Article

AMERICAN MARKETING ASSOCIATION

Journal of Marketing Research
2024, Vol. 61(5) 872-890

DOI: 10.1177/00222437231220327
journals.sagepub.com/home/mrj
Sage

Canary Categories

Eric Anderson, Chaoqun Chen, Ayelet Israeli, and Duncan Simester

Abstract

Past customer spending in a category is generally a positive signal of future customer spending. Analyses of historical data at two retailers demonstrate that there exist "canary categories" for which the reverse is true. Purchases in these categories are a signal that customers are less likely to return to that retailer. The authors propose an explanation for the existence of canary categories and then develop a stylized model that illustrates four contributing factors: the probability that a customer finds their favorite brand, customers' willingness to substitute brands, the cost and attractiveness of visiting other stores, and expectations about future brand availability. The analysis uses both field data and experiments to investigate these factors. The findings suggest that canary categories exist (at least in part) because store assortments are not completely adjusted to local preferences. An implication is that canary categories are endogenous to each retailer; the same category may be a canary category at one retailer and a destination category at a competing retailer.

Keywords

assortment management, customer retention, churn, destination categories

Online supplement: https://doi.org/10.1177/00222437231220327

Submitted August 10, 2022

> There is always the nagging concern that a slow seller you delete might be an important product to some of your best customers, prompting them to defect to competitors.
>
> — Fisher and Vaidyanathan (2012)

Adjusting local assortments to local preferences requires not only adding products but also removing products to make shelf space available. Yet, choosing which products to remove is complex. When Walmart implemented Project Impact, removing 15% of its stockkeeping units, it was forced to quickly reverse many of its changes (for additional examples, see Fisher and Vaidyanathan [2012]). Optimizing local assortments is challenging, as it requires a detailed understanding of both substitution among items in a category and substitution among competing stores.

In this article, we contribute to the literature on assortment management by identifying a novel empirical effect that illustrates substitution among both products and stores. We use data from a large consumer packaged goods retailer to distinguish two types of product categories. The purchase of an item from the first type of category is a strong signal that a customer will not return to the store in the future. We label these "canary categories," as in "a canary in a coal mine." The reverse is true for the second type of category, which we label "bellwether categories." A purchase in a bellwether category is a strong signal that a customer will return to the store in the future. We confirm the existence of both types of categories at two retailers selling different types of products.

Canary categories share similar characteristics at the focal retailer. Compared with bellwether categories, in canary categories customers are less likely to find their favorite brands when visiting a store. Moreover, when their favorite brands are unavailable, customers are more likely to purchase a substitute brand, rather than simply forgo a purchase. Consequently, two types of purchases occur in canary categories: (1) customers purchase a brand that is their favorite and (2) customers purchase a substitute brand because their favorite brand is unavailable. The attrition effect arises due to the second

Eric Anderson is Professor of Marketing, Kellogg School of Management, Northwestern University, USA (email: eric-anderson@kellogg.northwestern.edu). Chaoqun Chen is Data Scientist, Stripe (email: chaoqunchen@stripe.com). Ayelet Israeli (corresponding author) is Associate Professor, Harvard Business School, Harvard University, USA (email: aisraeli@hbs.edu). Duncan Simester is Professor of Marketing, Sloan School of Management, Massachusetts Institute of Technology, USA (email: simester@mit.edu).

Table 1. Retail Assortment Literature.

Article	Key Finding	Identification
Discontinuing Products		
Food Marketing Institute (1993)	No loss in sales	Field study
Drèze, Hoch, and Purk (1994)	4% sales increase (test vs. control stores)	Field experiment
Boatwright and Nunes (2001, 2004)	11% sales increase	Field study
Broniarczyk, Hoyer, and McAlister (1998)	Importance of shelf space and favorite brands to maintain sales	Laboratory study
Borle et al. (2005)	Heterogeneity in reduced shopping frequency and quantity purchased among categories	Field study
Store Assortments and Store Choice		
Briesch, Dillon, and Fox (2013)	Destination categories affect store choice	Field study
Briesch, Chintagunta, and Fox (2009)	Importance of assortment versus price, advertising for store choice	Field study
Chernev and Hamilton (2009)	Assortment attractiveness and store choice	Laboratory study

group of customers. Many of these customers do not return to that store in the future and instead visit another store.

We formalize this relationship between the availability of customers' favorite brands and the canary category effect in an illustrative model, which identifies several factors that contribute to the effect. These include the probability customers will find their favorite brands when visiting a store, customers' willingness to purchase a substitute brand, customers' expectations about the future availability of their favorite brand, and the cost and attractiveness of visiting a different store. We investigate the model predictions associated with these factors using a combination of historical data and experiments. In our analysis of historical data, we examine how customer behavior changes when a favorite brand is missing and how this varies between canary and bellwether categories. In the experiments, we randomly vary the factors in our proposed mechanism and observe changes in participants' stated purchase rates, including whether they intend to return to the focal store in the future.

Our theoretical model and empirical analyses recognize that canary categories are an endogenous outcome that is influenced by both local customer preferences and retail assortment decisions. We show that customers have localized preferences and the focal retailer in our study partially adjusts assortments. However, the adjustments are incomplete, which leads to some customers not finding their favorite brand when they visit a store.

The relationship between the canary category effect and local store assortments has an important managerial implication. Canaries in coal mines warn miners of the danger of carbon monoxide. Canary categories warn retailers of the dangers of customer attrition due to gaps in store assortments. They are in some respects the opposite of destination categories (Briesch, Dillon, and Fox 2013); a destination category attracts customers to a store, whereas canary categories are associated with customers not returning to a store. Our findings suggest that the two concepts are linked. A canary category can contribute to customer churn at the focal retailer, nudging customers to shop at a competing retailer for their favorite brand. For the competing retailer, the same category becomes a destination category, attracting customers to the store. Thus, a canary category for one retailer may be a destination category for a competitor.

The article continues with a literature review in the next section. In "The Canary Category Effect" section, we document the canary category effect and establish that the findings both survive a broad array of robustness checks and replicate at another retailer. Then, we describe our proposed mechanism for the effect. We provide support for this proposed mechanism using historical observations and experimental evidence.

Literature Review

There is an extensive literature on retail assortments. Within this literature, we review two sets of studies. First, there is a body of research measuring how discontinuing items in a category impacts revenue. Second, there is a literature investigating the relationship between retail assortments and store choice. We summarize the articles and key results in Table 1.

Discontinuing Items

A series of studies have investigated the impact of discontinuing items by removing them from a retailer's assortment. A 1993 industry study reports that reducing assortments in six grocery categories in three retail chains did not lead to a significant reduction in category sales (Food Marketing Institute 1993). Drèze, Hoch, and Purk (1994) report findings from a large-scale field experiment in which they removed items with low sales and replaced them with additional facings of more popular items (holding total shelf space constant). They report a 4% increase in aggregate sales in the test stores compared with the control stores. Similarly, Boatwright and Nunes (2001, 2004) investigate the outcome of a field study (i.e., natural experiment) at an online grocery store in which 94% of categories experienced large reductions in low-selling items. They report an 11% increase in sales across 42 categories and attribute this to a reduction in clutter following the removal of redundant items.

Broniarczyk, Hoyer, and McAlister (1998) report findings from two laboratory studies that measure the impact of reducing assortments on participants' perceptions of a retailer's assortment. They investigate the role of total item count, total category shelf space, and whether a participant's favorite item is available. Their findings suggest that retailers can reduce total item counts without adversely impacting customer perceptions, as long as total category shelf space is maintained and a customer's favorite item remains available. Together these findings suggest that, at least in some categories, retailers may be able to modestly reduce the number of items they sell without adversely impacting customer perceptions or sales.

When adjusting assortment size, research has documented heterogeneous effects among categories. Borle et al. (2005) study the impact of reducing assortments at an online grocery retailer and report that the negative impacts are large in some categories but small in other categories (particularly less frequently purchased categories). They conclude that if managers can identify categories in which the impacts are small, there is an opportunity for managers to target assortment reductions in these categories without harming overall store image or customer traffic.

Store Assortments and Store Choice

A second group of articles related to ours focuses on store choice and retail assortments. Briesch, Dillon, and Fox (2013) investigate the role of destination categories and make a similar point as Borle et al. (2005) about heterogeneous effects. They show that retailers' merchandising decisions—including assortment, feature advertising, display, and prices—have a larger impact on customer store choice in some categories than in other categories. An important managerial challenge is identifying the categories in which assortment reductions will have a large versus small effect on customer retention. As we will discuss, identifying canary categories offers a relatively straightforward process for accomplishing this.

Briesch, Chintagunta, and Fox (2009) estimate a structural model to compare how feature advertising, convenience, and prices impact store choice decisions. They conclude that, for most households, assortments are generally more important than retail prices in these decisions. While the number of brands positively impacts store choice, the number of items per brand tends to have a negative effect (consistent with the category clutter literature). Chernev and Hamilton (2009) conduct a series of laboratory experiments in which they vary assortment size among participants and measure the attractiveness of a retail store. They document that the relationship between assortment size is concave, implying that larger assortments may be less attractive.

Finally, the canary category effect can also be compared with the "harbinger customer" effect. Anderson et al. (2015) document that there are some customers who systematically purchase new products that fail. A purchase by these customers of a recently introduced product is a strong signal that the new product will not survive. The canary category effect is the inverse of the harbinger customer effect. Purchases by harbinger customers signal that a new product will fail. In contrast, purchases in canary categories signal that a "new" customer will fail (i.e., will not return to the retailer).

The Canary Category Effect

Our goal in this section is to establish the existence of the canary category effect. We do so by demonstrating that there exist categories for which purchases are a signal that customers are less likely to return to that retailer in the future. It is helpful to keep in mind that we only need to demonstrate an association between category purchases and customer retention to establish the canary customer effect. Therefore, throughout this section, we frame our findings as associations rather than causal results. In the "Support for the Proposed Mechanism from Online Experiments" section, we use online experiments that incorporate key features of our store data to establish causality.

Overview of the Analysis

The analysis proceeds in three steps. In Step 1, we identify customer retention by labeling whether customers survive. We split the data into two periods: preperiod and postperiod. We label customers as survivors or nonsurvivors by observing purchases in a specific store during the preperiod and measuring whether the customer purchased (in any category) in the same store in the postperiod.

In Step 2, we identify canary categories. We split customers into two subsamples (classification and holdout) and focus on purchases by the classification customers during the preperiod. For each category we observe which of the classification customers purchased the category in the preperiod and which of these customers survived (i.e., purchased from the same store in the postperiod). For example, if 100 classification customers purchased from a category in the preperiod and 30 of these customers returned to the same store in the postperiod, the average survival rate for that category is 30%. We use the average survival rate to classify categories into four groups (quartiles) and label the product categories with the lowest survival rates canary categories and the categories with the highest survival rate bellwether categories.

In Step 3, we use these categories to help explain variation in the survival of customers in the holdout set. For each of the holdout customers, we count the number of items purchased during the preperiod from the four groups of product categories. We show that if customers purchased from a canary category in the preperiod, there is a lower probability that they purchased during the postperiod.

ConvenienceStore Application

The primary analysis in this article uses data provided by a large U.S. chain of convenience stores located in urban and suburban locations, which sell consumer packaged goods in the grocery, health and beauty, and general merchandise categories.[1] For

[1] Data from this retailer, including this data set, has been used in several previous publications (the data was not obtained specifically for this study).

confidentiality reasons, we cannot identify the retailer, but for ease of exposition we refer to it as ConvenienceStore. Our analysis focuses on purchases a randomly selected sample of 103,366 customers made using ConvenienceStore's frequent shopping card between November 2003 and November 2005.[2] The customers represent a random sample of customers, and the data includes their complete purchase histories (when they purchased using the retailer's frequent shopping card). At ConvenienceStore, we treat the calendar year 2004 as the preperiod. The postperiod starts six months after the preperiod and runs until the end of the data period (July 1, 2005–November 12, 2005). We will show that our results are robust to different definitions of the pre- and postperiods.

Product categories are defined relatively finely (we later also replicate our findings using a more aggregate definition of categories). For example, there are over 30 beverage product categories, including "soda 2 liter," "water 12 pack," and "juices single." We restrict attention to product categories with at least 1,000 units sold across the entire two-year data period, and we also exclude a small number of categories labeled seasonal, services (film processing), or miscellaneous. After these omissions we have a total of 501 categories.

ConvenienceStore also provided weekly sales data for 111 stores for the same period. The stores are distributed across 13 states and were visited by customers in our panel. This weekly sales data is complete for every stockkeeping unit, and we use it to identify the product assortments offered by each of these 111 stores.

Step 1: Labeling Customers and Dividing Them into Classification and Holdout Groups

We start by labeling customers as "survivors" or not. If a customer made a purchase in any category in store s during the postperiod, we label the customer as having survived in that store. In particular, our outcome measure Survived_{is} for each customer i is equal to 1 if customer i purchased in store s during the postperiod and 0 otherwise. There are 227,268 unique customer × store combinations that purchased in the preperiod. The average postperiod survival rate across these combinations is 35.9%. The combinations that survived to the postperiod contributed 68.6% of the preperiod transactions.

To randomly divide customers into classification and holdout samples, we draw random numbers from a uniform distribution [0, 1], then we assign customers with random draws less than .5 to the classification sample and the remaining customers to the holdout sample. The classification sample contains 51,610 customers and 113,032 customer × store combinations, while the holdout sample contains 51,756 customers and 114,236 customer × store combinations.

We identify the items that each customer purchased during the preperiod. Using the purchases by customers in the classification set, we calculate an average survival rate for each product category c:

$$\text{Average survival rate}_c = \frac{\sum_i \sum_s q_{isc} \text{Survived}_{is}}{\sum_i \sum_c q_{isc}}. \quad (1)$$

This category-level measure is a weighted average of Survived_{is} using the customers in the classification set, weighted by the number of units from product category c that customer i purchased from store s in the preperiod (q_{isc}). In Web Appendix A, we show that our key findings hold when using different weighting schemes, including using binary indicators that describe whether a customer made any purchases in the category.[3]

Step 2: Identifying Canary Categories

We use a quartile split of the average survival rate to assign each of the 501 product categories to one of four groups. Group 1 contains the product categories with the highest average survival rate, and Group 4 contains categories with the lowest average survival rate.[4] It will be helpful to remember that Group 4 categories were purchased by classification customers who were least likely to purchase in the postperiod.

Step 3: Explaining Which Customers Survived in the Holdout Sample

Next, we switch the focus to customers in the holdout sample and investigate whether purchases of items in the different groups helps explain which customers were more likely to purchase in the postperiod. For each customer i and store s in the holdout sample we use purchases in the preperiod to calculate:

- Total Items_{is}: number of items purchased,
- $\text{Group } 1_{is}$: number of items in Group 1 categories purchased,
- $\text{Group } 2_{is}$: number of items in Group 2 categories purchased,
- $\text{Group } 3_{is}$: number of items in Group 3 categories purchased, and
- $\text{Group } 4_{is}$: number of items in Group 4 categories purchased.

[2] We replicate the main findings at several other retailers and in different time periods with data as recent as 2016 (for details, see Web Appendix A). We focus on ConvenienceStore in our analysis due to the richness of this data set, which includes both individual transactions and complete store assortments.

[3] The quantity purchased measure (q_{isc}) counts the number of packages purchased but does not adjust for package sizes. However, we also use revenue as an alternative weight when calculating the average survival rate. Revenue is sensitive to differences in package sizes.

[4] We subsequently investigate using a decile split instead of a quartile split of the average survival rate.

Table 2. ConvenienceStore Results.

	Model 1	Model 2	Model 1	Model 2
Total items	.074%**		.077%**	
	(.005%)		(.006%)	
Group 4: Canary category		−.151%**		−.161%**
		(.016%)		(.021%)
Group 3		.067%**		.115%**
		(.025%)		(.037%)
Group 2		.208%**		.209%**
		(.029%)		(.041%)
Group 1: Bellwether category		.197%**		.208%**
		(.019%)		(.029%)
Discount frequency	3.259%**	2.699%	1.252%*	.933%
	(.391%)	(.392%)	(.587%)	(.587%)
Average price	.235%**	.228%**	.198%**	.198%**
	(.028%)	(.028%)	(.041%)	(.041%)
Customer fixed effects	No	No	Yes	Yes
R^2	.3024	.3045	.5472	.5489

*$p < .05$. **$p < .01$.

Notes: The table reports coefficients from estimating Models 1 and 2 on ConvenienceStore data. Columns 1 and 2 are estimated without customer fixed effects, while Columns 3 and 4 include customer fixed effects. Fixed effects identifying the calendar months of the first and last purchase were estimated but are not reported. Eicker–Huber–White standard errors are reported in parentheses. The unit of analysis is a customer × store and the sample sizes are 114,236 (Columns 1 and 2) and 90,149 (Columns 3 and 4).

We then use the holdout sample to estimate two ordinary least squares (OLS) models:

$$\text{Model 1 Survived}_{is} = \alpha + \beta_1 \text{Total Items}_{is} + \text{Controls}_{is} + \varepsilon_{is}, \quad (2)$$

$$\text{Model 2 Survived}_{is} = \alpha + \beta_1 \text{Group1}_{is} + \beta_2 \text{Group2}_{is} + \beta_3 \text{Group3}_{is} + \beta_4 \text{Group4}_{is} + \text{Controls}_{is} + \varepsilon_{is}. \quad (3)$$

Model 1 is the benchmark model, which controls for the total number of items that customers purchased in the preperiod. This model measures whether there is a positive association between past and future purchases (which we would expect). In Model 2, we distinguish which items customers purchased. In particular, we investigate whether purchases of items from different category groups provide different information about whether customers returned in the postperiod. We also estimate analogous models using a logistic regression (see Web Appendix A).

The model includes four sets of control variables. For ease of exposition, in the equations for Models 1 and 2 we group the control variables under the label Controls:

$$\text{Controls}_{is} = \beta_a \text{Percent Discounted}_{is} + \beta_b \text{Average Price}_{is} + \boldsymbol{\tau}\,\textbf{First Month} + \boldsymbol{\eta}\,\textbf{Last Month}. \quad (4)$$

First, we control for the proportion of items that customer i purchased at a discount in store s (Percent Discounted). We use all the customer's transactions in the preperiod to measure the proportion of items purchased at a discount.[5] The second control variable, Average Price, controls for the average price paid for each item in the preperiod (in store s). This average is constructed by weighting the price of each item by the number of units purchased (we obtain similar results without weighting).[6] The **τ First Month** term refers to fixed effects identifying the calendar month that we first observed the customer making a purchase in store s in the preperiod.[7] We also include fixed effects identifying the last month that the customer purchased in store s in the preperiod (**η Last Month**). In Web Appendix A, we also investigate the inclusion of additional control variables that could play a role in customer retention, including variables to control for the distance to competing stores.

The unit of observation is a customer × store, and we estimate Models 1 and 2 using the 114,236 observations in the holdout sample. The findings are reported in Table 2. Because of the panel nature of the data, we use Eicker–Huber–White standard errors to adjust for heteroskedasticity.

The Model 1 findings confirm that the total number of items that customers purchased in the preperiod (Total Items) on its own explains significant variation in which customers will continue to purchase in the postperiod. As we would expect, purchasing more items in the preperiod is positively associated with customer retention. However, we see in Model 2 that when we distinguish among the categories that customers

[5] One explanation for why purchases of some products could help predict whether customers survive is customer deal sensitivity (or price sensitivity). If promotions are more common in some categories and customer deal sensitivity is correlated with survival, this could contribute to a canary category effect.

[6] The inclusion of both Average Price and Total Items effectively controls for total expenditure in the preperiod.

[7] Terms in bold are vectors. In this case they represent fixed effects by month.

purchased in the preperiod, there are different effects among the four groups. In particular, we observe positive effects for categories in Groups 1, 2, and 3, but a negative effect for canary categories (Group 4). Purchases of items in each group are informative, but they send different signals.

We caution that the findings do not show that canary categories are bought by more nonsurvivors than survivors. Instead, they show that, ceteris paribus, purchasing more products from canary categories is associated with a lower probability that customers returned in the postperiod. Holding the other variables constant, an increase of one purchase from Group 4 categories during the 12-month preperiod is associated with a .151% reduction in the probability that the customer survived (Model 2). This compares with a .197% higher probability of survival for every unit purchased in a bellwether category. The difference in these coefficients, .348%, highlights the importance of distinguishing between categories when using past purchases to predict churn.[8]

We can also illustrate this association using a simple comparison of means. Holding the total number of items purchased during the preperiod within .1 standard deviations of its mean, the average number of Group 4 items purchased during the preperiod is significantly lower among customers that survived than among customers that did not survive. Customers who survived purchased an average of 2.77 (.03) Group 4 items in the preperiod, compared with 3.21 (.04) for customers who did not survive (here and elsewhere, we report standard errors in parentheses).

It should not be surprising that purchases of some categories are more positively associated with customer retention than other categories. It is surprising that the association between purchases of canary categories and customer retention is negative. In particular, the coefficient on purchases from Group 4 categories in Model 2 is negative, not just smaller in magnitude than the coefficient associated with purchases from other categories.

During the preperiod, 27,669 of the customers visited multiple stores. For these customers, we can observe variation across stores in (1) whether the customer survived in each store and (2) which items the customers purchased in each store.[9] This variation enables us to estimate versions of Models 1 and 2 in which we include customer fixed effects. These findings are reported in Columns 3 and 4 in Table 2. Both models are estimated using the 90,149 customer × store observations for the 27,669 customers that visited more than one store. We see an essentially identical pattern of results. This result is notable because it reveals that the canary category effect is not explained by customer differences. For example, the result cannot be explained by more price sensitive customers purchasing canary categories, while less price sensitive customers purchase in bellwether categories.

Table 3. High Volume Top Five Canary and Bellwether Categories.

Canary Categories	Average Survival Rate	Bellwether Categories	Average Survival Rate
Juices single	57.67%	Milk gal 2%	80.96%
Water single	60.99%	Facial tissue	74.71%
Candy bars/nonchocolate	63.52%	Soup canned	74.13%
Chips single	60.99%	Nuts can/jar	73.52%
Soda 20-ounce singles	64.21%	Shelf stable seafood	73.04%

Examples of Canary Categories

If we restrict attention to high volume categories with at least 50,000 units sold in the preperiod, the categories with the largest and smallest average survival rates are shown in Table 3.

Among the high-volume canary categories, the category with the lowest average survival rate is “juices single” with an average survival rate of 57.67%. In contrast among bellwether categories, “milk gal 2%” has the highest average survival rate of 80.96%.

It is notable that the five canary categories are all products that are suitable for immediate consumption. Customers can consume individual drinks, candy, and snacks immediately after they leave the store. In contrast, the bellwether categories are more typical grocery products in which consumption is generally less immediate. If customers want to consume an item immediately, they may be more locked-in to the current store because traveling to a different store is less appealing. In the “An Explanation for the Canary Category Effect” section, we argue that lock-in could help explain the seemingly conflicting observation that customers buy on the current trip but do not return in the future.[10]

In Web Appendix B, we also compare the canary and bellwether categories on several additional dimensions. Canary categories appeared in 50% of baskets and contributed an average of \$1.94 to the \$11.96 overall average basket size, while bellwether categories appeared in 66% of baskets and contributed \$4.48. On average, canary categories generated less than half (43.6%) as much revenue as bellwether categories. This reflects a combination of fewer customers, lower

[8] Using the 114,236 observations in the holdout sample, the standard deviation of the number of units purchased in canary categories during the preperiod is 11.64 units. The standard deviation of the number of units purchased in canary categories during the same period is almost identical (11.65 units). The coefficients indicate that if a customer purchased one standard deviation more units from canary categories and one standard deviation fewer units from bellwether categories, then the expected probability of survival is approximately 4% lower.

[9] There is also variation across stores in each of the control variables.

[10] Given the opportunity for immediate consumption, it is possible that customers purchased in canary categories when traveling and stopping for snacks at stores that they do not often visit. This could explain why the probability of returning to these stores is lower. However, the data does not support this explanation. Customers purchased canary categories in approximately the same number of stores in which they purchased bellwether categories. Moreover, when purchasing items in canary categories, their baskets tended to contain more items from other categories and a wider range of categories. They were also less likely to purchase on impulse (see subsequent discussion).

units purchased per customer, and lower average unit prices paid.[11] The lower average price paid is attributable to lower regular prices, rather than larger discounts. The average discount received is smaller in canary categories than in bellwether categories (both in dollar terms and percentage terms).

The finding that customers received fewer discounts when purchasing canary categories does not necessarily imply that canary categories were promoted less frequently. Instead, it could reflect less deal sensitivity in these categories. Measuring the frequency of promotion is challenging, as we only observe prices in weeks that there were transactions. If we restrict attention to month × store × product combinations in which we observe transactions every week we can control for this confound. This analysis is available from the authors and indicates that canary categories were discounted at least as often as bellwether categories. It suggests that the evidence that customers received fewer discounts when purchasing canary categories is actually because they were less deal sensitive in these categories and thus more willing to purchase even when the item was not discounted. This in turn is consistent with both stronger brand preferences and a more urgent customer need.

Although canary categories contributed less revenue on average than bellwether categories, we should not conclude that they are an unimportant source of the retailer's revenue. The largest of the canary categories is over 50 times larger (in revenue) than the smallest bellwether category. Moreover, the average profit margin in canary categories (calculated as $\frac{\text{Revenue}-\text{Wholesale Cost}}{\text{Revenue}}$) is 48% higher in canary categories than in bellwether categories. Even though canary categories contributed less than half of the revenue of bellwether categories, their total profit contribution is only 24% less than bellwether categories.

We also compare spending in other categories by customers who purchased from canary and bellwether categories. Both groups of customers had similar overall spending in the chain, with a similar number of store visits and basket sizes (measured in both dollars and units). Using transactions from all categories, we see that customers who purchased in canary categories tended to pay slightly lower unit prices and receive smaller discounts. At least in part because they received smaller discounts, they contributed more total profit (per customer) than bellwether categories. We conclude that customers who purchase in canary categories are slightly more valuable (per customer) to the retailer than customers who purchase in bellwether categories.

Finally, three independent research assistants rate each category according to whether the products in the category tend to be (1) utilitarian or hedonic products, (2) search goods or experience goods, and (3) impulse purchases. While there are no differences between canary and bellwether categories on the first two sets of measures, we find that canary categories are significantly less likely to be purchased on impulse. Note that if a product category is always purchased on impulse, we might expect that purchases in the category (and the availability of brands in the category) would have less impact on store choices.

Table 4. Chain-Level Survival.

	Model 1	Model 2
Total items	.023%** (.003%)	
Group 4		−.075%** (.013%)
Group 3		.031% (.022%)
Group 2		.035% (.022%)
Group 1		.133%** (.022%)
R^2	.2406	.2413

$^{**}p < .01$.

Notes: The table reports the coefficients of interest from estimating modified versions of Equations 2 and 3 that use survival at the chain level instead of the individual store level (see discussion in text). Fixed effects and control variables were estimated but are not reported. Standard errors are reported in parentheses. The unit of analysis is a customer and the sample size is 103,366.

Robustness Checks

In Web Appendix A, we illustrate that the canary category effect is a very resilient finding that survives numerous robustness checks, such as:

- Grouping categories into deciles instead of quartiles,
- Alternative model specifications, including using a logistic model,
- Including additional controls, including the number of preperiod store visits, the amount spent per visit in the preperiod, whether a store is in an urban area, and the distance to alternative stores,
- Changing the measure used to calculate preperiod spending in the four category groups,
- Varying the length of pre- and postperiods,
- Changing the outcome measure to include whether there is a return visit, number of return visits, and time until a return visit, and
- Using a more aggregate definition of categories.

The key finding that preperiod purchases of Group 4 categories are associated with a lower probability of postperiod survival is extremely robust. The statistical significance of this finding survives every robustness check and every modification to the analysis that we have tried.

Chain-Level Survival

In the analysis in Table 2, the dependent variable measures whether a customer returned to the focal store in the postperiod. We can also measure whether a customer returned to any store in the chain the postperiod (Chain Survival). Across all customers and all stores, the average probability customers returned to the chain in the postperiod is 65.5%.[12] We repeat our analysis

[11] This result is robust to varying the length of the postperiod. Therefore, the result does not merely reflect that customers who purchased less frequently purchased items were less likely to purchase in the postperiod.

[12] By construction, there is a higher probability that a customer returned in the postperiod to any store in the chain than to a specific store in the chain.

Table 5. Out-of-Sample Predictive Performance.

	Base Model	Model 1	Model 2
Mean squared error	.1618	.1604	.1587
Mean Squared Error Difference			
Versus base model		−.00136** (.00009)	−.00303** (.00018)
Versus Model 1			−.00167** (.00016)
Classification Accuracy			
Precision	71.38%	71.53%	71.91%
Recall	59.84%	59.77%	61.23%
Other Accuracy Measures			
Akaike information criterion	75,725.8	75,221.4	74,642.4
Area under the receiver operating characteristic curve	.813	.816	.818

**$p < .01$.

Notes: The table reports out-of-sample predictive accuracy statistics for each model. The unit of analysis is a customer × store in the validation sample and the sample size is 76,107.

when using Chain Survival instead of Survival at an individual store. This results in three changes to the analysis: replacing the dependent variable, regrouping categories using Chain Survival, and recalculating the control variables at the chain level. In Table 4, we report the coefficients of interest and see that the findings replicate the pattern of coefficients reported in Table 2.

Further investigation reveals that the grouping of categories using chain-level survival (Chain Survival) is also similar to the grouping of categories using store-level survival (Survival).[13] If a category is a canary category at the chain level, it is also likely a canary category at the store level.

Replication at Another Retail Chain

To further demonstrate the robustness of the canary category effect, we replicate the results at a U.S. membership-based wholesale club that sells large package sizes at low prices. The results closely match the findings in Table 2. We provide complete details of the replication, including the results, in Web Appendix A. This replication offers several advantages. First, the data is more recent; the data period extends from January 2013 through July 2016. Second, we do not need to infer whether customers survive; instead, we can observe survival directly from whether they renew their memberships.[14] Third, the data period is longer, which enables us to measure whether the effect endures and is detectable more than one year after the preperiod. The results are even stronger when we consider more distant postperiods. The wholesale club data also enables us to rule out an alternative explanation (at least for this retailer). A limitation in the analysis in Table 2 is that we do not control for postpurchase marketing activities. It is possible that the canary category effect results from retailers changing their marketing actions based on the categories that customers purchase. We know with confidence that this did not occur at the wholesale club. As part of a research collaboration with this retailer, one of the authors helped this retailer improve how it targeted its marketing actions. At the time this data was collected, the firm targeted customers based on geographic location, membership type, and membership expiration date. It did not target its actions based on customers past purchases.[15]

Predicting Survival in a Holdout Sample

We can directly evaluate predictive performance by calibrating the model on a sample of customers and using it to predict customer survival in a separate holdout sample. This analysis uses three data sets:

1. Classification customers to classify categories into the four groups,
2. Calibration customers to estimate parameters, and
3. Validation customers to validate the models using a holdout sample of new customers.

We randomly assign a third of the customers to each sample. After classifying the categories, we estimate three logistic models: (1) a base model equivalent to Model 1 but without Total Items (or any measure of preperiod sales), (2) Model 1 (including Total Items), and (3) Model 2, where we replace each group variable with four binary variables identifying whether the customer purchased 0 times, 1 time, 2 times, or

[13] Of the 126 categories that are in Group 4 using chain-level survival, 84 (66.7%) are in Group 4 using store-level survival, with an additional 29 (23.0%) in Group 3. More generally, the pairwise correlation between the two average survival rates is $\rho = .76$.

[14] When using membership renewal decisions to define survival, we measure survival at the customer × chain level, rather than the customer × store level. For robustness, we also define survival at the store level by observing transactions in a postperiod that started 12 months after the preperiod. The pattern of findings is unchanged.

[15] In the "Support for the Proposed Mechanism Using Historical Observation" and "Support for the Proposed Mechanism from Online Experiments" sections, we investigate our proposed explanation for the canary category effect at ConvenienceStore. The evidence supports our explanation and appears unlikely to be explained by differences in postpurchase marketing activities.

more than 3 times in that group. We use the estimated parameters from each model to predict the survival of customers in the validation sample. In Table 5, we report and compare the out-of-sample mean squared error and several other (out-of-sample) measures of predictive accuracy.

We see significant reductions in prediction error in Models 1 and 2 compared with the benchmark model (using a paired comparison t-test). We also see a significant reduction in error in Model 2 compared with Model 1. Distinguishing which product categories customers purchase is more informative than simply counting how many products they purchase.

For completeness, we also investigate the robustness of the findings by randomly redrawing the calibration and validation samples 100 times. This enables us to calculate standard errors for various measures of fit—including precision, recall, Akaike information criterion, and the area under the receiver operating characteristic curve—and compare them across the three models. These results are reported in Web Appendix A and confirm that the improvement in the performance of Model 2 over the other two models is statistically significant ($p < .01$).

Summary

In this section, we have established the existence of the canary category effect. Purchases of canary categories signal that a customer is less likely to survive as a customer in the future. This is a resilient finding that survives numerous robustness checks, including replication at a U.S. membership wholesale club. In the next section, we propose an explanation for the canary category effect.

An Explanation for the Canary Category Effect

Our explanation for the effect argues that when customers are in a focal store, they face switching costs to visit an alternative store. If they discover their favorite brand is unavailable, they may purchase an alternative brand in the category but then switch to an alternative store for future visits. This can occur in any category. Therefore, to identify the distinguishing features of canary categories, we develop an illustrative model of this explanation. In this section, we describe the model and the key empirical predictions. We relegate a complete description of the model itself to Web Appendix C.

We consider a single focal store, two juice brands (Veryfine and Minute Maid), and a heterogeneous group of customers who differ in their brand preferences. Customers can purchase at most one unit of juice each period. Before visiting the focal store, customers have an outside option valued at W_A, which represents the option of purchasing at other stores. When at the focal store, the outside option has a value W_B. The relative values of W_A and W_B may be affected by a variety of factors, including geographic location of the stores. If the alternative store is located close to the focal store, then there is little additional cost of traveling from the focal store to the alternative store. In this situation, $W_A < W_B$ and visiting the alternative store is relatively attractive once you have arrived at the focal store. In contrast, suppose the alternative store is located on the other side of town from the focal store. In this case, traveling to store b is less attractive if you are at the focal store ($W_A > W_B$). The relationship between W_A and W_B may also depend on the purpose of the shopping trip. For example, if the customer's shopping basket at the focal includes perishables, switching to a competing retailer today may be less attractive ($W_A > W_B$). Similarly, if the category purchase is urgent, such as cough medicine for a sick child, an additional shopping trip to a competing store may be undesirable ($W_A > W_B$).

The model includes two time periods:

1. Period 1: Veryfine is unavailable at the focal store this period.
2. Period 2: Customers choose whether to return to the focal store.

In Period 1, all customers visit the focal store and upon arrival discover that Veryfine is not available. Customers who prefer Minute Maid purchase this brand, while customers with strong preferences for Veryfine leave the focal store and choose the outside option. Customers with weak preferences for Veryfine stay and purchase Minute Maid as an alternative (because Veryfine is unavailable).

In Period 2, all customers need to decide whether to visit the focal store again. However, now they are less confident that the focal store will carry Veryfine. The customers who prefer Minute Maid return, while customers with strong preferences for Veryfine do not return. The interesting question is: What do the customers with weak preferences for Veryfine do in Period 2? A key feature of the model is that in Period 1, when customers learn the focal store does not sell Veryfine, the cost of visiting the store (α) is sunk, which generates a lock-in effect. In contrast, in Period 2 they adjust their expectations about whether the store will sell Veryfine before visiting; the cost of visiting the store is not yet sunk, and customers are not yet locked in. As a result, customers with weak preferences for Veryfine who purchase Minute Maid in Period 1 divide into two groups. Those with the weakest preferences for Veryfine return to the focal store, but those with stronger preferences for Veryfine visit an alternative store in Period 2. This results in three segments of customers we will reference in the remainder of the article:

1. Segment A: Customers who do not purchase in either period. Customers who do not purchase from the focal store because they do not find their favorite brand. These are the customers with strong preferences for Veryfine.
2. Segment B: Customers who buy but do not return. These are the Veryfine customers who purchase Minute Maid from the focal store in Period 1 but do not return to the focal store in Period 2.
3. Segment C: Customers who buy and return. This segment includes two types of customers: customers

Table 6. Summary of the Empirical Analyses.

Market Feature	Historical Data	Experiments
Probability customers find their favorite brand	Frequency customers' favorite brands are unavailable Customer retention when favorite brand unavailable Local preferences and localized assortments	Experiment 1: Vary whether a customer's favorite brand is present.
Customers' willingness to substitute	Purchase of alternative brand if favorite brand is unavailable	All four experiments vary market features and measure willingness to substitute
Value of the outside option when at the focal store		Experiment 2: Vary urgency of purchase Experiment 3: Vary distance to alternative store
Expectations about future brand availability		Experiment 4: Vary whether stockout is temporary or permanent

whose favorite brand is available (customers who prefer Minute Maid) and customers who have weak brand preferences and are willing to switch brands if their favorite brand is unavailable. This second group of customers are the customers who weakly prefer Veryfine but are willing to buy Minute Maid if Veryfine is unavailable.

The Segment B customers are the source of the canary category effect; customers who purchase the alternative brand in Period 1 but do not return to the focal store in Period 2. The model highlights several market features that contribute to the canary category effect:

- Retail assortments are incomplete: A larger proportion of customers are unable to find their favorite brand when visiting a store.
- Customers' willingness to substitute: When customers have a strong preference for their favorite brand (compared with alternative brands), they will not purchase an alternative brand on the current trip, which reduces the size of Segment B.
- Value of the outside option when at the focal store: If the outside option becomes less attractive when the customer is already at the focal store (W_B is small relative to W_A), there is more lock-in, so more customers purchase the alternative brand in Period 1 but do not return in Period 2.
- Expectations about future brand availability: The canary category effect is larger when there is a larger adjustment in customer beliefs about Veryfine's availability in Period 2 (after it was unavailable in Period 1).

Note that some of these market features reflect retail characteristics, while other market features reflect customer characteristics.

In Table 6, we summarize how we investigate the role of these market features using data from two sources: historical transaction data from ConvenienceStore and experimental data from a series of Amazon Mechanical Turk and Prolific studies. In the "Support for the Proposed Mechanism Using Historical Observations" section, we use the historical data to document that the availability of customers' favorite brands varies among categories. We document the association between a missing favorite brand and customer purchase behavior. We also show how local preferences and adjustments in local assortments are associated with the canary category effect. In the "Support for the Proposed Mechanism from Online Experiments" section, we experimentally vary market features and measure how this changes the tendency to purchase on the current trip but not return to that store on future trips. For example, we show that if the unavailability of a preferred brand is temporary (due to a stockout), customers are more likely to return to the store on future trips, compared with when the item is permanently discontinued at that store.

Support for the Proposed Mechanism Using Historical Observations

We preface the analyses in this section by describing the relative difficulty of managing assortments in canary categories versus bellwether categories. In canary categories, there are slightly more brands in the market compared with bellwether categories, yet individual stores tend to offer slightly fewer brands in canary categories than in bellwether categories. As a result, in canary categories an average store offers 67.0% (2.0%) of the brands available in the market, compared with 75.2% (1.8%) in bellwether categories. The difference between 67.0% and 75.2% is statistically significant ($p < .01$) and has implications for both retailers and consumers. Retailers face a more difficult task of optimizing their brand assortments in canary categories; the retailer's choice set is larger, and they select a smaller set of brands. In turn, because a smaller proportion of the brands in the market are offered in a specific store, this increases the risk that customers are unable to find their favorite brands.

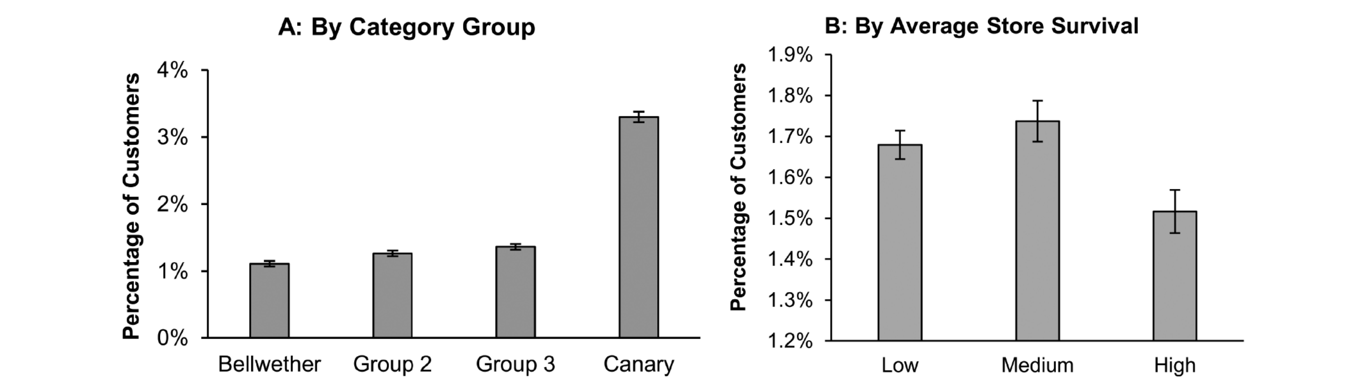

Figure 1. Proportion of Customers Who Do Not Find Their Favorite Brands.
Notes: The figure reports the percentage of customers whose favorite brands are not included in the assortment for a given store × category. The unit of analysis is a customer × category × store. Panel A groups observations by category group and Panel B groups observations by average store survival. The sample includes customers who visited a store and for whom we observed a favorite brand (using purchases at other stores). The sample size is 86,018. Error bars indicate 95% confidence intervals.

As assortments are more challenging to manage in canary categories, consumers are less likely to find their favorite brands in canary categories on a store visit. For each customer × store combination, we identify the customer's favorite brand using the customer's purchases in every store other than the focal store. We define the favorite brand as the brand purchased most often in other stores. On average, in canary categories there is a 3.30% (.04%) probability that a customer's favorite brand will be unavailable in a given store, compared with just 1.11% (.02%) in bellwether categories (Figure 1, Panel A).

The absence of a customer's favorite brand is a necessary condition in our explanation of the canary category effect. In the remainder of this section, we conduct three separate analyses that investigate subsequent customer behavior when a favorite brand is not available. Together, these provide convergent evidence that supports our explanation.

Category Purchase Rates

When a favorite brand is unavailable, customers have the option of purchasing another brand in the category or not making a category purchase in that store. We find strong evidence that customers are less likely to purchase when their favorite brand is unavailable. When a favorite brand is unavailable in the focal store, the purchase probability decreases by 3.1% (.14%) for bellwether categories and just 2.1% (.12%) for canary categories. As a robustness check, we repeat the analysis using all brands that a customer purchased at a store other than the focal store (not just the brand that the customer purchased most often). We obtain the same pattern of results.[16]

Collectively, these results illustrate two of the factors listed in Table 6 that contribute to the canary category effect. In canary categories, (1) customers are less likely to find a favorite brand when visiting a store, and (2) when customers do not find a favorite brand, they are more likely to substitute to an alternative brand (compared with bellwether categories).

Repeat Visits at a Store

We use the field data to investigate the relationship between the availability of a customer's favorite brand and customer retention. In particular, we ask whether customers are less likely to return to a store that does not sell their favorite brand. The analysis exploits variation in brand assortments within a category across stores. The results are reported in Web Appendix B and confirm that when a customer's favorite brand is unavailable, customers are significantly less likely to return to that store.[17]

These findings are consistent with evidence in the inventory management literature that customers are less likely to return to a store when an item they try to order is out of stock. For example, Anderson et al. (2006) measure stockouts during a 5-week treatment period and measure sales in a 13-month posttreatment period. They report that if all items a customer ordered during the treatment period were out of stock, posttreatment purchases were approximately 22% lower. Broniarczyk, Hoyer, and McAlister (1998) also investigate the impact when a customer's favorite brand is unavailable. Using two laboratory experiments, they report that removing a participant's favorite brand adversely impacts both perceptions of the retailer's assortment and stated store choice.

We can also exploit longitudinal variation in product assortments over time. We identify observations in which a customer's favorite brand was unavailable in a store at the

[16] Complete findings and robustness checks are reported in Web Appendix B.

[17] The results are estimated using OLS (for additional details, see Web Appendix B). The OLS coefficient estimates that if the proportion of favorite brands unavailable increases from 0% to 100%, the probability that the customer will return to the store decreases by 6.53% (2.21%).

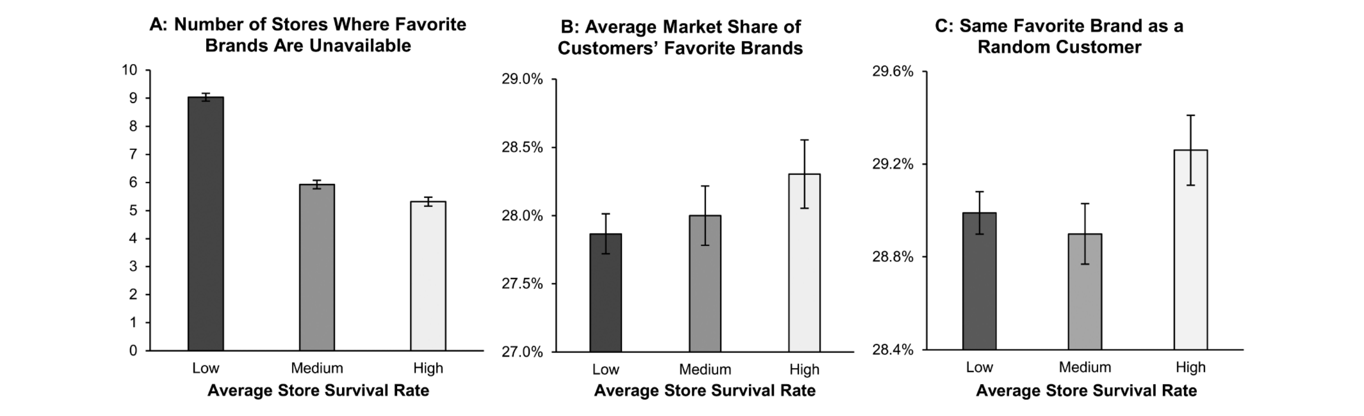

Figure 2. Canary Category Brand Preferences.
Notes: Panel A reports the average of the number of stores where a favorite brand is unavailable for low, medium, and high survival stores. Panel B reports the average market share of customers' favorite brands. Panel C reports the proportion of times a customer's favorite brand is the same as the favorite brand of a (different) randomly selected customer in a (different) randomly selected store. The unit of analysis is a customer × category × store. Sample sizes are 198,613 (Panels A and B) and 1,759,521 (Panel C). Error bars indicate 95% confidence intervals. Detailed findings are reported in Web Appendix B.

start of the preperiod. Then, we observe whether the brand was introduced at that store in the second half of the preperiod and measure whether the customer returned to that store in the postperiod. The findings confirm that the introduction of a customer's favorite brand is associated with increased customer retention.[18]

To complement this analysis, we investigate the relationship between the availability of a customer's favorite brand and the average customer retention rate at a store. To calculate an average store retention rate, we use the survivor labels for the customers that shopped in each of the 111 stores for which we observe product assortments (for the definition of survivor, see "The Canary Category Effect" section). We aggregate to the store level by averaging this survivor measure across customers, then we rank order the 111 stores using this average and split them into three groups of 37 stores each. We label these groups low survival, medium survival, and high survival stores.

Figure 1, Panel B, illustrates that in canary categories there is a strong association between whether customers can find their favorite brands and a store's average survival rate. Customers' favorite brands are more likely to be unavailable in low and medium survival stores than in high survival stores. As we previously discussed, the probability a favorite brand is unavailable is three times higher in canary categories than in other categories (see Figure 1, Panel A).

We caution that these results do not allow us to conclude that the unavailability of a customer's favorite brand causes the customer to not return and leads to low store survival. The historical variation in product assortments across stores and over time is endogenously determined by the retailer. We address this limitation with a series of online experiments in the "Support for the Proposed Mechanism from Online Experiments" section.

[18] This result is also estimated using OLS. The OLS coefficient estimates that if the proportion of favorite brands that were introduced in the second half of the preperiod increases from 0% to 100%, the probability that the customer will return to the store increases by 2.38% (1.19%). We provide complete details of this analysis in Web Appendix B.

Local Preferences and Localized Assortments

The previous analysis demonstrates that there is an association between average store retention rates and whether a customer finds their favorite brands. To further unpack the proposed mechanism, we consider two alternative explanations for this result. First, it is possible that stores with low retention rates also serve customers with more unique, or niche, preferences. Second, it is possible that the focal retailer recognizes variation in demand among stores but does not fully localize the assortment with a store. Both explanations could reduce the likelihood of finding a favorite brand.

To investigate heterogeneity in consumer preferences among stores, we use three different metrics. For the sake of brevity, we limit our analysis to canary categories and provide detailed descriptions of our complete analyses in Web Appendix B. First, we calculate how many of the 111 stores each customer's favorite brand is available in. Second, we calculate the market share (at the chain level) of each customer's favorite brand. Third, we calculate how often a customer in a store has the same favorite brand as another randomly selected customer in a different randomly selected store (within the same canary category). We average each of these metrics separately within low, medium, and high survival stores, and we report these averages in Figure 2 and Web Appendix B.

The findings reveal that in canary categories, customers in low survival stores tend to prefer brands that are less widely available and have lower market shares (Figure 2, Panels A

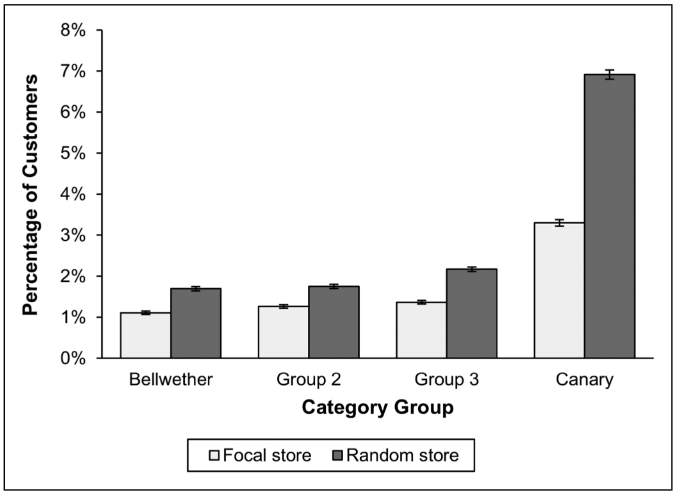

Figure 3. Probability That a Customer's Favorite Brand Is Unavailable (Focal Store vs. Random Store).

Notes: The figure reports the percentage of customers whose favorite brand is unavailable, comparing the assortment in the focal store with the assortment in a (different) randomly selected store. The unit of analysis is a customer × category × store, and the overall sample size is 980,878. Error bars indicate 95% confidence intervals. Detailed findings are reported in Web Appendix B.

and B). In Figure 2, Panel C, we show that customers in low survival stores are more likely to have favorite brands that are atypical, compared with customers in high survival stores. Among canary categories, there appears to be a strong association between niche preferences and low store survival rates.

Given this preference variation, we might expect the focal retailer to localize assortments among stores. If assortments were fully localized, we might expect that customers find their favorite brand and a canary category does not emerge. To investigate the extent of localized assortments, we compare the proportion of customers that are unable to find their favorite brands in a focal store (using the store's actual assortment) with the proportion that would not find their favorite brand if they visited a different randomly selected store.[19] The findings in Figure 3 illustrate the extent to which ConvenienceStore tailors its assortments to local preferences.

ConvenienceStore does appear to tailor its assortments to local preferences; customers' favorite brands are less likely to be unavailable in the focal store than in a randomly selected store. We also see that the difference between the actual store and random store unavailability rate is particularly large in canary categories. This suggests that in canary categories there is significantly more tailoring of local assortments to local preferences.[20] Yet, these assortment adjustments are incomplete. Customers are less likely to find their favorite brands in canary categories (Figure 1, Panel A) despite the retailer making large assortment adjustments in these categories.

Summary of Field Data Analyses

The field data establishes a clear association between missing favorite brands and customer behavior. First, we show that customers are more likely to substitute within a canary category (vs. a bellwether category) when a favorite brand is missing. Second, customers are less likely to return to a store when a favorite brand is missing. Third, customers in canary categories are more likely to prefer niche brands, and the retailer does not fully customize store assortments to these preferences.

Support for the Proposed Mechanism from Online Experiments

In this section, we experimentally investigate how the market features we identified in the theoretical model impact the canary category effect. Specifically, we experimentally vary market features and measure how this changes the tendency to purchase on the current trip but not return to that store on future trips. In each experiment, we measure the impact of our intervention on the sizes of Segments A, B, and C (as defined in the "An Explanation for the Canary Category Effect" section).[21]

Experimental Design

Recall from "The Canary Category Effect" section that the category with the lowest average survival rate is "juices single" (among categories with at least 50,000 preperiod purchases). We use this category to represent canary categories in most of our experiments. Similarly, we use the category with the highest average survival rate, "milk gal 2%," to represent bellwether categories in our experiments. To further connect our experiments with the historical data, we use the most popular juice and milk brands that we observe in the historical data to design the experimental stimuli. In particular, we use the juice and milk brands that were available in the largest number of stores.[22]

Although we design the experiments to closely mimic the field data and use juice to represent a canary category and milk to represent a bellwether category, we caution that these interpretations have to be treated with care. We have argued that the definition of a canary category is endogenous and

[19] Recall that each customer's favorite brand is identified using purchases outside the focal store. This means that we would not mechanically expect that a customer's favorite brand is more likely to be available in the focal store than in a randomly selected store.

[20] We caution that this does not imply that the assortments are tailored to individual stores. It is possible that the retailer merely adjusts assortments to groups of stores. However, this evidence confirms that the store assortments are at least somewhat matched to local preferences.

[21] Recall that if a customer's favorite brand is unavailable, Segment A customers do not purchase on the current trip and do not return on future trips, Segment B customers purchase an alternative brand on the current trip but do not return on future trips, and Segment C customers purchase an alternative brand on the current trip and return to the focal store on future trips. In the experiments, we include any customers that do not purchase on the current trip in Segment A, including those who may return to the store in the future.

[22] We exclude private label brands to avoid identifying the retailer. We also replace one juice brand because it no longer exists. We replace it with the highest-selling brand that is not already used in the experiment. We also consolidate "Dole juices" and "Dole" to "Dole."

could vary depending on both the store's actual assortment and the distribution of customer preferences. Therefore, while juice and milk are canary and bellwether categories, respectively, at ConvenienceStore, they may not be canary and bellwether categories at every retailer.

In our four experiments, we randomly vary four features of participants' decisions corresponding to the market features listed in Table 6. In Experiment 1, we vary whether a participant's preferred brand is present. In the second experiment, we randomly vary how urgently the product is required, which manipulates customer lock-in. In the third experiment, we vary the distance to an alternative store, and in the final experiment, we vary whether the brand is temporarily or permanently unavailable. We expect these different features to increase the size of Segment B. We also hypothesize that the effect on Segment B will be larger for canary categories than bellwether categories.

Experiment 1

In this experiment, we use a 2×2 experimental design, manipulating both the product category (juice vs. milk) and whether a participant's preferred brand is available or unavailable. The experimental procedure included three parts. In Part 1, we asked participants to identify their favorite brands in four categories (juice, smartphones, milk, and video streaming).

In Part 2, participants read the following scenario: "You **urgently need** [juice/milk] for an event that starts in one hour. You go to your local store and discover that it only sells the following brands."

To put participants in a lock-in mindset, the scenario describes an urgent need for this purchase (we vary urgency in Experiment 2). They were then shown a list of five brands, of which three were available in the store. The available brands were chosen according to the participants' choices in Part 1 and the randomly assigned treatment condition. To ensure that we only manipulated product assortments, without also manipulating participants' brand awareness, we showed participants the same list of brands in all treatment conditions and varied which brands were available. We then asked:

1. Question 1: Would you purchase [juice/milk] from this store on this shopping trip? Yes or no.
2. Question 2: Would you return to this local store for future shopping trips? Yes or no.

In Part 3, we asked participants to indicate their strength of preference for their first choice relative to their second choice out of the available assortment.[23] We also collected demographic measures (age and gender). Additional details, including the complete stimuli, are provided in Web Appendix D.

Participants were recruited for the experiment using Prolific. A total of 401 participants completed the study. The key outcome measure is how many participants were in Segment B: participants who indicated they would purchase on the current visit (responded "yes" to Question 1) but not buy again (responded "no" to Question 2). In Figure 4, we summarize the proportion of participants in each segment for each condition and product category.

In the juice (canary) category, when the preferred brand was unavailable, 20.8% (4.0%) of participants indicated that they would purchase in the category on the current shopping trip but would not return to that store the next time they need to purchase in the category (Segment B). These are the customers who contribute to the canary category effect. We also see a small segment of customers that did not buy at all (Segment A). However, when the favorite was available, the sizes of Segment A and B were essentially zero. These results are robust to adding age and gender controls (this is also true for our other experiments). For the milk (bellwether) category, when the preferred brand was unavailable, 13.9% (3.4%) of participants indicated that they would purchase in the category on the current shopping trip but would not return to that store the next time they need to purchase in the category (Segment B). To compare the two categories, we used OLS to estimate the following model:

$$\begin{aligned}\text{Segment B}_i = \alpha &+ \beta_1 \text{Unavailable}_i + \beta_2 \text{Canary Category}_i \\ &+ \beta_3 \text{Unavailable} \times \text{Canary Category}_i \\ &+ \text{Controls}_i + \varepsilon_i. \qquad (5)\end{aligned}$$

The dependent variable, Segment B_i, indicates whether participant i belonged to Segment B (responded "yes" to Question 1 and "no" to Question 2), Unavailable indicates whether they were in the unavailable or available condition, and Canary Category indicates whether they were asked about juice or milk. The coefficients of interest are β_1 and β_3: β_1 provides an estimate of the availability treatment effect for milk, while β_3 estimates the change (increase or decrease) in this effect for juice. The controls include the strength of preference variables as well as gender and age.

We find that β_1 is .092 (.039), suggesting that Segment B was 9.2% larger when the preferred brand was unavailable in the milk category. The estimate for β_3 is .096 (.058), suggesting that the effect size for the canary category is more than double the effect size for the bellwether category. The β_3 coefficient is marginally significantly larger than zero ($p < .10$). We provide the full regression table in Web Appendix D, where we also report findings for analogous models for Experiments 2 and 4.

For robustness, we separately replicated the results in four additional experiments using four canary product categories: juice, soda, chips, and candy. For each of these replications, we

[23] In the available condition, we showed participants the two (nonfavorite) brands and asked which they preferred. We then asked these participants to rate on a five-point scale (1 = "indifferent," and 5 = "very strongly prefer") their preference for their favorite brand compared with the preferred nonfavorite. In the unavailable condition, the favorite brand was unavailable, so we showed participants all three nonfavorite brands that were available and asked them to choose the most preferred brand then rate the strength of their preference. This data shows a positive correlation between preference and the likelihood of being in Segment B.

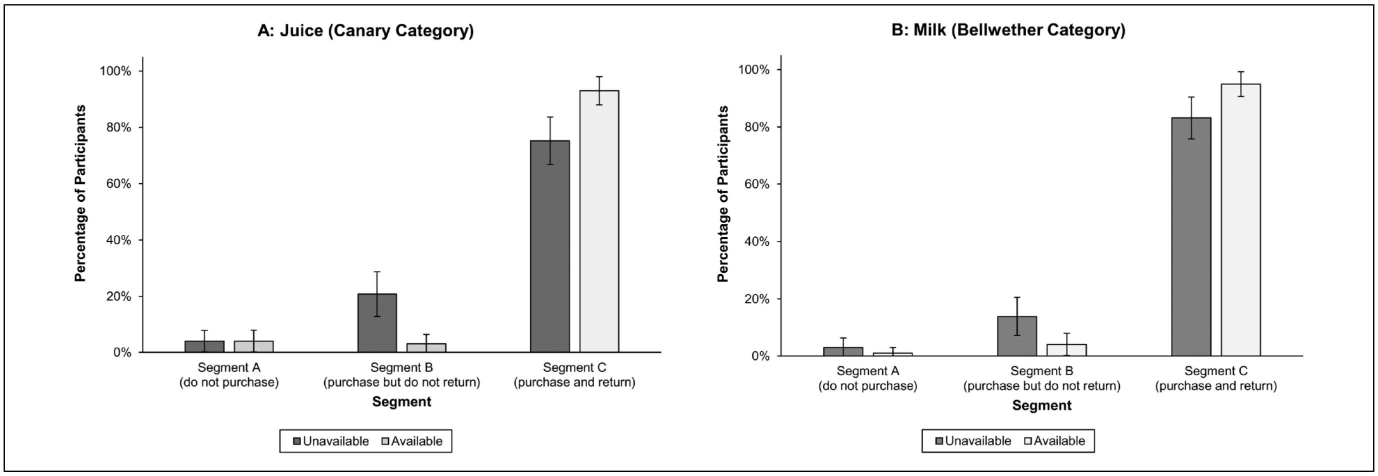

Figure 4. Results from Experiment 1.
Notes: The figure reports the percentage of participants in each segment in Experiment 1 by condition and category. Error bars indicate 95% confidence intervals.

recruited 200 participants using Prolific (800 participants in total). The findings are reported as Experiment 1B in Web Appendix D. They confirm that Segments A and B were both larger when the preferred brand was unavailable (compared with when it was available).[24]

Segment B participants in the unavailable condition, who would purchase on the current visit but not return to the store on future shopping trips, provide direct evidence of our proposed mechanism. Different customers have different preferences for canary categories. If customers visit a retailer and do not find their favorite brand, some of these customers will purchase on the current trip but not return in the future. In the next experiment, we investigate the extent to which the results depend on the urgency of customers' purchases.

Experiment 2

Recall that in the "An Explanation for the Canary Category Effect" section we interpret the urgency of the purchase as one reason that the value of the outside option may decrease when the customer is already at the focal store (W_B decreases relative to W_A). This increases lock-in on the current trip and can contribute to an increase in the size of Segment B. In this experiment, we measured whether varying the urgency of the purchase results in a reallocation of customers between Segments A and B. In particular, we repeated the unavailable conditions from Experiment 1. Within this setting, we randomly assigned participants to one of four experimental conditions (high vs. low urgency, juice vs. milk category), which varied the scenario participants were asked to consider:

High urgency condition: You are in your local store for a shopping trip. You **urgently need** [juice/milk]. You have an event that takes place in one hour and will need [juice/milk] for the event.

Low urgency condition: You are in your local store for a shopping trip. You **do not urgently need** [juice/milk]. You have an event that takes place in two weeks and will need [juice/milk] for the event.

The other elements of the experimental design were identical to Experiment 1. A total of 1,963 participants were recruited using CloudResearch, and the findings are summarized in Figure 5.

The findings in the high urgency conditions replicate the results of Experiment 1, with 20.4% (1.8%) of participants indicating that they would purchase on the current trip but not return to the focal store the next time they need juice (Segment B). In the low urgency condition, when we decreased the urgency of the purchase, the fraction of participants in Segment B decreased to 8.1% (1.2%). The difference between 8.1% and 20.4% is statistically significant ($p < .01$). This difference also occurs for the bellwether category in which Segment B increased from 6.5% in the low urgency condition to 17.6% in the high urgency condition. If there was no urgency, participants were significantly less likely to indicate they would purchase on the current trip, moving these participants from Segment B to Segment A. When there was urgency, the participants were a lot more likely to indicate they would purchase on the current trip, but many of them would not return to this store the next time they need juice. This is consistent with our interpretation that the urgency of the purchase can contribute to lock-in and increase the size of Segment B.

Next, we repeated our regression analysis by using OLS to estimate an analogous version of Equation 5. The findings are reported in Web Appendix D. They reveal that in the milk category Segment B was 11.2% (2.1%) larger when there was an urgency of purchase. This coefficient increased by 2.0% (3.0%) in the juice category (this difference is not statistically significant). The nonsignificance of this difference suggests that

[24] Segments A and B were essentially empty when the participants' favorite brands were available. The only exception is in the candy category; even when the preferred item was available, Segments A and B were statistically larger than zero (5% and 8%, respectively). One explanation could be that participants perceive that purchasing candy is less urgent than purchasing in other categories.

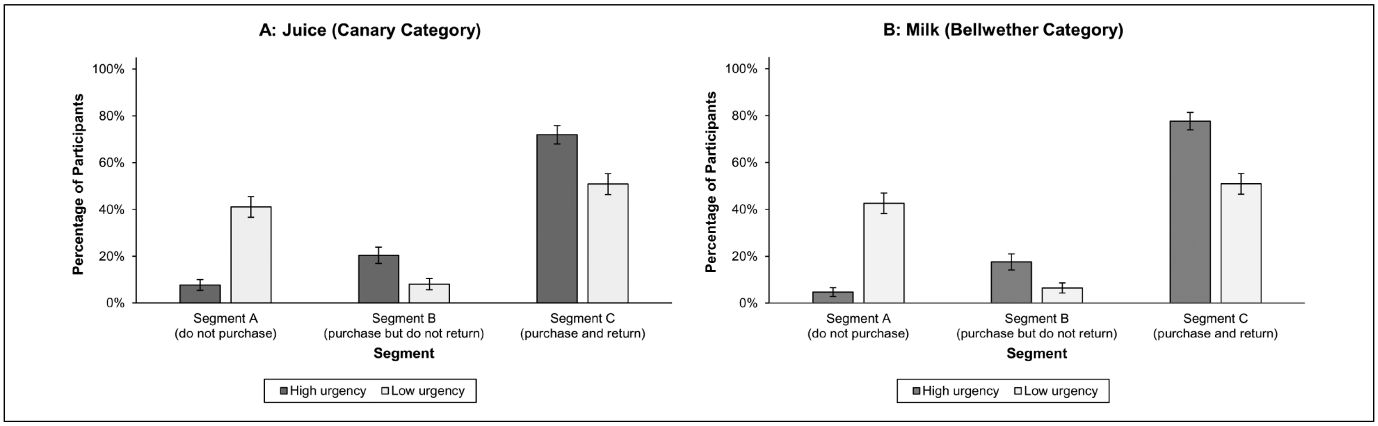

Figure 5. Results from Experiment 2.

Notes: The figure reports the percentage of participants in each segment in Experiment 2 by condition and category. Error bars indicate 95% confidence intervals.

there are no differences in the impact of urgency between canary and bellwether categories.

In Web Appendix D, we also report an experiment (2b) with a control condition where we did not indicate any urgency for the canary category condition. In that case, the size of Segment B was similar to the low urgency condition and was significantly lower than in the high urgency condition. Segment A was significantly larger when the scenario did not indicate any need for juice relative to the other two conditions (as we would expect).

Experiment 3

In this experiment, we manipulated the distance to an alternative store from the focal store. If customers have an alternative store close by, we might expect more customers to forgo purchasing altogether from the focal store. This represents a shift of customers from Segment B to Segment A. In particular, we again repeated the unavailable–juice condition from Experiment 1. Within this setting, we randomly assigned participants to one of two experimental conditions (near vs. far distance to an alternative store):

> Near condition: There are two local stores that are both the same distance from your house. The two stores are next to each other. You **urgently need** juice for an event that starts in one hour. You go to one of the local stores (see map).
>
> Far condition: There are two local stores that are both the same distance from your house. The two stores are on different sides of town. You **urgently need** juice for an event that starts in one hour. You go to one of the local stores (see map).

The map stimuli used in each condition are presented in Figure 6. We implemented two versions of the return intentions question. In Experiment 3a, we left the question as before and asked, "Would you return to this local store for future shopping trips? Yes/no." However, we were concerned that participants

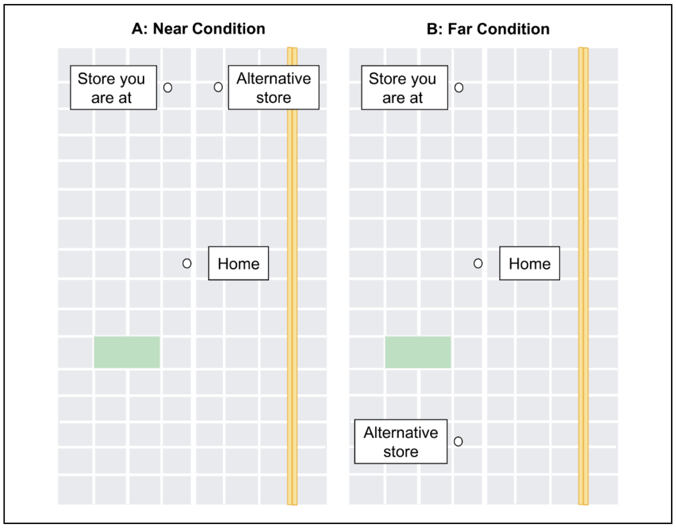

Figure 6. Map Stimuli Used in Experiment 3.

might be confused about which store we were referring to. Therefore, in Experiment 3b we asked, "Now, you are at home the following week and plan to go on a shopping trip. Which store would you go to for your shopping trip? Store I was at/alternative store." Then, we again showed the map and, for clarity, adjusted the map labels to "Store you were at" and "You are here: Home." We also decided to focus only on the canary category in this experiment. Other elements of the experiment were identical to Experiment 1.

A total of 200 participants (who had not participated in any of our other experiments) were recruited using Prolific for each of the two experiments, and the findings are summarized in Figure 7. While the results are directionally consistent with our predictions (Segment B was larger in the far condition than in the near condition), the sizes of Segment B were not significantly different between conditions in either experiment. However, the size of Segment A was significantly larger in the near condition,

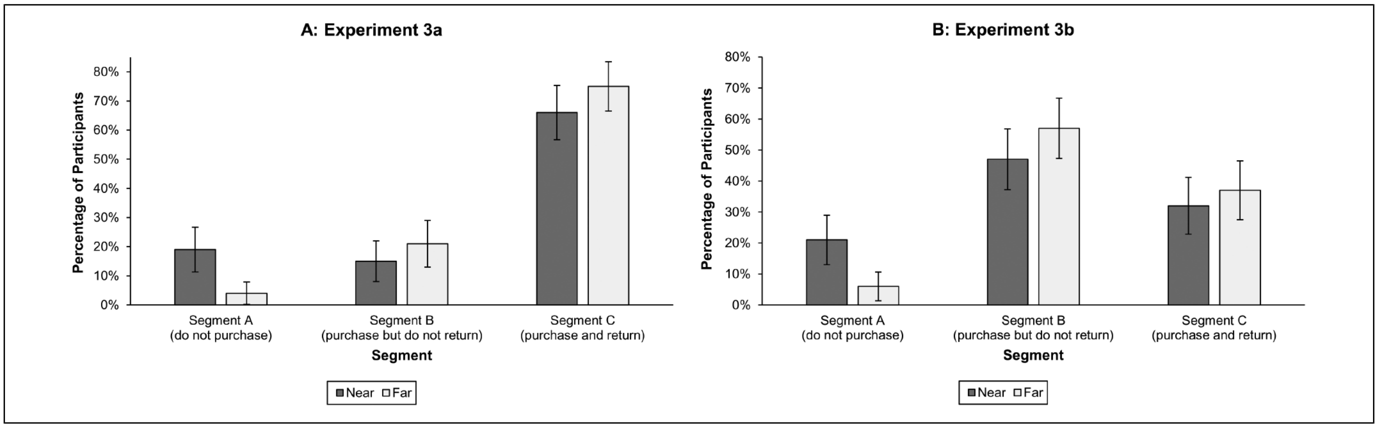

Figure 7. Results from Experiment 3.
Notes: The figure reports the percentage of participants in each segment in Experiments 3a and 3b by condition and category. Error bars indicate 95% confidence intervals.

suggesting that when there is an alternative store nearby, people are more likely to forgo purchasing at the focal store.

Together, these findings are directionally consistent with our proposed mechanism but not statistically significant. We believe the effect of distance to the alternative store is likely to be more subtle than the other market features we considered in our experiments.

Experiment 4

In our final experiment, we varied customers' expectations about future product availability. We hypothesize that when customers expect that their preferred brand is only temporarily unavailable, they are more likely to return to the store in the future, thereby reducing the size of Segment B (relative to if they expect the stockout to be permanent). We again repeated the unavailable conditions from Experiment 1. Within this setting, we randomly assigned participants to one of four experimental conditions (temporary vs. permanent stockout, juice vs. milk), which varied the scenario participants were asked to consider:

> Permanent condition: You **urgently need** [juice/milk] for an event that starts in one hour. You go to your local store and discover that it only sells the following brands (the brands available in this store are highlighted in green): … You learn from the store manager that the store will no longer carry the unavailable brands.

> Temporary condition: You **urgently need** [juice/milk] for an event that starts in one hour. You go to your local store and discover that it only sells the following brands (the brands available in this store are highlighted in green): … You learn from the store manager that the unavailable brands will be restocked later this week.

The other elements of the experimental design were identical to Experiment 1. A total of 2,000 participants (who had not participated in any of our previous experiments) were recruited using Prolific, and the findings are summarized in Figure 8.[25] The results are consistent with our expectations; the canary category effect is more likely to occur when customers infer that the unavailability of their preferred brand is permanent. In the permanent–juice condition, 23.0% (1.9%) of participants were in Segment B, compared with just 10.8% (1.4%) in the temporary condition. The difference in these proportions is significant ($p < .01$). In the milk category, 15.3% (1.6%) of participants were in Segment B in the permanent condition, compared with 8.4% (1.2%) in the temporary condition. The difference in these proportions is significant ($p < .01$).

We again used Equation 5 to estimate the difference in the treatment effects between the two product categories. The results are reported in Web Appendix D. They reveal that in the milk category Segment B was 6.6% (2.0%) larger when the brand was permanently unavailable. This treatment effect increased by 5.5% (3.0%) in the juice category (this difference is marginally significant [$p < .10$]). This indicates that the effect size is almost twice as large in the canary category (juice) than in the bellwether category (milk).

In Web Appendix D, we also report an experiment (4b) with a control condition where we did not indicate any expectations of availability for the canary category condition. The size of Segment B was similar to the permanent condition and was significantly higher than in the temporary condition. It appears that when we did not describe future availability, participants interpreted the brand's unavailability as permanent.

Summary

In this section, we reported findings from four online experiments. The experiments were closely tied to the historical data; we used product categories and brand assortments that occur in the historical data. Together, the experiments provide causal support for our explanation of the canary

[25] We used data from the 1,997 participants who completed the survey.

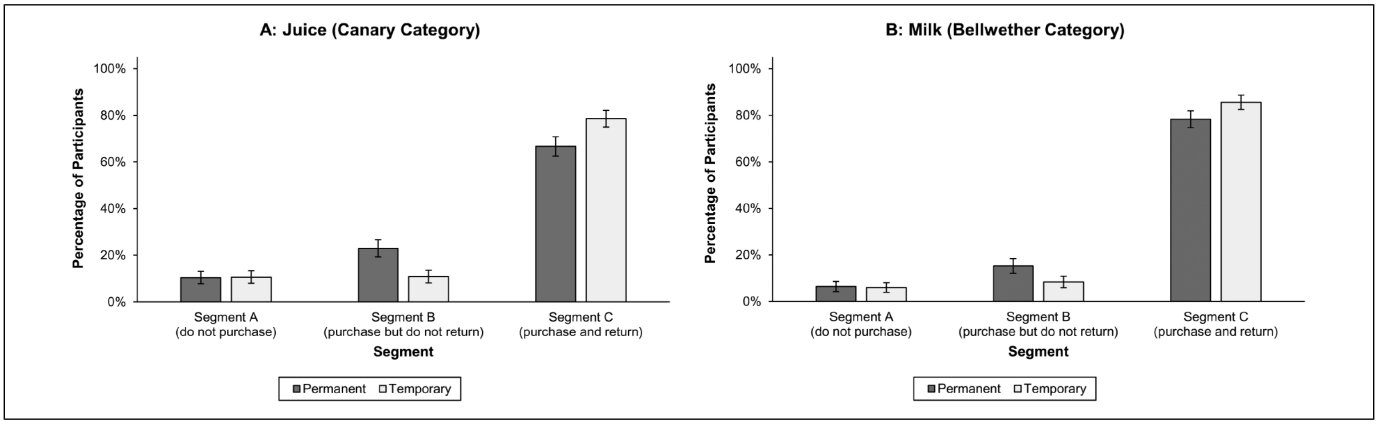

Figure 8. Results from Experiment 4.
Notes: The figure reports the percentage of participants in each segment in Experiment 4 by condition, by category. Error bars indicate 95% confidence intervals.

category effect. Different customers prefer different brands. If a store does not sell their preferred brands, some customers facing lock-in will purchase another brand, but many will not return to that store the next time they need to purchase in that category. The findings indicate that three market features contribute to the effect: (1) the likelihood customers will find their preferred brand, (2) customers' expectations about future product availability, and (3) the urgency of the purchase. The first two of these effects are larger in our representative canary category (juice) than in the representative bellwether category (milk).

Conclusion

We document a novel empirical finding that we label the canary category effect. When customers purchase products in canary categories, it is a signal that they will not return to that store on future trips. This contradicts conventional wisdom that past purchasing is a positive indicator of future purchasing. The effect is robust and replicates at different retailers.

While multiple mechanisms may contribute to the canary category effect, we focus on a mechanism that is well suited to the retailer in our study. When customers visit a store and their preferred brand is not available, some customers buy but are less likely to return in the future. Under this explanation, the trigger for the effect is the absence of a customer's favorite brand from the retailer's assortment. The implication is that canary categories are endogenous and can vary from retailer to retailer depending on their assortments. A canary category at one retailer may be a destination category at another retailer.

This explanation does not suggest that retailers should stop selling canary categories. These categories provide an important source of revenue and may also contribute an important source of differentiation that attracts customers. Instead, canary categories serve an important role, helping managers identify categories in which incomplete assortments are contributing to customer attrition. In the same way that canaries in coal mines warn miners of the danger of carbon monoxide, canary categories warn retailers that there are gaps in their assortments that are contributing to customer attrition.

Our findings also have implications for how retailers predict customer response when discontinuing a brand. Although temporary stockouts appear to provide a natural experiment to gauge how customers will respond, the findings in Experiment 4 highlight an important difference between temporary stockouts and permanently discontinuing a brand. We show that customers' willingness to return to a store depends on whether their favorite brand is temporarily or permanently unavailable. As a result, using temporary stockouts to predict the response to discontinuing a brand is likely to result in underpredicting customer attrition, and this prediction error is expected to be larger in canary categories.

Our findings suggest several additional opportunities for future research. We show that the canary category effect appears to be at least partly attributable to frictions that prevent retailers from fully adjusting their local assortments to local preferences. An obvious friction is limited shelf space. If there were no shelf space restrictions, the retailer could offer every brand in every category. Other frictions that could limit retailers' abilities to adjust local assortments might include supplier requirements, logistical challenges of adjusting assortments, information challenges of knowing which assortments to offer, and managerial challenges of identifying and implementing local adjustments in store assortments. There is currently little research investigating how these frictions influence the adjustment of assortments to local preferences.

The canary category effect may also offer new opportunities to enrich and extend different types of customer demand models. For example, customer lifetime value models typically assume a positive relationship between transactions and future value. For canary categories, we show that the effect can be negative. The canary category effect may also provide an important input to buy-till-you-die, which can serve an important role in estimating customer valuations and guiding resource allocations. There is also a well-established stream of research estimating customers' latent preferences. By estimating latent brand preferences, it may be possible to use the canary category effect to model how these latent preferences combine with variation in store assortments to affect customer store choices.

The extent to which the canary category effect is accounted for in existing assortment optimization models is unclear. These models typically do not consider store switching or individual customer behavior. However, it is possible that they indirectly account for at least part of the effect. For example, Rooderkerk, Van Heerde, and Bijmolt (2013) model a cannibalization effect based on product availability, which could capture part of the canary category effect. Further research could investigate the extent to which the effect is already incorporated into existing models.

A limitation of our analysis is that we do not have causal evidence in our field data of how customers will react to assortment changes, and this prevents us from identifying optimal assortments. While we document how purchasing varies with changes in assortments across stores and over time, these assortment changes are endogenous. It is for this reason that we complement our field data with online experiments. Future research using experiments conducted in the field would be valuable to further investigate the impact of the canary category effect on optimal store assortments.

Coeditor

Raghuram Iyengar

Associate Editor

Paulo Albuquerque

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Ayelet Israeli https://orcid.org/0000-0002-5996-3202
Duncan Simester https://orcid.org/0000-0003-2758-0116

References

Anderson, Eric, Gavan J. Fitzsimons, and Duncan Simester (2006), "Measuring and Mitigating the Costs of Stockouts," *Management Science*, 52 (11), 1751–63.

Anderson, Eric, Song Lin, Duncan Simester, and Catherine Tucker (2015), "Harbingers of Failure," *Journal of Marketing Research*, 52 (5), 580–92.

Boatwright, Peter and Joseph C. Nunes (2001), "Reducing Assortment: An Attribute-Based Approach," *Journal of Marketing*, 65 (3), 50–63.

Boatwright, Peter and Joseph C. Nunes (2004), "Correction Note for 'Reducing Assortment: An Attribute-Based Approach,'" *Journal of Marketing*, 68 (3), iv.

Borle, Sharad, Peter Boatwright, Joseph B. Kadane, Joseph C. Nunes, and Shmueli Galit (2005), "The Effect of Product Assortment Changes on Customer Retention," *Marketing Science*, 24 (4), 616–22.

Briesch, Richard A., Pradeep K. Chintagunta, and Edward J. Fox (2009), "How Does Assortment Affect Grocery Store Choice?" *Journal of Marketing Research*, 46 (2), 176–89.

Briesch, Richard A., William R. Dillon, and Edward J. Fox (2013), "Category Positioning and Store Choice: The Role of Destination Categories," *Marketing Science*, 32 (3), 488–509.

Broniarczyk, Susan M., Wayne D. Hoyer, and Leigh McAlister (1998), "Consumers' Perceptions of the Assortment Offered in a Grocery Category: The Impact of Item Reduction," *Journal of Marketing Research*, 35 (2), 166–76.

Chernev, Alex and Ryan Hamilton (2009), "Assortment Size and Option Attractiveness in Consumer Choice Among Retailers," *Journal of Marketing Research*, 46 (3), 410–20.

Drèze, Xavier, Stephen J. Hoch, and Mary E. Purk (1994), "Shelf Management and Space Elasticity," *Journal of Retailing*, 70 (4), 301–26.

Fisher, Marshall and Ramnath Vaidyanathan (2012), "Which Products Should You Stock?" *Harvard Business Review* (November), https://hbr.org/2012/11/which-products-should-you-stock.

Food Marketing Institute (1993), "Variety or Duplication: A Process to Know Where You Stand," report, Willard Bishop Consulting and Information Resources Inc.

Rooderkerk, Robert P., Harald J. van Heerde, and Tammo H.A. Bijmolt (2013), "Optimizing Retail Assortments," *Marketing Science*, 32 (5), 699–715.