PRICE STICKINESS AND CUSTOMER ANTAGONISM

PRICE STICKINESS AND CUSTOMER ANTAGONISM*

ERIC T. ANDERSON AND DUNCAN I. SIMESTER

Managers often state that they are reluctant to vary prices for fear of "antagonizing customers." However, there is no empirical evidence that antagonizing customers through price adjustments reduces demand or profits. We use a 28-month randomized field experiment involving over 50,000 customers to investigate how customers react if they buy a product and later observe the same retailer selling it for less. We find that customers react by making fewer subsequent purchases from the firm. The effect is largest among the firm's most valuable customers: those whose prior purchases were most recent and at the highest prices.

"It seems essential, therefore, to gain a better understanding of precisely what firms mean when they say that they hesitate to adjust prices for fear of antagonizing customers."

Blinder et al. (1998, p. 313)

I. INTRODUCTION

The assumption that nominal prices are sticky is fundamental to Keynesian economics and forms a basic premise of many models of monetary policy. A leading explanation for why prices are slow to adjust is that firms do not want to antagonize their customers. Yet there is little empirical evidence that antagonizing customers through price adjustments lowers either demand or profits. In this paper we study the effects of downward price adjustments and show that many customers stop purchasing if they see a firm charge a lower price than they previously paid for the same item. This customer boycott is concentrated among the firm's most valuable customers, which greatly magnifies the cost to the firm. Firms can mitigate these costs by limiting the frequency and/or depth of price adjustments.

The findings are replicated in two separate field experiments conducted in different product categories. The first experiment

*We thank Nathan Fong for valuable research assistance and the anonymous retailer for generously providing the data used in this study. We gratefully acknowledge valuable comments from Martin Eichenbaum, Nir Jaimovich, Aviv Nevo, Julio Rotemberg, Catherine Tucker, Birger Wernerfelt, Juanjaun Zhang, and numerous other colleagues. The paper has also benefited from comments by seminar participants at Chicago, Columbia, Cornell, Erasmus, INSEAD, LBS, Minnesota, MIT, NYU, Northeastern, Northwestern, Penn State, UCLA, UCSD, UTD, Yale, the University of Houston, the 2006 Economic Science Association conference, the 2007 QME conference, and the 2008 NBER Price Dynamics conference.

The Quarterly Journal of Economics, May 2010

was conducted with a publishing company. Customers were randomly chosen to receive a test or control version of a catalog (the "Test Catalog"). The versions were identical except that the test version offered substantially lower prices on 36 items (the "Test Items"). The loss of demand and profits was particularly dramatic among customers who had recently paid a higher price for one of those items. Lower prices under the test condition led to 14.8% *fewer* orders over the next 28 months, which equates to over $90 in lost revenue per customer (including revenue from the Test Catalog itself).

Decomposing the results reveals that the price adjustments had two effects. As expected, orders from the Test Catalog were higher in the test condition, as some customers took advantage of the discounted prices. However, this short-run effect was overwhelmed by a sharp reduction in orders from other catalogs. Price adjustments in a single catalog had negative spillover effects on future orders (and overall revenue).

To investigate the robustness and generalizability of the results, we replicate the key findings with a separate company in a different product category (clothing). This replication is particularly noteworthy because most customers expect to see lower prices on clothing at the end of a season. Despite this, we show that sending a "Sale" catalog to customers in the days immediately after Christmas reduced purchases by customers who had previously paid a higher price for one of the discounted items.

There are several possible explanations for these effects. One explanation is that customers are antagonized if they observe the firm charging lower prices than they previously paid. Another explanation is that lower prices may have prompted customers to update their price expectations and change their purchase behavior. A third possibility is that low prices may have influenced customers' beliefs about product quality. To evaluate these explanations, we exploit heterogeneity in the experimental treatment effects. For example, the reduction in demand and profits was restricted to customers who had recently purchased one of the discounted items at a high price. Customer antagonism is consistent with these boundaries. We also find some evidence that customers updated their price expectations and delayed purchasing, although this effect does not appear to explain the outcome fully. Other explanations are ruled out by the randomized allocation of customers to the two experimental conditions. The randomization ensures that competitive reactions, inventory constraints,

macroeconomic changes, and customer characteristics cannot explain the decrease in demand, as these factors were equally likely to affect sales under both conditions.

The results of the field experiments had a direct impact on the pricing policies of the two participating retailers. After learning of the results, the company that provided data for our first experiment responded by no longer sending catalogs containing discounts to customers who had recently purchased one of the discounted items. This effectively reduces the degree to which prices are varied for an individual customer. The company that participated in the second study also responded by restricting price changes. For example, the company removed a discounted item from the back page of a widely circulated catalog to avoid antagonizing approximately 120,000 customers who had previously purchased the item at full price. Even before learning of the findings, this company had policies that limited the frequency of discounts. Managers acknowledged that these policies reflected concern that frequent price adjustments may antagonize customers.

These reactions suggest that firms may be able to mitigate the reduction in demand from charging different prices to different customers. If the products are durables that most customers only purchase once, the cost of foregoing these price changes will be minimal. However, most retailers are not able to price discriminate as perfectly as the two catalog retailers in this study. When firms cannot charge different prices to individual customers, the effects that we report will tend to reduce the optimal frequency and/or depth of price adjustments.

I.A. Previous Research

The research on price stickiness can be broadly categorized into two topics: (1) are prices sticky? and (2) why are they sticky? The evidence that prices respond slowly to business cycles is now extensive (Gordon 1990; Weiss 1993). In recent years much of the attention has turned to the second question. One set of explanations argue that costs may be constant, either within a limited neighborhood or within a limited time period. Another common explanation studies the (menu) cost of changing prices. The more costly it is to change prices, the less frequently we would expect them to change. Other explanations have considered imperfections in the information available to price setters and asymmetries in the demand curve.

In this paper we study the role of customer antagonism. This explanation argues that firms do not want to change prices because doing so may antagonize their customers. Hall and Hitch (1939) were among the first to investigate this issue empirically. They interviewed a sample of managers to learn how prices are set. The managers' responses included statements such as "Price changes [are] a nuisance to agents, and disliked by the market" and "Frequent changes of price would alienate customers" (pp. 35 and 38). More recently, Blinder et al. (1998) also asked managers why they did not vary their prices. This study was conducted on a large scale and involved 200 interviews with senior executives conducted over a two-year period. The most common response was that frequent price changes would "antagonize" or "cause difficulties" for customers, leading the authors to conclude that we need to better understand the role of customer antagonism (pp. 85 and 308; see also the quotation at the start of this paper). The findings have since been corroborated by similar studies surveying managers in Canada (Amirault, Kwan, and Wilkinson 2004) and a broad range of European countries (Hall, Walsh, and Yates 1996; Fabiani et al. 2004; Apel, Friberg, and Hallsten 2005).[1] Although these studies have raised awareness that customer antagonism may contribute to price stickiness, there are limits to what can be learned from survey data. Notably, the data do not allow researchers to measure whether antagonizing customers through price adjustments reduces demand or profits.

The evidence in this paper is related to Rotemberg's theoretical research on firm altruism. Customers in Rotemberg's models only want to transact with firms that are "altruistic." Although customers interpret firms' actions generously, they boycott firms when there is convincing evidence that their expectations regarding altruism are violated. This reaction has been used to help explain why prices are sticky (Rotemberg 2005); investigate how customers react to brand extensions (Rotemberg 2008); and explore how fairness influences firms' pricing decisions (Rotemberg 2009). We will present evidence that the reductions in future purchases do not just reflect changes in customers' expectations about prices or product quality. Instead, the effects appear to influence

1. Other empirical evidence includes Zbaracki et al. (2004), who report findings from an extensive study of the cost of changing prices at a large industrial firm. Their findings include a series of anecdotes and quotations indicating that managers at this firm were concerned that frequent price changes would damage the firm's reputation with its customers.

how customers view less tangible characteristics of the firm and its brand. This evidence may be seen as direct support for Rotemberg's claim that customers' perceptions of whether firms are altruistic contribute to the value of a firm's brand (Rotemberg 2008).

Although our findings are clearly consistent with Rotemberg's models, this is not the only explanation for why customers may stop purchasing from a firm when they are antagonized by price adjustments. It is possible that firms are simply risk-averse and want to minimize the risk of future antagonism.[2] Other explanations are also suggested by previous research on the role of reputations and firm brands (Wernerfelt 1988; Tadelis 1999).

I.B. Customer Antagonism When Prices Increase

We measure the response to *downward* price adjustments. This is a natural place to start, as it invokes a strong strawman: lower prices lead to higher sales. There is a complementary stream of research that studies customer antagonism in response to upward price adjustments. The origins of this work can be traced to Phelps and Winter (1970) and Okun (1981). Okun introduces the label "customer markets" to describe markets with repeated transactions. He argues that if price increases antagonize buyers, then sellers may respond to unobserved price shocks by absorbing the cost changes rather than increasing their prices. Nakamura and Steinsson (2008) extend this intuition in a recent paper. In their model, price rigidity serves as a partial commitment device, which enables sellers to commit not to exploit customers' preferences to repeatedly purchase from the same seller. The commitment mechanism is endogenous and relies on the threat that a deviation would trigger an adverse shift in customer beliefs about future prices.[3]

As support for their model, Nakamura and Steinsson (2008) cite two experimental papers. Renner and Tyran (2004) demonstrate that when buyers are uncertain about product quality, sellers are less likely to raise prices in response to cost increases if there is an opportunity to develop long-term relationships. This effect is particularly pronounced if buyers cannot observe the cost increases. The explanation offered for these effects is that the

2. We thank an anonymous reviewer for this suggestion.

3. In a related paper, Kleshchelski and Vincent (2009) investigate how customer switching costs create an incentive for firms to build market share. They demonstrate that this may prompt firms to absorb a portion of transitory cost increases without increasing prices.

sellers are willing to absorb the additional costs rather than risk antagonizing the buyers. These results complement an earlier experimental study (Cason and Friedman 2002) that investigates how variation in customers' search costs affects both their willingness to engage in repeated relationships and the resulting variation in prices. The authors show that higher search costs tend to result in more repeated relationships, increasing sellers' profits and leading to less variation in prices.

I.C. Other Relevant Research

Our findings are also related to previous work on intertemporal price discrimination. Research on the timing of retail sales argues that firms can profit by occasionally lowering prices and selling to a pool of low-valuation customers (Conlisk, Gerstner, and Sobel 1984; Sobel 1984, 1991). These arguments do not consider the possibility that customers who observe the lower prices may be less likely to make future purchases. Our findings represent a countervailing force that may limit a firm's willingness to price discriminate.

The findings are also relevant to recent work on price obfuscation. This literature recognizes that many firms adopt practices that make it difficult for customers to compare prices (Ellison 2006; Ellison and Ellison 2009). Price obfuscation may allow firms to mitigate the reactions that we document in this paper. If customers could not easily compare prices, then we would not expect the same outcomes. In this respect, our findings may help to explain the use of price obfuscation mechanisms.

Finally, we can also compare our findings with research on reference prices (Kahneman, Knetsch, and Thaler 1986). A challenge in the reference price literature is to identify the correct "reference" price. Some customers may use later prices as a reference against which to evaluate the price paid in earlier transactions. Under this interpretation, the findings are easily reconciled with the reference price literature: customers who later see the firm charging a lower price may conclude that they experienced a loss from overpaying in the past. We caution that this definition of the reference price is not unique, and alternative definitions may lead to different predictions.

I.D. Plan of the Paper

The paper continues in Section II with a detailed description of the experimental setting and the design of the first field

experiment. We present initial findings in Section III, which provide both context and motivation for the results that follow. We investigate how the outcome is moderated by the recency of customers' prior purchases in Section IV. In Section V we consider the price paid in the prior purchases, together with the persistence of the results across the 28-month measurement period. In Section VI we investigate three alternative explanations for the results, and then in Section VI we present a replication in a different product category. The paper concludes in Section VIII with a summary of findings and limitations.

II. Study Design

The primary field experiment was conducted with a medium-sized publishing retailer that manufactures and sells a range of approximately 450 products targeted at well-educated retail customers.[4] All of the products carry the retailer's brand name and are sold exclusively through the company's catalogs. At the time of our study the firm also operated an Internet site, but few customers placed their orders online. The products are durables with characteristics similar to those of books, computer software, and music. It is rare for customers to buy multiple units of the same item (few customers buy two copies of *Oliver Twist*), and so incremental sales typically reflect purchases of other items (buying both *Oliver Twist* and *David Copperfield*). The average interpurchase interval is 48 weeks.

Interviews with managers revealed that the firm engages in intertemporal price discrimination, charging relatively high regular prices interspersed with frequent shallow discounts and occasional deep discounts. A review of the firm's historical transaction data confirms that there had been wide variation in prices paid for the same item. Approximately one-fifth of transactions occur at the regular price, with most of the remaining transactions occurring at either a shallow (20%–30%) or a deep (50%–60%) discount. In Figure I we report a frequency distribution describing the discounts received in the two years before the Test Catalog was mailed.

4. The results of this study for a small subset of customers were previously described in Anderson and Simester (2004). This earlier paper uses data from this study (and two other studies) to investigate a different research question: comparing how prospective and existing customers react to discounts.

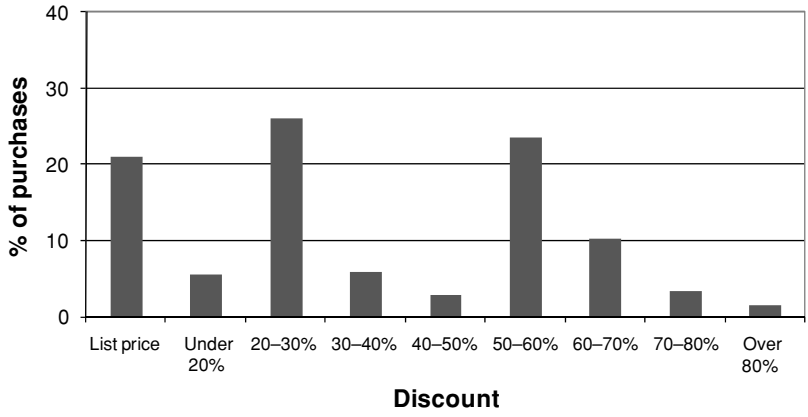

FIGURE I
Histogram of Pre-test Discounts
A histogram of the discounts received in purchases in the two years before the Test Catalog was mailed. The histogram was calculated using a randomly selected sample of customers (including customers not involved in the study).

II.A. Design of the Test Catalog

The field test encompassed two time periods separated by the mailing of the Test Catalog. Our data describe individual customer purchases in the eight years before the Test Catalog was mailed (the "pre-test" period) and the 28 months after this date (the "post-test" period). It will be important to remember that our measure of sales during the post-test period includes purchases from the Test Catalog itself. This allows us to rule out the possibility that the findings merely reflect intertemporal demand substitution.

There were two different versions of the Test Catalog: a "shallow discount" and a "deep discount" version. A total of 55,047 retail customers were mailed the Test Catalog, all of whom had previously purchased at least one item from the company. Approximately two-thirds of the customers (36,815) were randomly assigned to the Shallow Discount condition, and the remaining customers (18,232) were assigned to the Deep Discount condition.[5]

5. In Harrison and List's (2004) nomenclature, this is a "natural field experiment." They distinguish between "artefactual field experiments," which are the same as conventional lab experiments but with a nonstudent subject pool; "framed field experiments," which introduce field context; and "natural field experiments," which also occur in a field context and use subjects who do not know that they are participants in an experiment.

TABLE I
CHECK ON RANDOMIZATION PROCEDURES

	Shallow Discount condition	Deep Discount condition	Difference	*p*-value
Days since last purchase (recency)	646.04 (2.54)	647.92 (3.61)	−1.88 (4.42)	.67
Number of units purchased (frequency)	3.34 (0.02)	3.31 (0.03)	0.03 (0.04)	.42
Average price of units purchased (monetary value) ($)	133.26 (0.44)	132.69 (0.63)	0.57 (0.77)	.46
Sample size	36,815	18,232		

Note. Table I reports the mean values of each historical purchasing measure (calculated separately for each condition). The statistics are calculated using purchases during the eight-year pre-test period, prior to the mailing date for the Test Catalog. Standard errors are in parentheses. The *p*-values denote the probability that the difference between the deep discount and shallow discount averages will be larger than the observed difference, under the null hypothesis that the true averages are identical.

The decision to assign a larger fraction of customers to the shallow discount was made by the firm and was outside our control but does not affect our ability to interpret the results.

We confirm that the allocation of customers to the two conditions was random by comparing the historical purchases made by the two samples of customers. In Table I we compare the average *Recency*, *Frequency,* and *Monetary Value* (RFM) of customers' purchases during the eight-year pre-test period.[6] If the assignment were truly random, we should not observe any systematic differences in historical sales between the two samples. Reassuringly, none of the differences are significant despite the large sample sizes.

The Test Catalog was a regularly scheduled catalog containing 72 pages and 86 products. The only differences between the two versions were the prices on the 36 "Test Items." These 36 items were discounted in both versions, but the discounts were larger in the Deep Discount condition. In the Shallow Discount condition, the mean discount on the Test Items was 34%. In the Deep

6. "Recency" is measured as the number of days since a customer's last purchase. "Frequency" measures the number of items that customers previously purchased. "Monetary Value" measures the average price (in dollars) of the items ordered by each customer. The interpurchase interval (48 weeks) is much shorter than the average of the recency measure (646 days). These measures are not directly comparable: the interpurchase interval describes the time between purchases, whereas the recency measure includes some customers who will make no additional purchases.

Discount condition, the mean discount was 62%. These yielded mean prices of $133.81 and $77.17 on the 36 Test Items under the two conditions (compared to a mean regular price of $203.83). The sizes of the price discounts were chosen to be large enough to generate an effect, but not so large that they were outside the range of historical discounts (see Figure I). The other fifty items were all at their regular prices in both versions.

The prices of the Test Items were presented as "Regularly $x Sale $y" in the deep discount version, and as "Regularly $x Sale $z" in the shallow discount version. The regular price ($x) was the same under both conditions, but the sale price was lower in the Deep Discount condition ($y < $z). The use of identical text ensured that the discounted price was the only difference between the two versions and also explains why we used shallow discounts rather than the regular price as a control. There is considerable evidence that customers are sensitive to the word "Sale," even when prices are held constant (Anderson and Simester 1998, 2001). Charging the regular price as a control would have made it difficult to distinguish the effect of the price change from the "Sale" effect. By using the same wording under both conditions, we avoid this confound. As we will discuss, using shallow discounts as a control does not affect our ability to measure how price adjustments affect sales.

II.B. Mailing and Pricing Policies after the Test Catalog

The firm agreed to use the same mailing policy for all customers once the experimental manipulation was over. To confirm compliance, we obtained data describing post-test mailings for a random sample of 16,271 of the 55,047 customers involved in the test (approximately 30% of the full sample). Customers in the Deep Discount condition received a mean of 48.47 catalogs in the post-test period, compared to 48.77 catalogs in the Shallow Discount condition. The difference in these means is not close to statistical significance.

Notice also that the paper's key finding is that sales are lower under the Deep Discount condition (compared to the Shallow Discount condition). As we will discuss, the deep discounts led to an increase in sales from the Test Catalog itself. If this affected the firm's subsequent mailing decisions, it would tend to increase the number of catalogs mailed to customers under this condition. This could not explain the *decrease* in aggregate post-test sales.

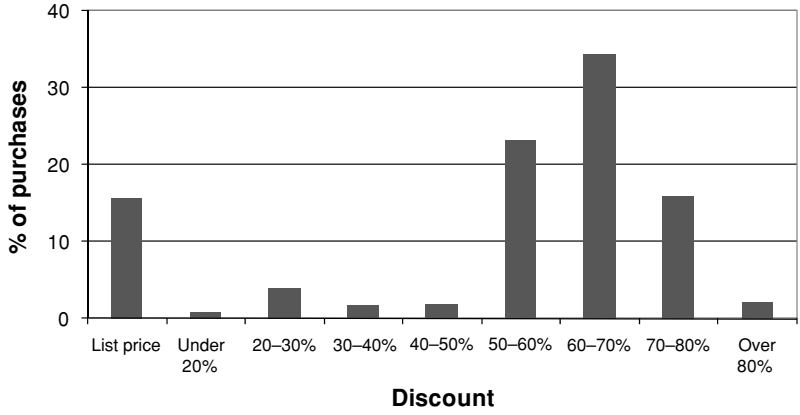

FIGURE II
Histogram of Post-test Discounts
A histogram of the discounts received in the 28-month post-test period. The histogram was calculated using a randomly selected sample of customers (including customers not involved in the study).

It is also useful to consider the firm's pricing policy after the Test Catalog was mailed. The distribution of discounts received in the post-test period is reported in Figure II. The use of discounts persisted, with a noticeable increase in the frequency of deep discounts. Most catalogs mailed during the post-test period contained at least some items with deep discounts. Because customers under the two conditions received the same downstream catalogs, this change in policy does not affect our ability to compare the behavior of customers under the two experimental conditions.

II.C. Predictions

Just over 47% of the customers (25,942) had not purchased any of the Test Items before receiving the Test Catalog. Of the remaining customers, very few customers had purchased more than one or two of the 36 Test Items, with less than 0.3% purchasing more than 10. As a result, the Test Catalog offered customers an opportunity to purchase discounted Test Items that they did not already own. Because our measure of post-test demand includes purchases from the Test Catalog itself, a standard model predicts higher demand in the Deep Discount condition (which offered lower prices). We label this the "Low Prices" prediction:

Low Prices. Post-test demand is *higher* under the Deep Discount condition than under the Shallow Discount condition.

Customer antagonism suggests an alternative prediction. Customers under the Deep Discount condition were more likely to see prices in the Test Catalog that were lower than what they paid for the same item. If this antagonized customers then we might expect lower sales in this condition. We label this the "antagonism" prediction:

Antagonism. Post-test sales are *lower* under the Deep Discount condition than under the Shallow Discount condition.

This prediction requires only a relatively simple model of customer behavior. It is sufficient that if customers see prices lower than what they paid they are less likely to make additional purchases. Because this is more likely to occur under the Deep Discount condition, we expect fewer sales under that condition.[7]

We will also consider three interactions. The first interaction distinguishes the 29,105 customers who had purchased one of the 36 Test Items before receiving the Test Catalog from the 25,942 customers who had not. We only expect customers to be antagonized if they see lower prices on items that they have purchased. Therefore, we expect a more negative (less positive) reaction to the deep discounts among the 29,105 customers with a prior purchase:[8]

Past Purchase Interaction. Deep discounts have a more negative (less positive) impact on post-test demand among customers who have previously purchased one of the Test Items.

There is an alternative explanation for this interaction. Because customers are unlikely to purchase the same item twice, customers who had already purchased may have been less likely to take advantage of the deep discounts in the Test Catalog. We will investigate this explanation in Section VI, together with other alternative explanations.

Customer antagonism may also depend upon how recently customers had purchased. Customers who purchased a long time ago may find it harder to remember the price that they paid. Moreover, those who can remember the price may be less antagonized

7. Notice that we do not need to rule out favorable responses if customers see a price higher than what they previously paid. We expect lower post-test sales under the Deep Discount condition irrespective of whether there is a positive response (or no response) to seeing higher prices. However, in Section IV we do investigate how the price that customers previously paid affects their response to the deep discounts.

8. It is possible that customers may be concerned about whether other customers were antagonized. However, previous studies have consistently reported that decision makers show little concern for the sufficiency of other customers' outcomes (Guth and van Damme 1998; Selten and Ockenfels 1998).

TABLE II
POST-TEST PURCHASES: COMPLETE SAMPLE

	Shallow discount	Deep discount	Difference
	Revenue and orders per customer		
Revenue ($)	159.28	157.16	−2.11
	(2.19)	(3.06)	(3.78)
Orders: all post-test orders	1.05	1.04	−0.02
	(0.01)	(0.02)	(0.02)
Orders: from the Test	0.022	0.036	0.014**
Catalog itself	(0.001)	(0.002)	(0.002)
Orders: from other catalogs	1.03	1.00	−0.03
	(0.01)	(0.02)	(0.02)
% of customers with at least	35.4	35.0	−0.4
1 post-test order	(0.3)	(0.4)	(0.4)
Number of customers	36,815	18,232	
	Composition of orders		
Average item price ($)	101.49	101.14	−0.35
	(0.51)	(0.74)	(0.90)
Average number of items	1.57	1.60	0.03
per order	(0.01)	(0.01)	(0.02)

Note. Table II reports the averages of each post-test sales measure for the respective samples. All statistics are calculated using only purchases from the 28-month post-test period (after the Test Catalog was mailed). The *orders*, *revenue*, and *any post-test order* statistics are calculated using the 36,815 and 18,232 customers under the Shallow and Deep Discount conditions, respectively. The average item price and number of items per order are calculated using the 13,038 and 6,388 customers from each condition who placed at least one order during the post-test period. Standard errors are in parentheses.

**Significantly different from zero, $p < .01$.

*Significantly different from zero, $p < .05$.

because they have had additional opportunities to consume. We will investigate whether the time between a customer's prior purchase and the mailing date for the Test Catalog contributes to the outcome:

Time Interaction. Deep discounts have a less negative (more positive) effect on post-test sales if the time since a customer's previous purchase is longer.

We also investigate whether there is evidence that customers who paid higher prices for Test Items were more antagonized by the deep discounts. This "past price" interaction is somewhat complicated by the experimental design and so we will defer discussion of this issue to when we present the results.

III. INITIAL RESULTS

In this section we present initial findings that provide both context and motivation for the main analysis that follows. To begin, we ask how the deep discounts affected post-test purchases by

TABLE III
POST-TEST PURCHASES: CUSTOMERS AT RISK OF BEING ANTAGONIZED

	Shallow discount	Deep discount	Difference
Revenue and orders per customer			
Revenue ($)	506.02	415.31	−90.71*
	(23.70)	(30.91)	(40.15)
Orders: all post-test orders	3.27	2.78	−0.48*
	(0.13)	(0.18)	(0.22)
Orders: from the Test Catalog	0.10	0.12	0.02
	(0.01)	(0.02)	(0.02)
Orders: from other catalogs	3.17	2.66	−0.50*
	(0.13)	(0.17)	(0.22)
% of customers with at least	72.9	66.0	−6.9*
1 post-test order	(1.5)	(2.2)	(2.6)
Number of customers	933	459	
Composition of orders			
Average item price ($)	97.61	97.82	0.21
	(1.98)	(2.87)	(3.53)
Average number of	1.66	1.64	−0.02
items per order	(0.06)	(0.06)	(0.06)

Note. Table III reports the averages of each post-test sales measure for the respective samples. The samples are restricted to customers who had paid a high price (above the shallow discount price) for a Test Item within three months of the Test Catalog mailing date. The *orders*, *revenue*, and *any post-test order* statistics are calculated using all of the customers in the respective samples. The average item price and number of items per order are calculated using customers who placed at least one order during the post-test period. Standard errors are in parentheses.

**Significantly different from zero, $p < .01$.

*Significantly different from zero, $p < .05$.

the full sample of 55,047 customers. These results are reported in Table II, where we describe the average number of orders placed by customers under the Deep and Shallow Discount conditions. We distinguish between orders from the Test Catalog itself and orders from other catalogs (not the Test Catalog) during the post-test period.[9] We also present the average revenue earned, the average size of customer orders, and the average price of the items that they purchased.

Because this overall comparison aggregates customers for whom the outcome was positive with others for whom it was negative, it reveals few significant differences between the two samples. We can illustrate this by returning to the motivating example

9. When customers call to place an order, they are asked for a code printed on the back of the catalog. This code is also printed directly on the mail-in order form. The data we received contain the catalog code for each order, so we can identify orders from the Test Catalog and orders from other catalogs. There are a small number of orders for which the catalog code is not available, including some Internet orders. Fortunately these instances are rare (there were very few Internet orders during the period of the study) and have little effect on the findings.

used in the Introduction. In Table III we focus on customers who paid a high price (above the shallow discount price) for a Test Item in the three months before the Test Catalog was mailed. In the Introduction we anticipated that these were the customers most likely to have been antagonized by the deep discounts.

The deep discounts resulted in fewer post-test orders from these customers. Under the Deep Discount condition customers placed an average of 2.78 post-test orders compared to 3.27 orders under the Shallow Discount condition. The difference (0.48 orders or 14.8%) is statistically significant and can be fully attributed to the prices in the Test Catalog. The findings also help clarify the source of the effect. The differences result solely from changes in the number of orders placed, rather than changes in the composition of those orders; there is no difference between conditions in either the average number of items per order or the average prices of the items purchased. The reduction in orders at least partly reflects an increase in the proportion of customers who placed no orders during the post-test period (34.0% under the Deep Discount condition versus 27.1% under the Shallow Discount condition). It is this result that led us to describe the effect as a customer boycott.

The decrease in orders is a strong result: merely sending these customers a catalog containing lower prices reduced purchases. The effect is large, and is precisely the outcome that managers appear to anticipate when stating they are reluctant to adjust prices for fear of antagonizing their customers (Hall and Hitch 1939; Blinder et al. 1998). Notice that customers who had purchased recently (Table III) are systematically more valuable than the average customer in the study (Table II). Although they place similar-sized orders, the recent purchasers order a lot more frequently in the post-test period. Because they are more valuable, the cost to the firm is greatly amplified. Although confidentiality restrictions prevent us from reporting detailed profit results, sending the deep discount version to these 1,392 customers would have lowered the profits earned from these customers by over $93,000 (compared to sending the shallow discount version).

It is also helpful to remember factors that cannot explain this result. Most importantly, the difference in orders cannot be explained by differences in the customers themselves. Our analysis compares customers under the two experimental conditions, and random assignment (which we verified in Table I) ensures that there are no systematic differences between customers under the two conditions. We can also see that the difference in post-test

orders is not due to a difference in the number of orders from the Test Catalog itself. Orders from the Test Catalog represent only a very small proportion (less than 3%) of overall post-test demand. Moreover, orders from the Test Catalog were actually slightly higher under the Deep Discount condition, presumably due to the lower prices. Finally, because our measure of sales includes purchases from the Test Catalog, the result cannot be explained by intertemporal demand substitution (forward buying). Acceleration in purchases to take advantage of the Test Catalog discounts would not affect this measure of total orders.

One explanation for the results, which does not depend on customer antagonism, is that the deep discounts may have changed customers' expectations about the availability of future discounts. If customers delayed their future purchases in anticipation of these discounts, it could explain a reduction in post-test sales. As we acknowledged in the Introduction, we cannot completely rule out this explanation. However, in Section VI we will present results that suggest this cannot be a complete explanation for all of the findings.

The initial analysis in this section focused on the sample of customers whom we expected to be most susceptible to the antagonism prediction. In the next section we extend the focus to all of the customers in the study, and measure how the effect was moderated by whether customers had previously purchased a Test Item and (if so) how recently they had purchased it.

IV. The Past Purchase and Time Interactions

To directly estimate the *past purchase* and *time* interactions, we use a multivariate approach. Because our initial analysis revealed that the primary effect is upon the number of orders placed, rather than the composition of those orders, we use a "count" of orders placed as the dependent variable (we later also use revenue and profit as dependent variables). We estimate this variable using Poisson regression, which is well suited to counting data.[10] Under this model, the number of orders from customer i (Q_i) is drawn from a Poisson distribution with parameter λ_i:

(1)
$$\text{Prob}(Q_i = q) = \frac{e^{-\lambda_i}\lambda_i^q}{q!}, \quad \text{where } q = 0, 1, 2, \ldots, \text{ and } \ln(\lambda_i) = \boldsymbol{\beta}\mathbf{X}_i.$$

10. In Section V we investigate the impact on revenue and profits. This analysis demonstrates that the results are also robust to using an OLS specification.

To estimate how the outcome was moderated by the time interaction since a customer's prior purchase, we use the following specification:

$$(2) \qquad \boldsymbol{\beta}\mathbf{X}_i = \beta_1 \text{DeepDiscount}_i + \beta_2 \text{DeepDiscount}_i \times \text{Time}_i + \beta_3 \text{Time}_i + \theta Z_i.$$

These variables are defined as follows:

Deep Discount$_i$. A binary variable indicating whether customer i was in the Deep Discount condition.

Time. The log of the number of months between the Test Catalog mailing date and customer i's most recent purchase of a Test Item.

For completeness, the vector Z includes the log of the historical RFM measures as control variables. Because these variables do not vary systematically across the two conditions (Table I), their inclusion or exclusion has little impact on the coefficients of interest (β_1 and β_2). Among customers who previously purchased a Test Item, β_1 describes how receiving the deep discounts affected post-test orders by a "benchmark" customer, who purchased a Test Item immediately before receiving the Test Catalog (*Time* equals zero). As the time since the customer's prior purchase increases, the estimated impact of the deep discounts is moderated by β_2. The time interaction predicts that β_2 will have a positive sign: the longer the time since the prior purchase, the smaller the reduction in post-test sales.

Notice that this model preserves the benefits of the randomized experimental design. The coefficients of interest measure the percentage difference in post-test sales between customers who received the deep and shallow discount versions of the Test Catalog. As a result, the coefficients of interest cannot be explained by differences in customer characteristics between the two experimental treatments, or by intervening competitive or macroeconomic events. We rely heavily on this feature of the study as it allows us to rule out a wide range of alternative explanations. The model is estimated separately on the 28,642 customers who had previously purchased a Test Item and on the 26,405 customers who had not previously purchased a Test Item.[11] Coefficients and standard errors for both models are reported in Table IV.

11. Joint estimation of the two models yields the same pattern of results, but the findings are more difficult to interpret (they require three-way interactions).

TABLE IV
POISSON REGRESSION: IMPACT OF DEEP DISCOUNTS ON POST-TEST ORDERS

	Customers with prior test item purchases (1)	Customers without prior test item purchases (2)
Deep discount	−0.078**	0.085
	(0.020)	(0.071)
Deep discount × time	0.026**	−0.017
	(0.007)	(0.012)
Time (since prior test item purchase)	−0.084**	−0.382**
	(0.006)	(0.007)
Recency (since any purchase)	−0.249**	
	(0.005)	
Frequency	0.620**	0.466**
	(0.005)	(0.012)
Monetary value	0.038**	0.048**
	(0.010)	(0.014)
Intercept	0.926**	1.247**
	(0.051)	(0.081)
Log likelihood	−53,680	−30,308
Sample size	29,105	25,942

Note. Table IV reports the coefficients from estimating equation (2) on each subsample of customers. The dependent variable measures the number of orders made during the post-test period. Asymptotic standard errors are in parentheses.
**Significantly different from zero, $p < .01$.
*Significantly different from zero, $p < .05$.

Column (1) focuses on customers with a prior Test Item purchase. For these customers β_1 is negative and significant, indicating that for our benchmark customer (for whom *Time* equals zero) the deep discounts led to a 7.8% reduction in post-test sales. This replicates our univariate results and confirms that merely sending customers a catalog containing lower prices led to fewer orders by these customers. For the same customers, β_2 is positive and significant. This result is consistent with the *Time* interaction and confirms that the loss of sales is smaller when more time has passed since a customer's earlier purchase.

The findings also reveal that the drop in post-test sales was limited to customers who had previously purchased one of the discounted items. For customers who had not previously purchased a Test Item (column (2)), the difference in sales between the two conditions was not significant. This is consistent with the *Past*

For customers who had not previously purchased a Test Item, we calculate the *Time* since their most recent purchase of any item. This is the same as our *Recency* measure, and so we omit *Recency* from this model.

Purchase prediction, which anticipated that the deep discounts would only reduce sales among customers who had previously purchased one of the Test Items.[12]

We can use the coefficients in Table IV to calculate how much time is required to elapse for the deep discounts to have a positive effect on post-test sales. Given coefficients of -0.078 for β_1 and 0.026 for β_2, the net impact of the deep discounts equals zero when *Time* is equal to 20.1 months.[13] We conclude that the estimated impact of the deep discounts was negative if customers purchased a Test Item within approximately two years of receiving the Test Catalog. This time interaction has an important additional implication for the firm. Recall that the initial findings in the preceding section (and the *Recency* and *Time* coefficients in Table IV) confirm that customers who purchased recently are systematically more valuable—they purchase significantly more frequently in the post-test period. Because the reduction in post-test sales was focused on these recent customers, there was a disproportionate effect on overall sales. Specifically, 13,830 customers purchased a Test Item within 20.1 months of receiving the Test Catalog. These customers represented just 25% of the sample but contributed approximately 52% of post-test revenue. If the firm mailed the deep discount version (rather than the shallow discount version) to each of these 13,830 customers, its profits would decrease by approximately $155,000.[14]

Equation (2) imposes a functional form on the interaction between the *Time* and *Deep Discount* variables. We can relax this restriction by grouping customers into segments based on the time since their earlier Test Item purchase and directly estimating the impact of the deep discount on each segment. In particular, we group customers into four segments based on the timing of their earlier purchases: less than 250 days, 250 to 500 days, 500 to 750 days, or over 750 days since the Test Catalog was mailed. We then estimate the following Poisson regression model for each segment:

$$\boldsymbol{\beta}\mathbf{X}_i = \beta_1 \text{DeepDiscount}_i + \theta Z_i. \tag{3}$$

12. We did not expect these customers to be antagonized by the deep discounts, but we would have expected them to take advantage of the lower prices. Further investigation confirms that when we restrict attention to orders from the Test Catalog itself, we see a strong *increase* in sales under the Deep Discount condition, but this was offset by lower sales from subsequent catalogs.

13. Calculated as $e^{0.078/0.026}$ (recall that *Time* is measured in months and has a log specification).

14. Calculated as the difference in the average profit in each condition multiplied by the 13,830 customers.

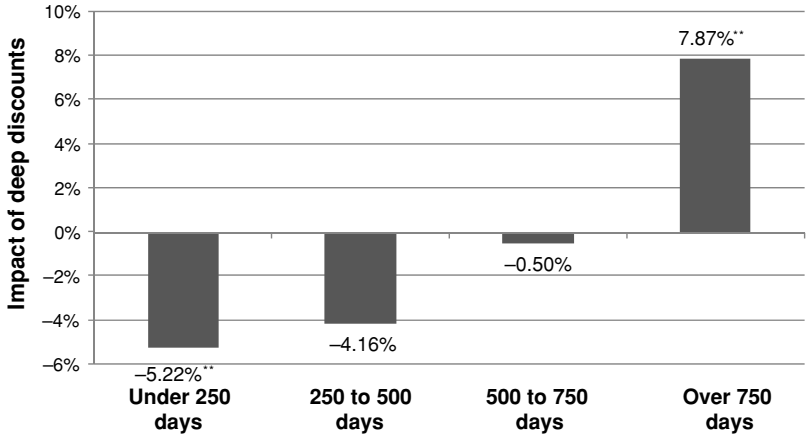

FIGURE III
Impact of Deep Discounts on Post-test Sales by Timing of Prior Test Item Purchase

The β_1 coefficients when estimating equation (3) on the four customer segments described above. Detailed findings are provided in Table V. *Significantly different from zero, $p < .05$. **Significantly different from zero, $p < .01$.

The β_1 coefficients for each segment are summarized in Figure III, and complete findings are provided in Table V.[15] The deep discounts led to a significant decrease in sales (-5.22%) from the 4,413 customers who had purchased within 250 days of receiving the Test Catalog, and a significant increase in sales ($+7.87\%$) from customers whose prior purchases occurred over 750 days ago. The results for the other two segments fall between these two results. This consistent monotonic relationship is reassuring, and confirms that we cannot attribute the role of time solely to the functional form imposed in equation (2).

The positive outcome for customers whose prior purchases were not recent (over 750 days ago) is worthy of comment. We offer two possible explanations. First, recall that customers had generally purchased at most one or two of the Test Items that were discounted in the Test Catalog. The deep discounts in the Test Catalog gave them the opportunity to purchase other items at low prices. This motivated our *Low Prices* prediction that sales would be higher under the Deep Discount condition. Second, it

15. In the next set of analyses we show that the effect is concentrated among customers who paid high prices in their previous purchases. We restrict attention to these customers in this analysis.

TABLE V
IMPACT OF DEEP DISCOUNTS ON POST-TEST SALES BY TIMING OF PRIOR TEST ITEM PURCHASE

	Segment 1 Under 250 days	Segment 2 250 to 500 days	Segment 3 500 to 750 days	Segment 4 Over 750 days
Deep discount	−0.052**	−0.042	−0.005	0.079**
	(0.015)	(0.027)	(0.037)	(0.029)
Recency	−0.152**	−0.182**	−0.262**	−0.306**
	(0.009)	(0.012)	(0.017)	(0.011)
Frequency	0.623**	0.707**	0.769**	0.677**
	(0.010)	(0.016)	(0.024)	(0.018)
Monetary value	−0.203**	−0.239**	−0.347**	−0.365**
	(0.016)	(0.026)	(0.033)	(0.025)
Intercept	1.724**	1.739**	2.397**	2.728**
	(0.085)	(0.132)	(0.170)	(0.136)
Log likelihood	−11,261	−6,364	−4,435	−8,459
Sample size	4,413	2,913	2,604	6,079

Note. Table V reports the coefficients from estimating equation (3) separately on the four customer segments. Asymptotic standard errors are in parentheses.
**Significantly different from zero, $p < .01$.
*Significantly different from zero, $p < .05$.

TABLE VI
IMPACT OF THE DEEP DISCOUNTS ON POST-TEST ORDERS FROM THE TEST CATALOG AND OTHER CATALOGS

Timing of prior purchase	Orders from the Test Catalog (%) (1)	Orders from other catalogs (%) (2)	All post-test orders (%) (3)
Under 250 days	31.6**	−6.5**	−5.2**
250 to 500 days	21.4	−4.9	−4.2
500 to 750 days	84.3**	−2.6	−0.5
Over 750 days	61.2**	6.6*	7.9**

Note. Table VI reports the *Deep Discount* coefficient (β_1) from estimating equation (3) separately on each customer segment and dependent variable. The results in column (3) are also reported in Table V (and in Figure III).
**Significantly different from zero, $p < .01$.
*Significantly different from zero, $p < .05$.

is possible the deep discounts persuaded some customers that the firm offered low prices, which may have prompted them to pay more attention to future catalogs (Anderson and Simester 2004). We investigate both explanations in Table VI, where we distinguish orders from the Test Catalog itself and post-test orders from other catalogs.

The deep discounts led to more orders from the Test Catalog for all customer segments. This is consistent with customers taking advantage of the deep discounts in that catalog to purchase items they had not already purchased (the *Low Prices* prediction). The effect is particularly strong for customers whose prior purchases were less recent. These are the same customers for whom we observe a positive outcome in Figure III. Among customers who had not purchased for over 750 days, the positive effect of the deep discounts extends beyond the Test Catalog to also increase orders from other catalogs by 6.6%. This is consistent with our second explanation: more favorable price expectations may have expanded the range of occasions on which these customers searched for products in the firm's catalogs.

We conclude that the findings offer strong support for the *Past Purchase* and *Time* interactions. The reduction in sales is limited to customers who had previously purchased a Test Item and is stronger for customers who had purchased more recently. We next consider how the price that customers paid in their earlier purchases influences the outcome. We will then investigate whether the findings persist throughout the post-test period, and whether they survive when we use post-test measures of revenue and profit as the dependent variable.

V. Additional Results

If price variation leads to customer antagonism, we might expect that customers who paid the highest prices in their earlier purchases would be most antagonized. There is considerable variation in the prices that customers paid for Test Items before receiving the Test Catalog. We can illustrate this variation by using this past price to group customers into three segments, which we label "Low," "Medium," and "High":

Segment	Past price level	Number of customers
Low	Less than the deep discount price	463
Medium	Between the deep and shallow discount prices	12,633
High	Above the shallow discount price	16,009

We do not expect the 463 customers who paid less than the deep discount price to be antagonized by the deep discounts. For the 12,633 customers in the Medium segment who paid between the

deep and shallow discount prices, the possibility of antagonism only arises under the Deep Discount condition, as this was the only condition where prices were lower than what customers previously paid. In contrast, all of the 16,009 customers in the High segment paid above the shallow discount price and saw lower prices when they received the Test Catalog. These customers may be antagonized under both experimental conditions.

Without knowing how customers respond to small and large price differences, we cannot make a clear *ex ante* prediction about how the outcome will vary across the High and Medium segments. We can illustrate this uncertainty by considering two extreme examples. If customers have the same negative reaction to any observed price reduction, then customers in the High segment will have the same reaction under both experimental conditions. In contrast, customers in the Medium segment will react negatively in the Deep Discount condition but not in the Shallow Discount condition. Our comparison of the two conditions will reveal a negative response in the Medium segment and no effect in the High segment. A second (equally extreme) behavior is that customers react only to large price differences. It is possible that none of the customers in the Medium segment observe a large enough price difference to prompt a reaction, whereas the outcome in the High segment depends upon the amount customers paid. Customers who paid the most may react under both conditions, whereas others may only react under the Deep Discount condition. Comparison of the two conditions will reveal no effect in the Medium segment and a negative effect for some customers in the High segment.

We can investigate this issue empirically. Our experimental design ensures that within each segment customers are randomly assigned to the Deep and Shallow Discount conditions. Therefore, we can measure how the deep discounts affected post-test sales in each segment by estimating equation (3) separately for each segment. The coefficients, which are reported in Table VII, reveal clear evidence that the past price plays an important role. Customers in the High segment reacted adversely to the deep discounts, but we do not observe any reaction in the Medium segment. Together, these results suggest that although customers react negatively to large price differences, they may be willing to overlook small price differences.

These results also have an important implication for the firm. Customers who pay higher prices tend to be systematically more

TABLE VII
THE ROLE OF THE PRICE PAID IN EARLIER TRANSACTIONS

	Past price segment		
	Low	Medium	High
Deep discount	0.469	0.008	−0.133**
	(0.256)	(0.036)	(0.024)
Deep discount × time	−0.169**	−0.004	0.050**
	(0.063)	(0.013)	(0.009)
Time (since prior test item purchase)	0.045	−0.069**	−0.102**
	(0.055)	(0.010)	(0.007)
Recency (since any purchase)	−0.113*	−0.253**	−0.246**
	(0.056)	(0.008)	(0.006)
Frequency	0.489**	0.648**	0.606**
	(0.061)	(0.008)	(0.007)
Monetary value	0.061	0.112**	0.008
	(0.073)	(0.017)	(0.013)
Intercept	0.056	0.477**	1.128**
	(0.037)	(0.090)	(0.068)
Log likelihood	−965	−21,449	−31,196
Sample size	463	12,633	16,009

Note. Table VII reports the coefficients from estimating equation (3) separately on the three customer segments. The Low segment includes customers who paid less than the deep discount price for a Test Item before the Test Catalog; customers in the Medium segment paid between the deep and shallow discount prices; customers in the High segment paid over the shallow discount price. The dependent variable in all three columns measures the number of units purchased during the post-test period. Asymptotic standard errors are in parentheses.

**Significantly different from zero, $p < .01$.

*Significantly different from zero, $p < .05$.

valuable: the 16,009 customers in the High segment represent approximately 29% of the sample but contribute 47% of the post-test profit. The concentration of the effect among these customers amplifies the impact on the firm's profits.

We conclude that the reduction in demand is concentrated in the segment of customers who had previously paid a high price (above the shallow discount price) for a Test Item. We will focus on this segment in the remainder of our analysis. We next consider whether the findings persist throughout the post-test period, and then investigate whether they survive when we use post-test measures of revenue and profit as the dependent variable.

V.A. Does the Adverse Outcome Persist?

To evaluate the persistence of the results we divided the 28-month post-test period into two 14-month subperiods. We estimate equation (2) on both subperiods and report the findings in

TABLE VIII
ADDITIONAL RESULTS: REVENUE, PROFITS, AND PERSISTENCE OF THE EFFECT

	Start of the post-test period (1)	End of the post-test period (2)	Revenue (3)	Profits (4)
Deep discount	−0.139**	−0.127**	−65.83**	−48.42**
	(0.033)	(0.035)	(19.72)	(14.55)
Deep discount × time	0.044**	0.056**	21.30**	15.46**
	(0.013)	(0.013)	(6.74)	(4.98)
Time	−0.117**	−0.085**	−31.07**	−21.00**
	(0.010)	(0.011)	(6.22)	(4.59)
Recency	−0.249**	−0.242**	−78.24**	−59.14**
	(0.009)	(0.009)	(5.43)	(4.01)
Frequency	0.598**	0.616**	204.23**	150.84**
	(0.009)	(0.010)	(4.26)	(3.14)
Monetary value	0.040**	0.025**	58.22**	45.61**
	(0.018)	(0.018)	(6.03)	(4.45)
Intercept	0.375**	0.495**	277.56**	190.79**
	(0.095)	(0.097)	(37.66)	(27.80)
Log likelihood	−19,674	−20,635		
Adjusted R^2			.229	.230
Sample size	16,009	16,009	16,009	16,009

Note. The dependent variables in the four columns are the number of units purchased in the first fourteen months of the post-test period (column (1)); the number of units purchased in the last fourteen months of the post-test period (column (2)); total revenue earned from each customer in the post-test period (column (3)); and total profit (calculated as revenue minus cost of goods sold) earned from each customer in the post-test period (column (4)). Columns (1) and (2) are estimated using Poisson regression (asymptotic standard errors are in parentheses). Columns (3) and (4) are estimated using OLS (standard errors are in parentheses). All of the models are estimated using the 16,009 customers who had previously paid a high price for a Test Item.

**Significantly different from zero, $p < .01$.

*Significantly different from zero, $p < .05$.

Table VIII (columns (1) and (2)). Comparing β_1 and β_2 between the two data periods reveals no significant difference in either coefficient (they are also not significantly different from the findings reported for these customers in Table VII). However, the results do imply a different time cutoff for which there is a negative response to the deep discounts. At the start of the post-test period, the negative response extends to customers who purchased within 23.5 months of receiving the Test Catalog. At the end of the period, the results indicate that the response is only negative for customers whose prior purchase occurred within 9.7 months of the Test Catalog. We conclude that although the negative effect survives for more than a year after the Test Catalog was mailed, there is evidence that the effect decays over time.

V.B. Profit and Revenue

The regression results reported so far consider only the number of orders purchased during the post-test period. To evaluate robustness, we also analyzed two additional dependent measures: the *Total Revenue* and *Total Profit* earned from each customer during the post-test period (where profit is calculated as revenue minus cost of goods sold). Both variables are continuous rather than count measures, and so we estimate the models using OLS. The findings are reported in Table VIII (columns (3) and (4)). They confirm that deep discounts lead to a reduction in both revenue and profits. This effect diminishes as the time since a customer's previous purchase increases. The interval at which the net effect is zero is approximately 22 months for both metrics.

We conclude that lower prices lead to fewer purchases by some customers. This effect is strongest among customers who had recently paid a high price to buy an item on which the price is later lowered. Unfortunately these include many of the firm's most valuable customers and this magnifies the importance of the effect. Readers who are solely interested in the existence of this effect may want to read ahead to Section VII, where we replicate the findings with a different company in a different product category. However, other readers may be interested in reviewing alternative explanations for these outcomes. We address this issue in the next section.

VI. Alternative Explanations

The findings reported in the preceding sections are consistent with customer antagonism. In this section we evaluate three alternative explanations for the findings. We begin by considering the possibility that customers delayed their purchases in anticipation of future discounts. We then consider the role of both quality signals and demand depletion.

VI.A. Delayed Purchases

Coase recognized that customers may respond to intertemporal discounts by delaying their purchases in anticipation of future discounts (Coase 1972). In our context, the discounts under the Deep Discount condition may have alerted customers to the possibility of future discounts. If customers responded by delaying their subsequent purchases, this might explain the reduction in

TABLE IX
WERE CUSTOMERS DELAYING IN ANTICIPATION OF FUTURE DISCOUNTS? POST-TEST PURCHASES AT DIFFERENT DISCOUNT LEVELS

	Coefficients of interest	
Discount threshold	(β_1) Deep discount	(β_2) Deep discount × time
Any discount level	−0.130**	0.055**
	(0.018)	(0.007)
Discounts of at least 10%	−0.115**	0.049**
	(0.020)	(0.008)
Discounts of at least 20%	−0.114**	0.048**
	(0.020)	(0.008)
Discounts of at least 30%	−0.112**	0.048**
	(0.021)	(0.008)
Discounts of at least 40%	−0.111**	0.050**
	(0.021)	(0.008)
Discounts of at least 50%	−0.115**	0.050**
	(0.021)	(0.008)
Discounts of at least 60%	−0.109**	0.049**
	(0.024)	(0.009)
Discounts of at least 70%	−0.091**	0.066**
	(0.034)	(0.013)

Note. Each row in Table IX reports the β_1 and β_2 coefficients from reestimating equation (2) when the dependent variable counts post-test units sold under each discount threshold. In each of these eight models the samples are restricted to customers who had paid a high price (the sample size for each model is 16,009). Asymptotic standard errors are in parentheses.

**Significantly different from zero, $p < .01$.

post-test orders. As we have already acknowledged, we cannot fully rule out this explanation, but we can investigate it using several approaches.

First, if customers were waiting for future discounts, the decrease in post-test sales should be larger at prices that represent either small discounts or no discounts. It is these purchases that we would expect customers to forgo while waiting for larger discounts. To investigate this prediction we recalculate post-test sales using different discount thresholds. For example, "Discounts of at least 60%" counts the number of units in the post-test period at discounts of at least 60%.[16] We reestimated equation (2) using eight different discount thresholds and report the results from each of these eight models in Table IX.

16. Notice that we shift from "orders placed" to "units purchased" because discounts are defined at the unit level.

Though not statistically significant, there is weak evidence of a trend in the results. At higher discount thresholds the *Deep Discount* coefficients are slightly less negative, which is consistent with customers waiting for larger discounts. Moreover, the implied *Time* at which the effect switches from negative to positive becomes shorter, which indicates that we only observe a negative effect among customers whose prior purchases were more proximate to receiving the Test Catalog. The findings also confirm that sales decrease even for deeply discounted items, and this is difficult to reconcile with customers waiting for future discounts.

Our second approach recognizes that deeply discounted prices were not unique to the Test Catalog. Over 38% of purchases in the pre-test period were made at discount levels of at least 50% (see Figure I).[17] If customers had purchased at deep discounts in the past, it seems less likely they would be surprised by the deep discounts in the Test Catalog or that they would change their behavior by delaying future purchases. Therefore, we can evaluate this alternative explanation by investigating whether the decrease in demand was smaller among customers who had purchased at deep discounts in the pre-test period. We do so by including a measure of past discounts in our model,

$$\begin{aligned}\boldsymbol{\beta}\mathbf{X}_i = {} & \beta_1 \text{DeepDiscount}_i + \beta_2 \text{DeepDiscount}_i \times \text{Time}_i \\ & + \beta_3 \text{Time}_i + \beta_4 \text{DeepDiscount}_i \times \text{MaximumPastDiscount}_i \\ & + \beta_5 \text{MaximumPastDiscount}_i + \theta \mathbf{Z}_i,\end{aligned} \tag{4}$$

where $\text{MaximumPastDiscount}_i$ measures the largest percentage discount *on any item* that customer i purchased during the pre-test period. We report the coefficients for this model in Table X. We also include an alternative model in which $\text{MaximumPastDiscount}_i$ is replaced with a binary variable (60% Past Discount_i) identifying customers who had previously purchased at a discount of at least 60%.

Neither of these interaction terms approaches statistical significance. The response to the deep discounts was apparently not affected by the size of the discounts that customers had received in the past. This is not what we would expect if the Test Catalog changed customers' price expectations; the effect should be

17. Many pre-test purchases were made at even deeper discounts (14.6% were made at a discount of at least 60%). Recall that we used a 60% discount level under the Deep Discount condition because it was consistent with past discount levels.

TABLE X
PRICE EXPECTATIONS

	(1)	(2)
Deep discount	−0.126**	−0.146**
	(0.053)	(0.026)
Deep discount × time	0.051**	0.051**
	(0.010)	(0.009)
Deep discount × maximum past discount	−0.022	
	(0.089)	
Deep discount × 60% past discount		0.049
		(0.032)
Time	−0.097**	−0.102**
	(0.007)	(0.007)
Maximum past discount	1.004	
	(0.062)	
60% past discount		−0.018
		(0.019)
Recency	−0.239**	−0.246**
	(0.006)	(0.006)
Frequency	0.528**	0.607**
	(0.008)	(0.007)
Monetary value	0.143**	0.008
	(0.015)	(0.013)
Intercept	0.098	1.134**
	(0.089)	(0.068)
Log likelihood	−31,024	−31,194
Sample size	16,009	16,009

Note. Table X reports the coefficients from estimating equation (4) on the customers who had previously paid a high price for a Test Item. Asymptotic standard errors are in parentheses.
**Significantly different from zero, $p < .01$.
*Significantly different from zero, $p < .05$.

strongest for customers who had not previously purchased at a large discount.

We make a final observation about this explanation. At the start of Section III we reported the average price of the items purchased in the post-test period (Table III). Increased willingness to wait for discounts should lead to customers paying lower average prices. However, the average price paid (per unit) was not significantly different between the two conditions. We conclude that delay in anticipation of future discounts does not appear to be a complete explanation for the results.

VI.B. Quality Signals

It is possible that customers interpreted the deep discounts on the Test Items as a signal that these items were of inferior

TABLE XI
QUALITY SIGNALS

	Non-test items (1)	Test items (2)
Deep discount	−0.152**	−0.047
	(0.021)	(0.039)
Deep discount × time	0.057**	0.046**
	(0.008)	(0.015)
Time (since prior test item purchase)	−0.153**	−0.176**
	(0.007)	(0.013)
Recency (since any purchase)	−0.196**	−0.180**
	(0.006)	(0.012)
Frequency	0.752**	0.514**
	(0.006)	(0.011)
Monetary value	0.116**	−0.177**
	(0.012)	(0.021)
Intercept	0.483**	0.917**
	(0.062)	(0.111)
Log likelihood	−40,946	−17,817
Sample size	16,009	16,009

Note. Both models are estimated using the 16,009 customers who had previously paid a high price for a test item. In column (1) the dependent variable measures the number of non-Test Items purchased during the post-test period. In column (2) the dependent variable measures the number of test items purchased. Asymptotic standard errors are in parentheses.

**Significantly different from zero, $p < .01$.

*Significantly different from zero, $p < .05$.

quality. To investigate this possibility, we can restrict attention to non-Test Items ("other items"). We do not expect discounts on the Test Items to signal information about the quality of these other items, and so if the effect persists for these other items, it is unlikely to be due to an adverse quality signal.

The Test Items accounted for only 22% of the 94,487 post-test purchases, so the remaining 78% of the purchases were for the approximately 400 other items sold by the firm. We repeat our earlier analysis when distinguishing between post-test sales for Test Items and other items and report the results in Table XI. The pattern of findings is unchanged. Indeed, the findings are stronger for these other items than for the Test Items, presumably because some customers took advantage of the deep discounts (on the Test Items) in the Test Catalog. We conclude that the decrease in sales cannot be fully explained by customers using the deep discounts as a signal that the Test Items are poor quality.

An alternative interpretation is that the deep discounts lowered the perceived quality of all of the products sold by the firm.

This is a relatively implausible explanation in this setting. Most customers in the study had made multiple previous purchases and had received a large number of catalogs from the firm. On average, each customer had purchased approximately 3.3 units prior to receiving the Test Catalog. This increases to 5.2 units among customers who had purchased a Test Item within 20.1 months of receiving the Test Catalog. Moreover, customers' first purchases occurred on average over two years before they received the Test Catalog and in the intervening period they had received many of the firm's catalogs (approximately once every two to four weeks). Given this extensive experience with the firm, it seems unlikely that customers' overall perceptions of the quality of the firm's products would be affected by the prices in a single catalog.

We conclude that it is unlikely that the results can be explained by changes in customers' expectations about product quality. It also appears unlikely that customers were merely waiting in anticipation of future discounts. It is for this reason that we suggested in the Introduction that the reputation effects reflect intangible brand attributions, rather than inferences about future prices or product quality. We next consider an alternative explanation for the *Past Purchase* and *Time* interactions.

VI.C. Demand Depletion

Because customers are unlikely to buy the same item twice, those who had already purchased a Test Item had fewer opportunities to take advantage of the deep discounts on these items in the Test Catalog. Similarly, customers with *recent* purchases may have had their immediate needs satisfied, which may also have diminished demand for discounted items in the Test Catalog.

It is helpful to consider the limits of this explanation. Demand depletion may explain why there was *no increase* in sales in the Deep Discount condition (compared to the Shallow Discount condition): customers could not take advantage of the lower prices because their demand was depleted. However, the two versions were mailed to equivalent groups of customers (whose demand was equivalently depleted) and so demand depletion *cannot* explain why sales *decreased* in the Deep Discount condition. Notice also that orders from the Test Catalog account for less than 3% of the post-test orders. If we restrict attention to orders from catalogs other than the Test Catalog, the findings survive and are actually strengthened by omission of the Test Catalog orders. This result cannot be explained by a diminished response to the Test Catalog

itself.[18] We conclude that depletion of demand for Test Items from the Test Catalog cannot fully explain the reduction in aggregate post-test orders.

In the next section we present the findings from a second study conducted with a clothing retailer. The study provides an opportunity to replicate the results with a different company and product category.

VII. Replication

To investigate whether the effects generalize, we conducted a second study with a separate catalog firm that sells private label clothing. The study investigates the impact of mailing a "Sale" catalog containing discounts in the days immediately after Christmas, a period when many customers expect clothing to be discounted. If we can show that the key findings extend to this setting we can be more confident that customer antagonism plays an important role in contributing to price stickiness.

Approximately 110,000 customers who had previously purchased from the firm were randomly assigned to equal-sized Test and Control conditions.[19] Those in the Control condition received a 16-page "Sale" catalog containing 132 items, all of which were discounted below their regular prices. The average regular price of the 132 items was \$67.24 and the average discount was 45%. Customers in the Test condition were not mailed this catalog. We measured (post-test) sales over the subsequent thirty months, including purchases from the Sale catalog.

Before describing these findings, we recognize that there are two important differences from the previous study in the experimental manipulation. First, it was standard for this company to mail a catalog containing discounts immediately after Christmas, and so the experimental manipulation was *suppression* of the Sale catalog. Because lowering the price was standard behavior, it appears implausible that merely receiving this Sale catalog changed customers' expectations about the frequency of future discounts. The second difference is that in the previous study all customers received the Test Catalog; we merely varied prices between the

18. Similarly, diminished demand from the Test Catalog cannot explain the findings in the second half of the post-test period. Over 99.7% of purchases from the Test Catalog occur within the first fourteen months of the post-test period.

19. Comparison of historical purchases confirmed that the assignment of customers to these two conditions was random.

TABLE XII
REPLICATION RESULTS: IMPACT OF SALE CATALOG ON POST-TEST ORDERS FROM THE APPAREL CATALOG

	Customers who had previously paid more than the sale price (1)	Customers who had paid a price no higher than the sale price (2)	Customers who had not previously purchased a sale item (3)
Received sale catalog	−0.024**	0.012	0.016**
	(0.005)	(0.013)	(0.003)
Recency	−0.208**	−0.177**	−0.220**
	(0.002)	(0.006)	(0.001)
Frequency	0.624**	0.673**	0.541**
	(0.003)	(0.007)	(0.002)
Monetary value	0.002**	0.081**	0.075**
	(0.011)	(0.028)	(0.005)
Intercept	0.478**	0.047	0.511**
	(0.048)	(0.116)	(0.021)
Log likelihood	−74,498	−10,392	−362,018
Sample size	14,699	1,630	93,142

Note. The dependent variable in all three models measures the number of orders made during the post-test period. Column (1) is estimated using the 14,699 customers who had previously paid a higher price for one of the items in the Sale catalog. Column (2) is estimated using the 1,630 customers who had previously paid a price equal to or less than the price in the Sale catalog. Column (3) is estimated using the 93,142 customers who had not previously purchased one of the items in the Sale catalog. Asymptotic standard errors are in parentheses.

**Significantly different from zero, $p < .01$.

*Significantly different from zero, $p < .05$.

two conditions. This ensured that any effects could be attributed to variation in prices and not to other information in the catalog. In this second study, the variation is sending the Sale catalog itself. As a result, we cannot distinguish which feature(s) of the Sale catalog prompted the observed outcomes.[20]

The findings are reported in Table XII. We report three models using different samples of customers. In column (1) we estimate the model using customers who had previously purchased one of the items in the Sale catalog and paid more than the Sale catalog price. In column (2) we use customers who had previously paid a price equal to or less than the price in the Sale catalog, whereas

20. If customers under the Test condition (who did not receive the Sale catalog) called to order one of the items in the Sale catalog, they were also charged the discounted price. Therefore, under a strict interpretation, the experiment manipulates the information that customers received about price changes, rather than the prices themselves. In a catalog setting this distinction is less important, because customers cannot respond to price changes that they do not know about.

in column (3) we use customers who had not previously purchased one of the items in the Sale catalog.

Customers who had previously paid a higher price for one of the discounted items (column (1)) made 2.4% *fewer orders* if they received the Sale catalog. This is again a strong result. Our measures include purchases from the Sale catalog, and the drop in sales can be fully attributed to merely receiving the Sale catalog containing lower prices. Customers who had not previously purchased one of the discounted items (column (3)) placed 1.6% *more orders* during the 30-month post-test period if they received the Sale catalog. These outcomes are both significantly different from zero and significantly different from each other ($p < .01$). They suggest that customer antagonism contributed to a net reduction in orders of 4.0%, which is the difference between a 1.6% increase and a 2.4% decrease.

The results in this replication may initially appear counterintuitive. Customers would not be surprised to receive a Sale catalog in the days after Christmas, and so we might not expect them to react adversely to the low prices in this catalog. On the other hand, customers do not know *which* products will be discounted in the Sale catalog, and it is likely that they experienced regret if they saw that they could have purchased the same item at a lower price. This is precisely the situation anticipated by Rotemberg (2009). Recall that customers in his model only want to transact with firms that are "altruistic." An altruistic firm would not allow customers to experience regret, and so regret represents evidence that a firm is not altruistic.

The findings in this second study offer reassurance that the effects are replicable. They also provide evidence that they generalize to other markets, including markets in which customers recognize that prices vary over time. Without this evidence it would be more difficult to claim that the effects contribute to price stickiness to an extent that is relevant to our understanding of monetary policy.

VIII. Discussion and Conclusions

Although customer antagonism is recognized as a possible explanation for price stickiness, there has been virtually no empirical evidence that price adjustments can cause customer antagonism and reduce demand. We present findings from two large-scale field tests that investigate how customers respond to

downward price adjustments. The findings reveal that many customers stop purchasing if a firm charges a lower price than they previously paid for the same item. We characterize the loss in demand as a customer boycott of the firm. The loss in profits was sufficient to cause the firms that participated in our two field experiments to reduce the frequency of price adjustments.

Both of the studies reported in the paper focus on products that are durables for which repeat purchases are rare. It is possible that the results might be different if we had used products that consumers purchase repeatedly. Customers may be less antagonized by lower prices on items they have already purchased if they can take advantage of the discounts. However, it is important to remember that the focus on durables in these two studies did not prevent customers from taking advantage of the discounts. Most customers had previously purchased very few of the discounted items, and so they could take advantage of the discounts by purchasing other items. Indeed, many customers did take advantage of the discounts: in the first study purchases from the Test Catalog itself were over 60% higher under the Deep Discount condition, whereas in the second study many customers who received the Sale catalog purchased from it.

Our measures may underestimate how customers react to seeing lower prices on items that they have previously purchased. When customers receive catalogs many of them simply throw them out without opening them. Random assignment ensures that for the Test Catalog this is likely to have occurred at equivalent rates under both conditions. However, if customers did not open the Test Catalog then the variation of prices within the catalog could not have affected their behavior. This will tend to weaken the aggregate differences between the two treatments and underestimate the individual effects on the customers who did see the discounts.

Although the data and the randomized experimental design allowed us to rule out many alternative explanations, we recognize that there are limitations. The most important limitation is that we do not have access to direct measures of customers' psychological reactions. This absence of intermediate process measures is a limitation common to many field studies.

NORTHWESTERN UNIVERSITY
MIT

REFERENCES

Amirault, David, Carolyn Kwan, and Gordon Wilkinson, "A Survey of the Price-Setting Behavior of Canadian Companies," *Bank of Canada Review* (2004), 29–40.

Anderson, Eric T., and Duncan I. Simester, "The Role of Sale Signs," *Marketing Science*, 17 (1998), 139–155.

——, "Are Sale Signs Less Effective When More Products Have Them?" *Marketing Science*, 20 (2001), 121–142.

——, "Long-Run Effects of Promotion Depth on New versus Established Customers: Three Field Studies," *Marketing Science,* 23 (2004), 4–20.

Apel, Mikael, Richard Friberg, and Kerstin Hallsten, "Micro Foundations of Macroeconomic Price Adjustment: Survey Evidence from Swedish Firms," *Journal of Money, Credit and Banking*, 37 (2005), 313–338.

Blinder, Alan S., Elie R. D. Canetti, David E. Lebow, and Jeremy B. Rudd, *Asking about Prices: A New Approach to Understanding Price Stickiness* (New York: Russell Sage Foundation, 1998).

Cason, Timothy N., and Daniel Friedman, "A Laboratory Study of Customer Markets," *Advances in Economic Analysis and Policy*, 2 (2002), Article 1.

Coase, Ronald H., "Durability and Monopoly," *Journal of Law and Economics*, 15 (1972), 143–149.

Conlisk, John, Eitan Gerstner, and Joel Sobel, "Cyclic Pricing by a Durable Goods Monopolist," *Quarterly Journal of Economics*, 99 (1984), 489–505.

Ellison, Glenn, "Bounded Rationality in Industrial Organization," in *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress, Richard Blundell, Whitney Newey, and Torsten Persson, eds. (Cambridge, UK: Cambridge University Press, 2006).

Ellison, Glenn, and Sara Fisher Ellison, "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 77 (2009), 427–452.

Fabiani, Silvia, Martine Druant, Ignacio Hernando, Claudia Kwapil, Bettina Landau, Claire Loupias, Fernando Martins, Thomas Matha, Roberto Sabbatini, Harald Stahl, and Ad Stokman, "The Pricing Behavior of Firms in the Euro Area: New Survey Evidence," Paper Presented at Conference on Inflation Persistence in the Euro Area at the European Central Bank, 2004.

Gordon, Robert J., "What Is New-Keynesian Economics?" *Journal of Economic Literature*, 28 (1990), 1115–1171.

Güth, Werner, and Eric van Damme, "Information, Strategic Behavior and Fairness in Ultimatum Bargaining: An Experimental Study," *Journal of Mathematical Psychology*, 42 (1998), 227–47.

Hall, Robert L., and Charles J. Hitch, "Price Theory and Business Behavior," *Oxford Economic Papers*, 2 (1939), 12–45.

Hall, Simon, Mark Walsh, and Anthony Yates, "How do UK Companies Set Prices?" *Bank of England Quarterly Bulletin* (1996), 180–192.

Harrison, Glenn W., and John A. List, "Field Experiments," *Journal of Economic Literature*, 42 (2004), 1009–1055.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler, "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review*, 76 (1986), 728–741.

Kleshchelski, Isaac, and Nicolas Vincent, "Market Share and Price Rigidity," *Journal of Monetary Economics*, 56 (2009), 344–352.

Nakamura, Emi, and Jón Steinsson, "Price Setting in Forward-Looking Customer Markets," Columbia University Working Paper, 2008.

Okun, Arthur M., *Prices and Quantities: A Macroeconomic Analysis* (Washington, DC: Brookings Institution, 1981).

Phelps, Edmond S., and Sidney G. Winter, "Optimal Price Policy under Atomistic Competition," in *Microeconomic Foundations of Employment and Inflation Theory*, E. S. Phelps, ed. (New York: W. W. Norton, 1970).

Renner, Elke, and Jean-Robert Tyran, "Price Rigidity in Customer Markets: An Empirical Study," *Journal of Economic Behavior and Organization*, 55 (2004), 575–593.

Rotemberg, Julio J., "Customer Anger at Price Increases, Changes in the Frequency of Price Adjustment and Monetary Policy," *Journal of Monetary Economics*, 52 (2005), 829–852.

——, "Expected Firm Altruism and Brand Extensions," Harvard Business School Working Paper, 2008.
——, "Fair Pricing," *Journal of the European Economic Association*, forthcoming, 2009.
Selten, Reinhard, and Axel Ockenfels, "An Experimental Solidarity Game," *Journal of Economic Behavior and Organization*, 34 (1998), 517–539.
Sobel, Joel, "The Timing of Sales," *Review of Economic Studies*, 51 (1984), 353–368.
——, "Durable Goods Monopoly with Entry of New Consumers," *Econometrica*, 59 (1991), 1455–1485.
Tadelis, Steven, "What's in a Name? Reputation as a Tradeable Asset," *American Economic Review*, 89 (1999), 548–563.
Weiss, Yoram, "Inflation and Price Adjustment: A Survey of Findings from Micro Data," in *Optimizing Pricing, Inflation, and the Cost of Price Adjustment*, E. Sheshinski and Y. Weiss, eds. (Cambridge, MA: MIT Press, 1993).
Wernerfelt, Birger, "Umbrella Branding as a Signal of New Product Quality: An Example of Signalling by Posting a Bond," *Rand Journal of Economics*, 19 (1988), 458–66.
Zbaracki, Mark J., Mark Ritson, Daniel Levy, Shantanu Dutta, and Mark Bergen, "Managerial and Customer Costs of Price Adjustment: Direct Evidence from Industrial Markets," *Review of Economics and Statistics*, 86 (2004), 514–533.