0.3)	
	(Current Super A feature)*feature propensity	4.186	(4.0)	
	(Current AB feature)*display propensity	3.912	(4.3)	
	(Current Lobby)*display propensity	3.574	(4.0)	
	(Current Front_Back)*display propensity	4.564	(5.8)	
	(Current Mid)*display propensity	2.554	(3.2)	
	(Current Promo)*display propensity	4.852	(3.5)	
		Smoothing (TV Advertising)	.10	--
		TV Advertising	1.141	(1.1)
	Previous Displays	Smoothing (Display)	0.045	(12.3)
Front/Back Display Stock		0.153	(1.2)	
Front/Back Display Stock*Display Propensity		1.030	(1.6)	
F/B Display Stock*Display Propensity*Price		-8.452	(-2.5)	
	Log Likelihood	-1450.0		
	Adjusted p^2	0.522		

Table 5

Next we consider the effect of lagged features. Given our previous explanation for the positive price*feature propensity interaction coefficient, we do not expect good results for lagged super A features. That is, it will be difficult to measure changes in price sensitivity since we have less accurate measurement of out-the-door price for products with store coupons. However, A and B features should not have such problems. Table 6 contains results from a model which includes the lagged effects of A and B features. Again, we find confirming evidence of our theory. The main effect for all households is small in magnitude, negative, and insignificant (-0.07, $t= 1.4$). The main effect for feature prone households is positive, large, and significant. (1.42, $t=2.9$). We also find a significant negative price interaction effect for feature prone households. (-6.9, $t=-2.9$) The highly significant main effect and interaction effect lends support to our theory.

		A or B Feature Stock	
		Variable	
		Coef	T
Current Marketing		Smoothing (Loyalty)	.40* --
		Loyalty	2.185 (32.8)
		Price	-15.680 (-6.1)
		(Current Price)*feature propensity	21.821 (3.1)
		(Current Price)*display propensity	-13.542 (-2.7)
		Super A Feature	0.395 (0.9)
		A or B Feature	-0.216 (-0.6)
		Lobby Display	-0.595 (-1.2)
		Front or Back Aisle Display	-0.078 (-0.2)
		Mid Aisle Display	1.165 (3.0)
		Promotional Aisle Display	-0.114 (-0.1)
		(Current Super A feature)*feature propensity	4.319 (4.1)
		(Current AB feature)*display propensity	3.947 (4.2)
		(Current Lobby)*display propensity	3.407 (3.8)
		(Current Front_Back)*display propensity	4.250 (5.7)
		(Current Mid)*display propensity	2.578 (3.1)
		(Current Promo)*display propensity	4.543 (3.3)
	Previous Features		Smoothing (TV Advertising)
		TV Advertising	1.010 (0.9)
		Smoothing (AB Feature)	0.003 (19.2)
		AB Feature Stock	-0.070 (-1.4)
		AB Feature Stock*Feature Propensity	1.421 (2.9)
		AB Feature Stock*Feature Propensity*Price	-6.911 (-2.9)
		Log Likelihood	-1449.5
		Adjusted p ²	0.522

Table 6

Discussion

Our main finding is that there is a positive main effect and a negative price interaction effect among display prone and feature prone households. Our theory is based on household's observing two pieces of information when they are exposed to a display or feature. First, the reminder of the product serves as a weak form of product advertising. Consumers who respond to displays and features are more likely to recall, recognize, and consider this product on future purchase occasions. The second piece of information a consumer is exposed to is price. Consumers gain knowledge of the frequency and distribution of promotional prices which makes them more price sensitive.

An further contribution of the paper is our methodology. In particular, researchers have generally only considered displays and features to have an immediate impact on share. By using shopping trip data, we show that displays and features also have weak advertising-like properties which to this point have been ignored. The richness of the shopping trip data may allow researchers to explore alternative advertising models.

The negative effect in our model is consistent with the reference price theories of "brand erosion". The strong interaction with price suggests that brand erosion may be primarily price image and not quality image. It should be noted that our model specification does not allow us to estimate both a positive advertising-like effect and a negative quality erosion effect. Thus the true advertising effect may be attenuated by a decrease in quality image. While our results support a theory of decreased price image, alternative model specifications may help sort out competing theories of brand erosion.

The careful segmentation of display/feature types and household response to trade promotions is a critical part of our analysis. In particular, accurate modeling of the price sensitivity was very important. One can imagine that failure to control for different household price sensitivities would make it difficult to accurately model the effect of lagged displays/features on price sensitivity. It is encouraging that we primarily find significant effects among those households that respond to displays and features. Given that the exposure to display and feature are very brief, one would not expect all households to respond to lagged displays and features.

In addition to the household propensity segmentation, we also explored other types of segmentation. In the marketing jargon, displays and features should have a greater impact on switching customers than loyal customers. This can be approached

in several ways. Let $P(x)$ be the probability of purchase which is increasing in x . Switching customers can be thought of as those for whom $dP(x)/dx$ is large. Switching customers may also be those who are indifferent between several brands. By construction, one would expect these customers to be more price sensitive and more influenced by any advertising-like effects.

We operationalize the household switching variable using the Guadagni and Little (1983) loyalty variable. Letting L be the loyalty variable, switchers are households for whom $L*(1-L)$ is larger. We explored several models of that included this variable. The effect of including this in our model is to allow for a non-linear response to loyalty. Indeed, we find that household response to loyalty is non-linear. However, these models did not shed additional light on the lagged effects of displays and features.

A final issue with household segmentation arises in our definition of propensity. Since we use the calibration data to develop the variable, there is some degree of endogeneity. We don't dispute this, however, we note that the propensity variable primarily contains household specific information and not brand size information. Since we look at a household's share of purchases across all brand sizes, the propensity variable contains only weak brand specific information. There are several ways to avoid problems with this variable. One method is to operationalize the variable like a stock variable. A second method is to look at another product category and to develop the household specific variable¹⁶. We will explore both methods in the future.

¹⁶ This was suggested by David Scharfstein.

Conclusion

We began this research project with the agenda of disentangling the long-term effects of trade promotions. We develop a theoretical framework based on the information content of displays and features which suggests that lagged displays and features have a positive main effect and a negative price interaction effect. Our results lend support to the notion that displays and features are a weak form of advertising that increases awareness, recognition, and consideration of the brand size but also increases price sensitivity.

Our finding of two different effects is a compelling result, but leaves us striving for further understanding of the phenomenon. We recognize that the proposed theoretical framework is not unique. Future experimental and empirical research are needed to sort out alternative theories.

Previous research has suggested that the primary effect of displays and features was purely short-term. If there was a positive long-term effect, it was only through purchase feedback. The present research suggests a new type of positive long term effect which has gone unrecognized and a negative price interaction effect that counters it.

Appendix 1: Equivalence of Adstock and Guadagni Little Variables

Consider the Guadagni Little loyalty variable:

$$\text{Loyalty}_{j,t} = \gamma \text{Loyalty}_{j,t-1} + (1-\gamma)r_{j,t-1}$$

and the Stock variable:

$$\text{Stock}_{j,t} = e^{-\lambda} \text{Stock}_{j,t-1} + r_{j,t-1} \quad (6)$$

Claim:

If

$$e^{-\lambda} = \gamma$$

$$\text{Stock}_{j,0}(1-\gamma) = \text{Loyalty}_{j,0} \quad (7)$$

Then

$$\text{Stock}_{j,t} (1-\gamma) = \text{Loyalty}_{j,t} \quad \forall t \quad (8)$$

Proof:

By induction.

True for $t=0$ by (7).

Assume true for t . Then by (6) and (8)

$$\text{Loyalty}_{j,t+1} = \gamma(1-\gamma)\text{Stock}_{j,t} + (1-\gamma)r_{j,t}$$

$$\text{Loyalty}_{j,t+1} = (1-\gamma)[e^{-\lambda}\text{Stock}_{j,t} + r_{j,t}]$$

$$\text{Loyalty}_{j,t+1} = (1-\gamma)\text{Stock}_{j,t+1}$$

QED

Appendix 2: Estimation

In the basic model we must estimate the following parameters: α_i , β ($k \times 1$), θ , and ϕ ($z \times 1$) where ($\alpha_1 = 0$). If one models the probability of exposure, one must also estimate the vector Δ . I use the analytic gradient procedure suggested in Fader, Lattin and Little (1992). The basic idea is to differentiate the utility function with respect to each parameter¹⁷, choose a starting value for each parameter, and evaluate all the derivatives using these parameters. Next, evaluate the utility function at the given parameter values and call this \bar{U} . Finally, let $\tilde{U} = \bar{U} + D\delta + \text{error}$, where D are the derivatives of U with respect to each parameter. One can estimate the δ vector using any standard methodology. I calculate the gradient and hessian explicitly. (see Ben-Akiva and Lerman (1985) for explicit equations) If $\delta \equiv 0$ the chosen set of parameters is at the optimum. If not, update the parameters and iterate.

One can easily see that this general algorithm allows one to estimate all types of linear and non-linear parameters providing that U is differentiable. The basic difference between this methodology and the one described in Fader, Lattin, and Little is that I differentiate the utility function and not individual variables. This is an issue when parameters appear in more than one variable in the model. For example, including a stock variable and a stock interaction variable would lead to two variables containing the same non-linear parameter.

A final issue is convergence. These models are highly non-linear and often fail to converge. Perhaps the most frustrating part of this paper has been arriving at the office in the morning only to find a model failed to converge after many hours of processing. To increase the likelihood of convergence, I have implemented a pseudo-line search algorithm for calculating the parameter step size. The basic idea is to shrink the parameter updates until there is an improvement in the likelihood function. In linear models this seems to work quite well. In non-linear models, convergence is more difficult to achieve. The theoretical solution of calculating the optimal step size seems prohibitively expensive. My alternative solution is more art than science. I simplify the model to its basic components and then add variables one at a time. This “build from the ground” approach is slow, but effective. Algorithms can be written to try several starting values until one converges.

¹⁷ The procedure is described slightly differently in Fader, Lattin, and Little.

Appendix 3: Parallel Computing Algorithm

One may notice that the size of the data sets in this paper are extremely large. In particular, variables dimensioned by household trips are over 20,000 observations. Computing time increases significantly since most of these variables involve non-linear parameters. At each iteration of the algorithm in Appendix 1, a new set of variables must be re-computed based on the new parameter values. That is, the stock variables and all derivatives must be computed at each iteration. Computing these variables is extremely processor intensive as it involves looping over each household.

In this paper, I developed a methodology to utilize the processing power of other computers¹⁸. The basic idea is to send out groups of households to each computer. Each computer then generates a data matrix. These data matrixes are collected by one computer which then continues the algorithm in Appendix 1. If the data matrix needs to be re-computed, parameter updates are sent to each computer and a new data matrix is generated at each computer.

The methodology allows for two scenarios. In either scenario, each computer is utilized to manipulate the raw data for a given set of parameter values. Parameter updates are sent to each computer until convergence. In the first scenario, the data matrix consists of the variables in the model. The second methodology relies on the independence of the multinomial logit model. If all observations are independent in the likelihood function, the gradient and hessian are additive across observations. Thus we can calculate a piece of the gradient (hessian), at each computer. One computer then collects the gradient (hessian) and adds the pieces up. Thus the difference between the scenarios is whether the computers return just variables or the gradient and hessian.

This methodology was implemented on four Dell PCs running Windows and Novell Personal Netware 1.0. One machine is designated as a client, the other three machines are servers. On each server, a section of RAM is set aside as a RAM Drive. This greatly improves performance. Each machine initiates the algorithm from within Matlab for Windows. (i.e., you have to physically go to each machine and start the algorithm.)

The client serves as the “Master” in this design. Initially, the client parses the data by household and sends out the basic raw data to each server. Each server then generates a data matrix which is written to the RAM disk. The client waits for the data matrix to arrive at each server and then processes the information. Finally, the

¹⁸ The general idea was suggested by John Little. The algorithm is the author's.

client writes a file to each servers RAM disk that includes either updated parameter values or a message that convergence has been reached. In this algorithm, the data matrix can be either a variable matrix or the gradient and hessian.

A significant problem with the algorithm is that information needs to be passed from one computer to another computer. If both computers try and access the same data file at the same time, the algorithm will halt. For example, each server writes a data matrix and the client is looking for that data matrix. If the server is writing the data matrix file and the client tries to read the data matrix file a sharing violation will occur. I solve this problem by developing a timing algorithm. I specify windows of time where only the server can read/write and only the client can read/write. If the windows are large enough, sharing violations can be avoided. Unfortunately, the window size also determines the efficiency of the algorithm. Large windows lead to a reliable but slow parallel computer.

A critical part of the algorithms is to optimize each computer's read/write ability. In particular, the client should be extremely fast at reading and writing. Using the RAM disk greatly improves performance. One should also ensure that common network features such as caching delays, which are normally beneficial, are turned off. For example, in Novell the default delay for the network cache is 5000 milliseconds. Another consideration may be to locate the machines physically closer together. In general, anything which improves read/write by the client will allow smaller timing windows and increased performance.

References

- Abe, Makoto (1991), "A Marketing Mix Model Developed from Single Source Data: A Semiparametric Approach," Unpublished Ph.D. Dissertation, MIT.
- Alba, Joseph W. Susan M. Broniarczyk, Terance A. Shimp, and Joel E. Urbany (1994), *Journal of Consumer Research*, Vol 21, September, 219-235.
- Baker, William E. and Richard J. Lutz (1987), "Levels of Processing and Levels of Explanation," in *Nonverbal Communication in Advertising*, ed. S. Hecker and D.W. Stewart, Lexington: Lexington Books, 84-89.
- Bawa, Kapil, and Robert W. Shoemaker, (1987) "The Effects of a Direct Mail Coupon on Brand Choice Behavior," *Journal of Marketing Research*, 24, November, 370-376.
- Blattberg, Robert C. and Scott A. Neslin (1990), *Sales Promotion: Concepts, Methods, and Strategies*, Prentice Hall.
- Bucklin, Randall and Sunil Gupta (1992), "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach," *Journal of Marketing Research*, 19, (May), 201-215.
- Davis, Scott, J. Jeffrey Inman, and Leigh McAlister, (1992) "Promotion Has a Negative Effect on Brand Evaluations - Or Does it? Additional Disconfirming Evidence," *Journal of Marketing Research*, 19, February, 143-148.
- Dodson, Joe A., Alice M. Tybout, and Brian Sternthal (1978), "Impact of Deals and Deal Retraction on Brand Switching," *Journal of Marketing Research*, 15 (February), 72-81.
- MacInnis, Deborah J. and Bernard J. Jaworski (1989), "Information Processing from Advertisements: Toward an Integrative Framework," *Journal of Marketing*, 53, 1-23.
- Neslin, Scott A. and Robert Shoemaker (1989), "An Alternative Explanation for Lower Repeat Rates After Promotional Purchases," *Journal of Marketing Research*, 26, (May), 205-213.
- Little, John D.C. and France Leclerc, "The Effect of Advertising in FSIs on Attitudes Toward the Promoted Brand," Unpublished Working Paper, MIT.
- Spethmann, Betsy (1995) "Trade Promotion Redefined", *Brandweek*, March 13, pp 25-32.
- Srinivasan, Srini S., Robert P. Leone, and Francis J. Mulhern, (1994) "The Advertising Exposure Effect of Free Standing Inserts," Unpublished Working Paper.

Uncles, Mark (1991), Presentation at Marketing Science Conference, Wilmington, Delaware, March.

Zajonc, Robert B. and Hazel Markus (1982), "Affective and Cognitive Factors in Preferences," *Journal of Consumer Research*, 9, 123-131.