American Economic Journal: Microeconomics 2012, 4(3): 96–120
http://dx.doi.org/10.1257/mic.4.3.96

Shopping Cost and Brand Exploration in Online Grocery[†]

By Andrea Pozzi*

This paper compares consumers' brand exploration when shopping online versus in a brick-and-mortar store. I use a new scanner dataset to compare the behavior of households shopping online and in-store at the same chain, for identical items and prices. I find that brand exploration is more prevalent in-store. My model quantifies the role of features of e-commerce, like the existence of "favorites lists" and the difficulty in verifying item quality. Limited exploration online implies higher barriers to entry on the Internet channel. Counterfactual exercises suggest that online advertising could make the Internet channel more competitive. (*JEL* D12, L11, L81, M31, M37)

Brand exploration—that is, purchasing a brand not tried in the past—is an important phenomenon in retail markets. Learning how much consumers can be induced to explore new products is key for manufacturers because it informs their pricing (Klemperer 1995) and marketing strategies (Bayus 1992). It also has policy implications, given that markets where brand exploration is low have higher de facto barriers to entry (Schmalensee 1974). In this paper, I study how the choice of the shopping channel (online or brick-and-mortar store) affects buyers' propensity to explore new brands. I show that customers tend to try new products less often when they shop online. I then estimate a structural model that quantifies this result. My findings imply that firms can exploit their customer base on the online channel, and that they also can minimize the threat of successful entry from outsiders.

Typically, the Internet is characterized as reducing the cost of search. Such online tools as shopbots and price search engines make it easy to gather information on alternative products and to compare them (Bakos 1997; Brynjolfsson and Smith 2000; Clemons, Hann, and Hitt 2002). More search may mean a higher probability of trying a new item. Therefore, we should expect e-commerce to foster exploration and competition (Brown and Goolsbee 2002).

On the other hand, it is recognized that Internet and brick-and-mortar retailing are different beyond their search costs. For example, the online shopper cannot

* Einaudi Institute for Economics and Finance, via Sallustiana 62, 00187 Rome, Italy (e-mail: andrea.pozzi@eief.it). I am deeply indebted to Liran Einav, Tim Bresnahan, and Jakub Kastl for their advice and guidance. I also benefited from comments and suggestions by Paul David, Wesley Hartmann, Jon Levin, Sridhar Moorthy, Andres Musalem, Harikesh Nair, Sridhar Narayanan, Franco Peracchi, Fabiano Schivardi, Frank Wolak, two anonymous referees and participants in the IIOC 2009, the seventh ZEW Conference on the Economics of ICT, the EARIE '09 and in seminars at several universities. I thank the managers of the supermarket chain for their help in accessing the data and for sharing with me their knowledge of the industry and of their business practices. I gratefully acknowledge financial support from SIEPR, in the form of the Schultz Graduate Student Fellowship and of the B.F. Haley and E.S. Shaw dissertation fellowship, and from the NET Institute. All errors are mine.
† To comment on this article in the online discussion forum, or to view additional materials, visit the article page at http://dx.doi.org/10.1257/mic.4.3.96

physically inspect the product, which makes it harder to verify the quality of online purchases. Furthermore, the features of e-shopping Web sites affect consumers' choices in ways that cannot be replicated in a store. E-retailers provide recommendations based on customers profiling, or they may create personalized lists of favorite goods. These features can affect product exploration and the competitive environment more generally. Although the outcome will depend on the Web site design, these practices are becoming ubiquitous, and firms frequently adopt them.[1] Still, we do not know much about the impact of these features on consumers' choice.

Identifying the effect of online shopping on consumer behavior is challenging because it involves many confounding factors. First, the choice of becoming an online shopper is endogenous. To disentangle the effect of the channel from self-selection, we must ask how an Internet customer would behave if shopping instead in a brick-and-mortar store. Additional complications arise from the different pricing policies across the two channels (goods are typically cheaper online). On the supply side, there are also more new entrants with little or no brand name online.

I explore a growing online shopping market: groceries (Goettler and Clay 2011). I analyze new scanner data for two years of grocery purchases made by a panel of some 11,000 households who shop *both* online and in-store at the same supermarket chain. The store purchases are tracked through use of a loyalty card. The card information is also required when the customer registers for the online service on the grocer's Web site, which allows me to link in-store and online trips for the same household.

This is an attractive setting for isolating the effect of online shopping. I can observe the same household shopping on both channels, whereas most previous studies compared a sample of online shoppers with a different sample of traditional shoppers.[2] My data then reduces concerns related to sample selection. Moreover, both the in-store and online purchases occur at the same supermarket chain. This ensures that the set of brands carried, prices and promotions, and loyalty to the retail chain are identical in both channels.

The focus of my analysis is the breakfast cereals product category. This category is highly concentrated (Nevo 2001) and presents a classic example of niche filling strategy (Schmalensee 1978; Scherer 1979). Each manufacturer markets a large number of differentiated brands. As a consequence, there is great potential for brand exploration. I document that, in my sample, households are 8 percentage points more likely to try a cereal brand they have not purchased before when they shop in a brick-and-mortar store than when they are shopping online. This finding is robust to different specifications and is consistent with previous results.[3] My goal is to understand the causal mechanisms behind this finding. I consider three potential explanations of the pattern seen in the data.

[1] In 2006, Netflix, an online video rental company, offered a $1M prize to those who could achieve a substantial improvement on their algorithm for rental recommendations. Contestants were granted access to proprietary data on previous customer rentals.

[2] Exceptions are Chu, Chintagunta, and Cebollada (2008) and Brynjolfsson, Hu, and Simester (2007) who also use a panel of cross-channel purchases, but with a different focus.

[3] Degeratu, Rangaswamy, and Wu (2000) use data from Peapod and observe that brand switching is lower in Peapod orders than in brick and mortar stores. Danaher, Wilson, and Davis (2003), with data from a grocery chain that also offers online delivery, find that brand loyalty is higher when customers choose to purchase on the Internet.

First, households may sort their transactions. When everybody shops both online and offline, selection is less of a concern. However, the decision of when to shop online is probably endogenous: purchasing groceries online is time efficient, so customers might be more likely to choose the online channel when they are short of time. In that case, they may also be less likely to explore new brands.

Second, certain features of the Internet environment, including list of favorite or recommended items, may reduce the cost of shopping and make the online shopping experience less flexible. In my analysis, the grocer's Web site offers the option of saving time browsing by purchasing from a "one click" list of items already selected by customers in the past. This time-cost reduction comes at the expenses of the freedom to browse, and potentially to uncover new brands. In fact, by design, this option reduces the amount of online exploration. A *past shopping history* list may have an effect given the evidence that the cost of browsing a Web page, or typing, can be substantial (Hann and Terwiesch 2003). For online goods that are seldom purchased and expensive, such as airplane tickets or electronics, this option is unlikely to be relevant. But in the case of a repeated activity involving relatively cheap items, like groceries, the appeal of a reduction in shopping costs can be much greater.

Finally, quality verification is more difficult online than in-store (Bhatnagar, Misra, and Rao 2000; Jin and Kato 2007). The information content of online searches is not as good as in the in-store setting. Exploration of new brands is by definition the search for items with more desirable attributes than those currently in the consumer's basket. If assessing the attributes of new brands is more difficult online, then the probability of successful exploration is lower for online purchases. As a result, a consumer will be less inclined to buy new brands when she shops on the Internet.

I quantify the importance of these three effects by estimating a structural model of consumer behavior. In the model, consumers first select the channel where they want to shop. Conditional on that choice, they pick a cereal brand. My identification of the channel selection equation exploits the supermarket chain's policy of offering discounted delivery fees for Internet orders, which introduces exogenous variation in the cost of shopping online. My results indicate that the sorting element definitely plays a role. The residuals from the channel selection equation are negatively correlated with unobserved shocks to the taste for unknown brands. In other words, specific circumstances that make it more likely for an agent to go online (e.g., lack of time) similarly make her less likely to try new brands. The reduction in shopping cost from a past shopping history list is also an important factor in explaining why consumers avoid exploring new brands on the Internet. My estimates imply that the benefit of not having to browse the Web site—by resorting to the past shopping history list—is worth about \$4 to the typical consumer.

Some counterfactual exercises provide additional insights on the economic significance of my results. For example, altering the design of the grocer's Web site to remove the past shopping history list boosts online brand exploration by 23 percent. I also show that new entrants' penetration rate is slower online. Within 18 months, a new brand achieves 60 percent of its potential in-store market share but only reaches 40 percent of its target market share online. This finding indicates additional entry barriers that online technology creates in this environment. At the same time, the introduction of features that facilitate the exploration of new brands online (for example, context ads or recommendations), results in the entrant penetrating the market much more quickly on the Internet.

My research is tied to the literature on the impact of popular features of e-commerce Web sites, including Ellison and Ellison's (2009) study of obfuscation techniques and Ackerberg, Hirano, and Shahriar's (2006) analysis of the Buy-it-now option in eBay auctions. This paper is also related to work that tests the "frictionless commerce" paradigm (Brynjolfsson and Smith 2000; Chen and Hitt 2002; Brynjolfsson, Dick, and Smith 2010).

The remainder of the paper is organized as follows. Section I presents the data and provides institutional details on the online grocery business. In Section II, I describe the cross-channel rate of trial of new brands. Section III develops a structural model of channel choice and demand for brands. I explain the estimation strategy in Section IV. Section V reports the results; Section VI presents the counterfactuals. Section VII concludes.

I. Data and Institutional Background

My data come from a national supermarket chain that operates more than 1,500 stores across the United States. This chain also offers the option of shopping for groceries online, through the company Web site, and having them delivered to your home. I observe scanner-level data for each purchase made by the 11,640 households who shopped at least once in a supermarket store and at least once through the online service between June 2004 and June 2006. For each one of those 1,829,254 shopping trips, I observe the date, the list of all items purchased (as identified by their bar code or UPC), price, and quantity purchased. Coupons and discounts are also recorded in the data. Finally, I have information on whether the purchase took place in a brick-and-mortar store or online. Customers are identified through their loyalty card number;[4] the chain matches cards belonging to different members of the same household under the same identifier.

A. *The Online Service*

The online service was first offered in 1999 but was substantially re-organized in 2002. The option of buying online is only available in selected metropolitan areas. Nevertheless, in areas reached by the service, the online channel plays a significant role. In my sample, Internet trips account for 9 percent of the total and generate about 25 percent of the revenues. Online orders are fulfilled using stock available in regular stores for each area covered by the service. Thus, the stockout process and

[4]Purchases made by households without a loyalty card are not recorded in the data. This is not a big concern, because the chain encourages use of the loyalty card (that entitles buyers to discounts and special offers on many items) and estimates that more than 85 percent of the customers have one. As for use of the loyalty card, the figures are also very high. Einav, Leibtag, and Nevo (2010) analyze shopping trips reported by Homescan households for a particular retailer who tracks customers through loyalty cards. Less than 20 percent of the shopping trips recorded by a household in Homescan do not find a match in the retailer's data. This is an upper bound, since the difference can be explained by trips where the card was not used but also by mistakes made by the consumer while recording the trip. For example, she can erroneously report the identifier of the store she visited, or the date of the trip.

the variety offered should not have any systematic impact on differences between the two channels.[5] The online service runs seven days a week, customers pay a delivery fee; and they can pick the delivery time, conditional on availability.

To access the online service, customers must register with their address, phone number, and loyalty card number. That allows me to link a customer's online transactions with in-store purchases. Online, the customer can choose how to shop: she can browse the online store, where aisles are represented as a series of nested links, and consider items displayed in alphabetical order or by price. Alternatively, she can shop from the list of items bought during her last visit or in her entire shopping history. The items' descriptions include characteristics, price and presence of discounts, a photograph of the item, and nutrition information. The retailer commits to offering goods at the same price online and in-store; therefore, customers face identical prices in the two environments.[6]

B. *Trip Characteristics: Online versus In-Store*

I focus on the breakfast cereal category. Because of the large number of existing brands, breakfast cereals represent an excellent category for analyzing product choice and brand switching decisions by households. Furthermore, they are popular and frequently purchased, and feature prominently in both online and in in-store sales. In my data, over 140,000 trips include the purchase of one item from this category. Finally, this product choice eases the bias from relying on data from a single retailer, because this grocery chain is a major player in this product category[7].

In the data, one observation refers to purchase of a good identified by its Universal Product Classification (UPC) code. For example, a UPC description might be "Kellogg's Cocoa Krispies, 18 oz. box." My definition of brand requires aggregating over multiple UPCs; therefore, UPCs referring to Rice Krispies sold in different sizes (18 ounce versus 12.5 ounce boxes) and varieties (Frosted versus Berries) identify the same brand in my analysis.

Table 1 reports descriptive statistics for trips involving cereal purchases, by distribution channels. The average size of the basket and the net expenditure are higher online, reflecting both the $50 minimum order constraint to access the online service and the incentive to stock-up in online orders, because of the fixed cost of delivery. Once we focus on cereal purchasing behavior, though, the differences are less marked. The number of cereal boxes purchased per trip is just higher when the customer orders online; the difference in the number of distinct cereal brands purchased in an online versus offline trip is even smaller. Both online and in-store trips appear to be comparable in terms of the most popular product categories purchased in each channel. Table 2 shows that nine out of the top ten categories purchased are the same.

[5] Online customers can specify instructions to be followed in case one item in their shopping list is not available. The options offered are: "no substitution"; "same size, different brand"; and "same brand, different size."

[6] Note that this does not imply that prices are the same in all stores. The retailer's price strategy is based on location, and the online customer is offered prices that match those in the area of her IP address.

[7] Based on a sample of Homescan data for 2004, the chain is among the top three retailers, both for market share and for the number of trips, in the breakfast cereal category in the United States. If we restrict the analysis to markets where the chain is actually operating, it becomes by far the largest retailer in the product category.

Table 1—Trip Descriptive Statistics by Channel of Purchase, for All the Trips Involving Purchase of Breakfast Cereals

	In-store	Online
Total expenditure	110.9	179.2
	(78.6)	(89.5)
Basket size	40.5	66.3
	(27.8)	(35.1)
Number of unique items	30.6	41.6
	(20)	(19)
Number of unique categories	23.3	30.4
	(14.4)	(12.7)
Number of cereal boxes	1.5	1.9
	(0.9)	(1.5)
Number of unique cereal brands	1.3	1.4
	(0.56)	(.80)
Observations	106,378	35,647

Notes: Total expenditure is net of any discount and expressed in dollars. Basket size refers to the total number of items purchased in the trip, including multiple purchases of the same good. Standard deviations are reported in parentheses.

Table 2—Product Categories Most Frequently Purchased in In-store and Online Trips

In-store	Online
Milk and substitutes	Milk and substitutes
Carbonated soft drinks	Fresh bread
Fresh bread	Bananas
Bananas	Carbonated soft drinks
Refrigerated yogurt	Salad vegetables
Salad vegetables	**Cold cereals**
Cold cereals	Eggs and substitutes
Cooking vegetables	Refrigerated yogurt
Still water	Still water

Note: The ranking is computed by counting the number of trips in which *at least one item* of a given product category has been purchased.

C. *Pricing of the Online Service*

The supermarket offers the same prices for goods sold online and in-store. However, Internet shoppers must also pay a delivery fee. Variation in the charge for delivery will be an important source of identification in the model; therefore, I describe the main features of the pricing of the online service here.

In the period under examination, the full price for delivery was \$9.95. In an attempt to promote the relatively new and still expanding online service the grocer issued frequent coupons, sent by mail and entitling shoppers to a discount on the delivery fee or waiving it altogether. Instead of targeting specific customers for the discount, the coupons were mailed to all registered shoppers in a given area, roughly corresponding to a zip code. Such a policy has two important implications for this analysis. First, it makes ownership of a coupon exogenous from the point of view of an individual household. Second, it allows me to back out coupon

Table 3—Distribution of Internet Delivery Fee

		Mean	Variance	Free	$4.95	$7.95	$9.95
				(*percent*)			
All trips							
	All discounts	7.14	13.99	17.38	14.89	11.87	52.81
	Only coupons	7.27	14.41	18.24	10.09	12.50	55.92
Cereal trips							
	All discounts	6.69	12.69	15.57	28.35	10.22	43.24
	Only coupons	7.14	14.71	19.00	11.09	12.59	53.98

Notes: The figures refer to the potential delivery fee; that is the one the household would have paid if it had decided to shop online. The mean fee is expressed in dollars. The *All discounts* rows include fee reductions granted to large orders. The *Only coupons* rows only consider discounts originated by mailed coupons.

holding even for customers who did not use one. In fact, whenever I observe a household using a coupon in a given week, I can assume that all the other households living in the same area had received one. For each in-store trip, I can impute a "potential fee"—that is, the fee the household would have paid had it decided to make that order online.

Table 3 presents summary statistics for the potential fee for all grocery trips in the sample and for trips involving cereal purchase only. The most common fees paid are $9.95, $7.95, and $4.95; free deliveries are also fairly common. In general, the couponing strategy generates a substantial amount of variation in the cost of the online service. Only half of the trips are charged the full delivery price, even fewer if we focus only on cereal trips. More importantly, there is significant within-household variation. Customers see the potential price of Internet orders vary over time, according to whether their area was targeted for coupon distribution. Some of the variation in the delivery fee could still be endogenous because shoppers also obtain discounts in case of large orders. However, knowing the rules that trigger such discounts, I can isolate them from the truly exogenous variation coming from coupons. In Table 3, I also present the distribution of potential fee when size discounts are excluded from the sample.

Coupons aim at promoting use of the new online service to create critical mass. Over 75 percent of the zip codes reached by the Internet service are targeted for a delivery fee discount at least once. Table 4 column 1 reports correlations between the potential delivery fee and the characteristics of the market. In column 2, a similar exercise is repeated for the probability of issuing coupons in the zip code. The most significant results are a slight preference for areas where younger households are prevalent and for markets characterized by higher income customers. Both groups are more likely to become regular e-shoppers. My analysis of the month fixed effects (not reported in the Table) reveals that discount campaigns are more likely to occur between March and June.

The last two columns of Table 4 indicate that these promotions are effective. I fit a linear probability model to detect the correlation between households' choice of shopping channel for a particular trip and the potential fee (column 3) or a coupon-holding indicator (column 4). The coefficient on the potential fee variable implies that moving from full price to free delivery increases the probability of selecting the

Table 4—Correlation of the Delivery Fee with Zip Code Characteristics and with Choice of the Shopping Channel

Dependent variable	Fee (1)	Coupon dummy (2)	Online dummy (3)	Online dummy (4)
Black	0.014	−0.002		
	(0.0012)	(0.0002)		
Hispanic	−0.004	0.000		
	(0.0011)	(0.0000)		
Age 35–54	−0.041	0.007		
	(0.0037)	(0.0005)		
Age 55–64	0.017	−0.002		
	(0.0045)	(0.0006)		
Age $\geq$ 65	−0.014	0.004		
	(0.0020)	(0.0003)		
Income	−0.013	0.001		
	(0.0010)	(0.0001)		
Fee			−0.007	
			(0.0010)	
Coupon				0.015
				(0.0040)
Fixed effects				
Household	No	No	Yes	Yes
Month	Yes	Yes	No	No
Day of the week	No	No	Yes	Yes
N	64,771	64,771	141,396	141,396
R^2	0.02	0.02	0.03	0.02

Notes: All coefficients are significant at the 1 percent level. The unit of observation is a zip code-week in columns 1 and 2 and a trip in columns 3 and 4. Standard errors are reported in parentheses.

online channel by almost 25 percent. Similarly, conditional on holding a coupon for discounted delivery, an online shopping trip is 6 percent more likely.

D. *Demographic Information*

For a random subsample of households in my data, the grocer has provided information on the address, edited to prevent identification of the specific household. I have this information for 6,155 of the households who purchased breakfast cereals at least once. I match those households with demographic data from the Census 2000 at the block-group level[8]. The demographic data include the share of African-American and Hispanic people in the block, the share of families, the fraction of the population with a college degree, the fraction employed, the age of the

[8] Matching at the block-group level, rather than the five-digit zip code level, has two main advantages. Block groups are smaller: their boundaries never cross county or state limits (as opposed to census tract boundaries), and are designed to include relatively homogeneous population. Hence, I am not only averaging demographic characteristics over a smaller set of people, but also over a set of people that is more likely to be similar.

Table 5—Demographic Information

	Percentiles				
	5th	25th	50th	75th	95th
Number of trips	48	102	164	250	451
Number of cereal trips	4	12	23	38	70
Trips online (percent)	0.4	1.7	5.3	16.2	50.6
Cereal trips online (percent)	0	2.4	12.5	38.5	86.4
Black	0	0.2	1.7	4.5	19.7
Hispanic	0.4	3.8	7.4	14.4	36.6
College degree	18.8	35	49.6	63.2	78.7
Employed	49.7	60.4	66.7	71.9	79.2
Age 15-34	0	4.6	9.7	17	31.6
Age 35-54	27.1	44.5	53.6	62.5	75.1
Age 55-64	0	10.4	16.1	22.1	32.9
Age $\geq$ 65	0	8.2	15.8	25	42.2
Commute $\leq$ 30 minutes	36.18	49.3	57.3	66	79.3
Commute 30 to 59 minutes	14.7	25.1	32.2	38.6	49.3
Commute $>$ 60 minutes	1.5	5	8.9	14.3	23.7
Per capita income	16,261	23,837	30,791	41,884	69,045
Distance	0.26	0.65	1.09	1.72	3.38

Notes: Information on number of trips, number of cereal trips and number of online trips are constructed from the scan data provided from the grocer. Distance is calculated as miles between the domicile of the household and the closest brick-and-mortar store of the supermarket chain and was also provided by the retailer. All the other variables (*black, Hispanic,* c*ollege degree, employed, per capita income, age, and commuting time*) are shares for the 9-digit zip code of residency of the household, with the exception of per capita income, which is expressed in dollars.

head of the household, and average income per capita. Table 5 provides an overview of the demographics for the households in the sample.

II. Descriptive Results

I first document the relationship between the choice of the shopping channel for the grocery purchase and brand exploration. I define brand exploration as a household buying a cereal brand that it has not bought previously. Under this definition, even a cereal brand that was introduced long ago can be the object of exploration, as long as it is new to this buyer.

It is difficult to identify brand exploration using transaction data because of left-truncation. I do not observe the brands the household purchased before the beginning of my data series. Therefore, I cannot know definitively when a consumer is purchasing a brand for the first time. To address this problem, I define "brand trial" as the purchase of a cereal brand that the household has not bought in the past three months. This way, I can use the first three months of data to simulate the set of brands known to the consumer at time t_0, and then exploit the remaining 21 months for the estimation. The rolling three-month window ensures that, at any point in time, trials are defined with respect to a past history of the same length. The choice of the three-month period is arbitrary and is subject to the usual problems faced in defining initial conditions (Erdem and Keane 1996). In the online Appendix, I show that a three-month window provides a good approximation of past choices, and that the results are robust to varying the length of this interval.

I observe 142,025 supermarket trips involving the purchase of breakfast cereals, performed by 9,175 different households. In 52,461 trips, the customer buys a cereal

Table 6—Brand and UPC Exploration for In-store versus Online Trips

	In-store	Online	*p*-value
Panel A. Brand exploration			
All trips	0.36	0.26	0.000
	(0.001)	(0.002)	
Trips worth > \$100	0.35	0.26	0.000
	(0.002)	(0.002)	
Basket size > 50	0.36	0.26	0.000
	(0.002)	(0.002)	
Panel B. UPC exploration			
Same brand, same UPC			
Number of trips	56,296	29,630	
Share (percent)	65.4	80.2	
Same brand, different UPC			
Number of trips	29,824	7,314	
Share (percent)	34.6	19.2	

Notes: Brand trial is defined as purchase of a brand not bought in the previous three months; UPC exploration is analogously defined. Panel B conditions on trips in which the household purchased a known brand. *P*-values refer to *t*-test where the hypothesis is equality of the means. Standard deviations are reported in parentheses.

brand she has never tried before. In 42,957 transactions, she buys more than one brand of cereal is bought, which allows for multiple brand explorations in the same transaction. Separately considering multiple cereal purchases in the same shopping trip, I obtain 61,216 trials of new brands in 184,982 purchases.[9]

A look at the raw data (Table 6, top panel) suggests that the average amount of brand exploration is significantly higher in stores than in online trips. This difference persists when I consider only "large" trips, making online and offline trips more comparable. To assess whether the result is robust to the inclusion of controls, I estimate the following linear probability model of trials:

$$Trial_{it} = \alpha + \beta Online_{it} + \tau_i + \varepsilon_{it}, \tag{1}$$

where $Trial_{it}$ is a dummy variable that equals one if consumer i explores a new brand in shopping trip t. $Online_{it}$ is an indicator variable for online trips, τ_i is a household-specific indicator that controls for time-invariant unobserved heterogeneity. I find (Table 7, column 1) that consumers are 8 percentage points less likely to try a new brand of cereals when they are shopping online, a 20 percent reduction.

[9] This implies a 33 percent probability of exploration for each trip. To understand whether this figure is plausible, it may be useful to compare it to measures of state dependence from other studies on scanner data. Brand exploration and simple brand switching are different phenomenon. However, we know that the former should be less frequent than the latter. Consistent with that, several studies report switching probabilities much higher than the probability of exploration observed in my sample. (Dube, Hitsch, and Rossi 2009) uses six products (distinguishing between different sizes within a brand) of frozen orange juice and four brands of margarine and observe that the probability of repeated purchase for a given brand oscillates between 77 percent and 90 percent. (Shin, Misra, and Horsky 2012) restricts analysis to seven toothpaste brands whose probability of repeated purchase ranges between 46 percent and 57 percent. Shum's (2004) study of the cereal market adopts a measure of brand loyalty similar to my definition of exploration. He finds that the probability of switching is around 50 percent.

Table 7—Probability of Exploration

	Dependent variable: New brand dummy OLS	IV	Dependent variable: New UPC dummy OLS	IV
	(1)	(2)	(3)	(4)
Online	−0.08***	−0.04***	−0.11***	−0.95***
	(0.004)	(0.010)	(0.004)	(0.008)
N	163,324	162,625	129,333	101,644

Notes: Trial of a new brand is defined as purchase of a brand not bought in the previous three months; trial of a new UPC is analogously defined. Specifications for new UPC's trial only include trips in which the household did not try a new brand. All specifications include households and day of the week fixed effects. The estimates come from a linear probability model. In column 2 and 4 the instrument is the "potential" fee that would be paid for an online order. Standard errors are clustered at the household level and reported in parentheses.

Potential Explanations for Lower Brand Exploration Online.—The features of my sample make it unlikely that sample selection, differences in price or quality, and reputation of the retailer can create the wedge between online and in-store behavior. There are some other potential explanations:

Sorting of Trips.—The choice of ordering groceries online is endogenous and can be determined by factors correlated with the probability of exploring a new brand. For example, if the customer views online shopping as a time-saving technology, then she will be more likely to do it when she feels under time pressure. In this sense, the quote here from a user of the online service which was collected by the grocer in a customer survey, is enlightening: "I can't do without this service. It is a necessity for busy families."

However, being short of time does not favor experimentation, and thus would generate a spurious negative correlation between brand exploration and online shopping.

The evidence in Figure 1 appears consistent with this story. On the left, it shows the share of in-store[10] trips in which customers try a new brand, by day of the week. The second plot displays the share of online orders, also by day of the week.[11] The two series are negatively correlated. The amount of new brand trials spike during weekends, when people have more time to do their grocery shopping. In contrast, the share of online orders is lower on weekends.

To provide a first assessment of the importance of the bias induced by sorting, column 2 of Table 7 presents an instrumental-variables estimate of the linear probability model of brand trial. The instrument I use is the dollar amount of the delivery fee that the household would have paid if she had decided to order online. The strategy adopted by this retailer gives us a strong reasons to consider this variation as an exogenous shifter for the appeal of online shopping. My results indicate that trip sorting

[10] Online orders are excluded to avoid composition effects.

[11] Figures refer to the delivery day. However, most orders are placed for quick delivery: either same day or next day.

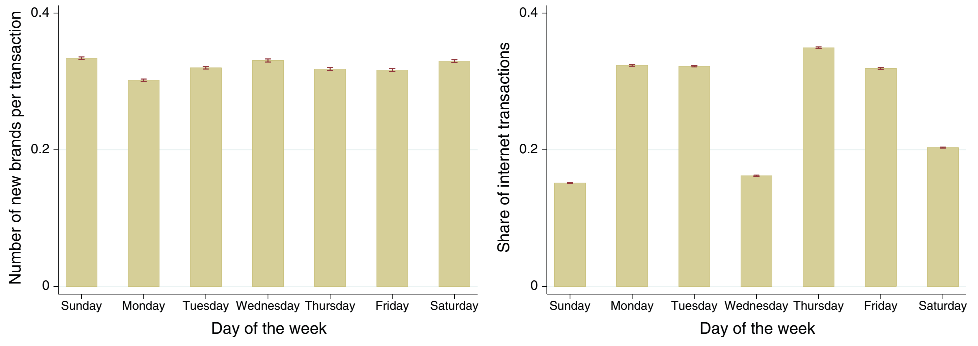

Figure 1. Propensity to Brand Exploration and Online Shopping, by Day of the Week

Notes: Confidence intervals are displayed on top of each bar. For online orders, the day of the week is intended as the day of delivery.

is responsible for half of the wedge in the probability of brand exploration between the online and the in-store context. In the model, I explicitly take this problem into account by modeling the household's decision on the shopping channel.

Uncertainty over Quality.—While the consumer knows the utility she will derive from consuming a brand she has already tried, she can only have expectations about the quality of new brands (Erdem and Keane 1996; Crawford and Shum 2005). This uncertainty can lead her to discount the utility she would derive from switching. Of course, this cost is not specific to the online environment, but verifying the quality of a good is harder online. The higher cost of experimentation online may justify the lower inclination toward brand trial when shopping on the Internet.[12]

Design of the Website and Shopping Cost.—Many e-commerce Web sites feature a favorites list, recommendations, banners or pop-ups, knowing that Web site design can affect consumer behavior (Burke, Harlam, Kahn, and Lodish 1992; Ellison and Ellison 2009). The supermarket chain in this study has a Web site that offers the option of shopping from the list of items (UPCs) already bought in the past. Consumers can save on time spent browsing the site but cannot search for new items this way. Opting for a recommendation system or pop-ups, the chain may have achieved the opposite effect, stimulating brand exploration.

Together, uncertainty over quality and Web site design will bias customers towards purchasing only known brands online. I can separately identify these two stories by exploiting the fact that the "shopping history list" includes UPCs, not brands. Thus, if lack of brand exploration online is caused by web design, then it should lead the customer to not only stick to previously purchased brands, but also to choose the same variety and size of the package. This latter would not occur if

[12] It is worth noting, though, that Mazar, Herrmann, and Johnson (2007) provide experimental evidence that customers evaluate fruit cereals based on non-sensory attributes. They also seem more accurate in matching their declared preferences when they shop for cereals online.

behavior were driven by uncertainty over brand quality. In Table 6 (panel B), I condition on instances in which the household decided to purchase a known brand; I ask whether persistence in purchase of the same UPC differs across channels. Indeed, I find that the fraction of consumers choosing to purchase the same brand in the same size and variety is much higher for online purchases.

Column 3 in Table 7 formalizes this point by showing that online shoppers are 11 percentage points less likely to try new varieties, or box sizes, of a known brand. I interpret this as evidence that households use the past shopping history list to reduce their cost of shopping, which in turn affects brand exploration. Once I account for the endogeneity in the choice of the shopping channel (column 4) again using the delivery fee as an instrument, I find that the cross-channel difference in UPC stickiness is even stronger. This finding is not surprising. Existing estimates of the monetary cost of searching online are quite large (Hong and Shum 2006; Brynjolfsson, Dick, and Smith 2010). It also squares with the evidence that consumers with more experience in Internet shopping tend to develop routines and to cut the length of their transactions (Johnson, Bellman, and Lohse 2003). The logic of these arguments is reinforced in the context of a frequent activity, such as grocery shopping.

III. Model

I develop a model of consumer behavior to disentangle the effect of sorting from the causal impact of the online channel. I assume that households face two sequential decisions. First, they have to select the channel where they want to shop: each trip can take either of the two forms; a visit to a brick-and-mortar supermarket or an online order through the chain's Web site. Conditional on the choice of the channel, consumers select a brand of cereal.

I index each of the N consumers with i, each of the J UPCs with j, and each of the T_i trips made by customer i with t. With the notation Ω_{it} I refer to the set of cereal brands purchased in the past by consumer i, as of trip t. Finally, h_{it} indicates the set of UPCs purchased by consumer i prior to trip t.[13]

A. *Choice of the Shopping Channel*

The latent utility from making trip t online for consumer i is

$$c^*_{it} = \mathbf{C}_{it}\gamma + \mu_i + \theta_{it}. \tag{2}$$

The vector of regressors $\mathbf{C}_{it}$ includes variables relevant for the decision between shopping online and in-store. The fee (in dollars) that the household would have

[13] Construction of both Ω_{it} and h_{it} involves the initial-conditions problem discussed at length in the online Appendix. I adopt here the same approach as in Section II by defining Ω_{it} as the set of brands purchased by household i in the three months prior to the time t trip and h_{it} as the set of UPCs purchased by household i in the same period. This potentially allows for some "forgetting": consumers treat a brand that they purchased farther back in the past as new to them. However, such instances are of little empirical relevance. In my data, fewer than three percent of the cases involve purchase of a brand which the consumer has bought before but forgotten.

to pay for home delivery if it were to order online plays an important role in the decision about the channel. Moreover, that fee is the only variable that varies both in time for the same household and across households at a given point in time. I also include the distance in miles between the house of the customer and the closest grocery store in the chain. I expect that households living further away from a store would be more likely to shop online. A dummy for weekends further captures the fact that time pressure is lower on non-working days, which makes e-commerce less valuable. The other controls are census demographics on the block-group of residence of the household, including education, employment status and age of the head of the household. μ_i is a random effect accounting for unobservable taste for online trips by agent i, which would make this choice correlated over time. The random effect is normally distributed with zero-mean and variance σ_μ. Moreover, it is independent from the i.i.d. error θ_{it} which is also a zero-mean normally distributed disturbance whose variance is normalized to one.

In the data, I observe the choice of the shopping channel (where, with $c = 1$, I refer to an online order) which follows the rule

$$c_{it} = \begin{cases} 1 & \text{if } c_{it}^* \geq 0 \\ 0 & \text{if } c_{it}^* < 0. \end{cases}$$

B. *Demand for Cereals*

The utility consumer i derives from purchasing UPC j in trip t on channel c is modeled as follows for in-store purchases:

$$U_{ijt}^{store} = \beta_0 + \beta_1 Price_{jt} + \beta_2^{store} \times 1\{j \notin h_{it}\} + M_j^{store} + (\delta + \xi_{it}) \times 1\{j \notin \Omega_{it}\} + v_{ijt}. \tag{3}$$

Similarly, for an online trip:

$$U_{ijt}^{online} = \beta_0 + \beta_1 Price_{jt} + \beta_2^{online} \times 1\{j \notin h_{it}\} + M_j^{online} + (\delta + \xi_{it}) \times 1\{j \notin \Omega_{it}\} + v_{ijt}, \tag{4}$$

where *Price* is the price per ounce paid by the household, net of discounts. To simplify notation, I only indexed the price with subscripts for product and time, as if the same product were available at the same price for every household in a particular week. As pointed out before, there is some price variation even within week, because the retailer may post different prices for the same good in different price areas. The indicator $1\{j \notin h_{it}\}$ singles out UPCs that consumer i has not bought in the past three months. Therefore, β_2 represents the impact on utility from purchasing a box of cereals whose UPC did not belong to the shopping history of the household. I allow this effect to be different according to whether the consumer shops online or in a store. M_j is a UPC-fixed effects that also is interacted with the channel dummy. The unobserved shock ξ_{it} only hits brands with which the customer is not familiar $(1\{j \notin \Omega_{it}\})$, representing an instantaneous taste for variety. Finally, v_{ijt} is the usual logit error which is assumed to be uncorrelated with ξ_{it}. The model is estimated using only grocery trips involving purchase of breakfast cereals. The outside good is represented by a basket of all the cereal brands outside the top 100 according to revenues generated in the time period under consideration.

The link between the two parts of the model is given by the generating process of the unobserved shock ξ_{it}. I assume that ξ_{it} and θ_{it}, the disturbances in equation (2), are jointly distributed according to a bivariate normal

$$(5) \qquad \begin{pmatrix} \theta \\ \xi \end{pmatrix} \sim BN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}\right] \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho\sigma_\xi \\ \rho\sigma_\xi & \sigma_\xi^2 \end{bmatrix}.$$

The parameter δ can be interpreted as the mean of the ξ_{it} shock. In some specifications, I assume that the mean taste for new goods can vary according to the channel selected and the time of the week (weekend versus weekday) imposing the following parametric form:

$$(6) \qquad \delta = \alpha_0 + \alpha_1 Internet + \alpha_2 Weekend.$$

C. *Discussion*

The model captures the salient features presented in the descriptive section of the paper. The potential bias deriving from the sorting of trips is addressed by allowing correlation between θ_{it} and ξ_{it}. This acknowledges the fact that unobserved determinants of the choice of shopping online (e.g., lack of time) also can affect the attitude towards brand experimentation.

The change in utility derived from purchase of a new UPC (β_2) can differ for online and in-store trips. Inertia in the choice of the UPC operates both online and in-store: customers tend to purchase the same box size or variety of their brand of choice over time. For this reason both β_2^{online} and β_2^{store} should be negative. If distaste for UPCs never purchased before is only driven by habits or unobservable characteristics, then it should be the same for online and in-store purchases. That is, $\beta_2^{online} = \beta_2^{store}$. I interpret the additional stickiness in the choice of the UPC online (i.e., any difference between β_2^{online} and β_2^{store}) as the effect of Web site design.

Finally, allowing the mean of the ξ_{it} shock (δ) to vary according to the channel of purchase, I capture structural differences in the taste for new brands for in-store versus Internet orders. This difference can be interpreted as coming from higher uncertainty, which makes exploration less attractive online.

D. *Identification*

The detailed information in the scanner data provide several sources of variation that help me to identify the model. Time invariant factors, such as distance from the closest store and household characteristics, help to explain heterogeneity across individuals in inclination toward online shopping. Availability of data from more than 1,000 stores ensures plenty of variation in household location when estimating these parameters.

However, disentangling the sorting effect requires a variable that changes the individual propensity to shop online through time. One such key shifter is the delivery fee. Recall that I can recover coupon-holding even for customers who do not use them. Therefore, I can construct the expected delivery fee that a household faces when deciding where to shop. Because of frequent couponing, there is much variation in the cost of online shopping for the same household through time. This identifies the selection effect under the assumption that a discount on the delivery fee does not affect product choice in the second stage.[14]

The correlation between cost-of-delivery fee and the household's choice of shopping channel is informative about its current utility of time. For example, a customer deciding to shop in-store despite being entitled to free delivery signals that she is not under extreme time pressure. Conversely, a household that pays full price for home delivery must need it enough to accept the extra cost. The weekend dummy is another variable that affects the propensity to shop online and varies through time. However, it also influences the taste for brand trial in some specifications of the demand model. Therefore, it would not represent a valid exclusion restriction in those cases.

Identification of the price coefficient exploits variation across different price areas in the same week as well as changes in price through time within a price area. Once again, the large number of price areas and the frequency of sales provide plenty of this type of variation. The panel structure of the data also allows for including brand dummies that account for quality and ease the problem of endogeneity of price. I allow brand quality to differ across channels acknowledging that differences in the environment (e.g., absence of preferential display) can induce differences between mean utility online and in-store for the same brand.

Finally, the effect of the shopping history list is identified separately from the ease of evaluation, exploiting differential stickiness at the UPC and brand level across shopping channels. Most brands are available in different box sizes or varieties, and I can observe whether the customer purchases the same brand in the same format. Comparing individual persistence in repurchase of brand and UPC online versus in-store identifies the lack of online brand exploration that is due to the higher cost of evaluating new brands on the Internet. Comparison of the stickiness of purchases

[14] This amounts to assuming that the within-category income effect as documented by Gicheva, Hastings, and Villas-Boas (2010) is not too strong or, at least, does not systematically concern cereals in all the categories included in the shopping basket. Milkman and Beshears (2009) uses data from an online grocery Web site to show that receiving a \$10 coupon increases grocery spending by \$1.59. The surge seems to be related to inclusion in the basket of product categories that the household does not usually purchase, rather than to adjustments in regularly bought product categories.

of the same UPC across channels is informative instead about the role of the repurchase feature of the Web site in lowering brand trial.

IV. Estimation

In this section, I describe the procedure for estimating the parameters of the shopping-channel-selection equation and the demand equation.

The data consist of a series of observed channel and product choices for each individual, $\mathbf{c}_i$ and $\mathbf{y}_i$ respectively, as well as observables entering the channel selection and demand equations ($\mathbf{C}$ and $\mathbf{X}$). The model has a complex heterogeneity structure. The richest version requires integrating out two random coefficients and a time-specific shock for each household in the demand equation, as well as the random effect in the channel selection equation. For the case of the mixed logit, which represents the second stage of this model, Train (2001) has shown that Bayesian methods deliver significant advantages in terms of computation time. Therefore, I choose a Bayesian approach for the estimation rather than using simulated maximum likelihood.

I present below the hierarchical Bayes model that is used to estimate the model with random coefficients; the version with fixed coefficients is a simplified case of this model. The details are left to the online Appendix.

A. *Specification of the Priors*

The parameter to estimate in the channel-selection equation is the vector γ. I assume it is normally distributed with mean $\underline{\gamma}$ and covariance matrix $\underline{\mathbf{V}}$. In the Bayesian approach, the random effect is also treated as a parameter: the prior on it is $\mu \sim N(0, \sigma_\mu)$. In the demand equation, the parameters of interest are divided into a set of random coefficients $\boldsymbol{\beta}_i$ and a set of fixed coefficients $\tilde{\boldsymbol{\beta}}$. I assume that $\boldsymbol{\beta}_i \,|\, \mathbf{b}, \mathbf{W} \sim N(\mathbf{b}, \mathbf{W})$ where $\mathbf{b}$ is normally distributed with mean $\underline{\mathbf{b}}$ and covariance $\underline{\mathbf{B}}$ and $\mathbf{W}$ is distributed according to an inverse Wishart with parameters w_1 and w_2. Finally, the distribution of the unobserved shocks ξ_{it} is implied by the joint distribution in equation (5). The covariance matrix of that bivariate normal ($\boldsymbol{\Sigma}$) is distributed as an inverse Wishart with two degrees of freedom and identity inverse scale matrix.

B. *Sampling Scheme*

Let $\mathbf{C}$ and $\mathbf{X}$ be the matrices of the covariates included in the channel selection and demand equation respectively, and recall that $\mathbf{c}$ is the vector of observed channel choices.

Iteration k of the algorithm unfolds as follows:

(i) Apply data augmentation (Tanner and Wong 1987) to the channel selection equation and draw $\mathbf{c}^{*(k)}$ conditional on $\mathbf{C}$, $\mathbf{c}$ and $\boldsymbol{\mu}^{(k-1)}$.

(ii) Draw γ conditional on $\mathbf{c}^{*(k)}$, $\mathbf{C}$.

(iii) Draw $\mu_i^{(k)}$ for each i conditional on $\sigma_\mu^{(k-1)}$. Now implied residuals $\theta_{it}^{(k)}$ can be calculated.

(iv) Draw $\sigma_\mu^{(k)}$ conditional on $\boldsymbol{\mu}^{(k)}$.

(v) Draw $\xi_{it}^{(k)}$ for each i conditional on $\theta_{it}^{(k)}$ and $\boldsymbol{\Sigma}^{(k-1)}$.

(vi) Draw $\boldsymbol{\beta}_i^{(k)}$ conditional on $\mathbf{b}^{(k-1)}, \mathbf{W}^{(k-1)}, \tilde{\boldsymbol{\beta}}^{(k-1)}, \boldsymbol{\xi}_i^{(k)}$.

(vii) Draw $\mathbf{b}^{(k)}$ conditional on $\boldsymbol{\beta}_i^{(k)}, \mathbf{W}^{(k-1)}$.

(viii) Draw $\mathbf{W}^{(k)}$ conditional on $\mathbf{b}^{(k)}, \boldsymbol{\beta}_i^{(k)}$.

(ix) Draw $\tilde{\boldsymbol{\beta}}^{(k)}$ conditional on $\boldsymbol{\beta}_i^{(k)}, \boldsymbol{\xi}^{(k)}$.

(x) Draw $\boldsymbol{\Sigma}^{(k)}$ conditional on $\boldsymbol{\theta}^{(k)}, \boldsymbol{\xi}^{(k)}$.

(xi) Obtain $\boldsymbol{\alpha}^{(k)}$ by Bayesian regression, conditional on $\boldsymbol{\theta}^{(k)}, \boldsymbol{\xi}^{(k)}, \boldsymbol{\Sigma}^{(k)}$.

V. Results

The model is estimated using a random subsample of 500 households[15], for a total of 2,778 trips[16], of which 783 are online orders. The results are displayed in Table 8 and Table 9 for channel selection and demand respectively. I set a burn-in period of 10,000 draws for the MCMC algorithm and estimate parameters using another 10,000 draws. The chain is thinned to reduced autocorrelation by retaining only one out of every ten draws. A Geweke (1992) test fails to reject that the chain has converged.

The covariates enter the channel selection equation (Table 8) with the expected signs. Higher delivery fees make it less appealing to shop on the Internet. Wealth positively affects preference for online shopping, whereas the *weekend* dummy is negative and large (the implied elasticity is 57 percent). Both results are consistent with the utility of time playing a role in the selection of the shopping channel. As observed in the previous literature (Chiou 2009), customers living farther away from a grocery store in the chain are likelier to order online. The variables related to age enter with a negative coefficient because the reference group is 18 to 35 years of age; that is the youngest and most likely to be at ease with technology.

Column 1 of Table 9 presents estimates of the parameters of the demand function obtained by assuming that the mean of the ξ shock to the taste for new brands (δ)

[15] This is aimed at saving computation time. Given the large number of brands included in my analysis, each shopping situation implies listing the covariates for more than 100 brands. This requires to limit the number of trips considered to avoid the matrix of regressors becoming too large and slowing down the estimation routine.

[16] Because demand is modeled as a discrete choice problem and estimated using household level data, I cannot handle trips where multiple brands of cereal were purchased. This leads me to discard about 20 percent of the trips in the sample. Descriptive statistics for single and multibrand trips are similar, especially when it comes to brand exploration. Average rate of brand exploration is .328 in single brand trips and .338 in multibrand trips.

Table 8—Estimates of the Parameters of the Channel Selection Equation

Variable	Coefficient	Variable	Coefficient
Items count	0.0002	Income per capita	0.0049
	(0.0002)		(0.0004)
Distance	0.0087	Age 35–54	−0.0030
	(0.0009)		(0.0002)
Fee	−0.1001	Age 55–64	−0.0071
	(0.0006)		(0.0003)
Weekend	−0.4929	Age $\geq$ 65	−0.0028
	(0.0002)		(0.0002)
Married	−0.0086	College	−0.0036
	(0.0001)		(0.0045)
Employed	−0.0003	σ_μ	0.124
	(0.0001)		(0.0011)

Notes: The unit of observation is a trip (2,778 trips). The dependent variable is an indicator equal to one if the trip is made online. Burn-in period: 10,000 draws, mean and standard deviation of the posterior are computed on the basis of the next 10,000 draws. Marginal effects reported.

Table 9—Demand Estimates

Variable	(1)	(2)	(3)
Price	−0.21	−0. 21	−0.15
	(0.02)	(0.02)	(0.03)
$1\{j \notin h_{it}\} \times Online_t$ (β_2^{online})	−6.96	−6.88	−4.33
	(0.12)	(0.13)	(0.11)
$1\{j \notin h_{it}\} \times (1 - Online_t)$ (β_2^{store})	−6.11	−5.94	−3.35
	(0.07)	(0.07)	(0.10)
δ	0.93		
	(0.05)		
α_0		0. 54	0.61
		(0.04)	(0.04)
α_1		−0.70	−0.47
		(0.17)	(0.17)
α_2		0.07	0.08
		(0.04)	(0.04)
σ_ξ	1.13	1.58	2.11
	(0.06)	(0.06)	(0.34)
ρ	−0.13	−0.16	−0.26
	(0.06)	(0.06)	(0.06)

Notes: The unit of observation is a cereal trip (2,778 observations). Standard deviation of the posterior is reported in parentheses. Burn-in period: 10,000 draws, mean and standard deviation of the posterior are computed on the basis of 10,000 draws of which 1 out of 10 is retained. UPC fixed effects are included in the specification and interacted with the channel dummy. The specification in column 3 includes random coefficients for the β's.

is the same for each trip. I relax this assumption in column 2, where δ is parameterized as in equation (6). The coefficient on price is negative and changes little across specifications. Estimation of the demand parameters offers one way to highlight the competitive consequences of the gap in the cross-channel rate of brand trial. In Table 10, I report the ratio between the in-store and online values of both own and

Table 10—In-store vs. Online Price Elasticity

	$\varepsilon_{xx}^{store}/\varepsilon_{xx}^{online}$	$\varepsilon_{xy}^{store}/\varepsilon_{xy}^{online}$
All brands	1.69	2.98
Top 10 brands	1.7	2.6
Brands 30 to 40	1.69	5.66

Notes: The first column reports the ratio between own price elasticity in-store and own price elasticity online, as implied by demand estimates (Table 9, column 2) for a different subset of brands. The second column analogously displays the ratio between cross elasticity in-store and online.

cross-price elasticities implied by the estimates in column 2 of Table 9.[17] On average, competitive effects are much stronger in-store: own-price elasticity online is 50 percent higher and cross-price elasticity is almost three times greater in-store than in the online context. Further insights can be gained by comparing similar figures for popular brands (the top ten most sold), which are likely to be in the Ω set of a large number of households, with products that are unknown to many (between the thirtieth and fortieth spot in the sales ranking). The gap between in-store and online elasticity is wider for the lesser known brands; this suggests that it is, at least in part, due to differences in the attitude towards brand exploration across channels.

The difference between the estimates of the disutility from purchasing a new UPC online and in-store—that is, the difference between β_2^{online} and β_2^{store} in equations (3) and (4)—identifies the effect of web design on brand choice. This estimate ranges between -0.85 and -0.91. None of the retained draws is positive: the 1 percent symmetric Bayesian confidence interval for the difference is $[-1.26; -0.55]$. Using the estimated coefficient for price, I can compute the dollar value of the effect of web design as

$$\text{Dollar value of the list} = \frac{\beta_2^{online} - \beta_2^{store}}{\beta_1}. \tag{7}$$

This comes to roughly four dollars which is a large number, given that the average cereal box in my dataset costs \$3.55.

The parameter δ represents the mean of the unobserved shock to taste for new brands. The estimate is positive, which is not totally surprising, since the brand proliferation in the industry is consistent with consumers having a taste for variety. However, its impact on utility is tiny compared to that of web-design driven state dependence. Allowing the parameter to vary according to some trip characteristics, I find that that the mean taste for variety is lower on the Internet, possibly capturing the higher uncertainty over quality that is part of online exploration. The parameter α_2 is positive, implying that customers are more likely to value new brands on weekends. This would be the case if, for example, parents are more likely to shop with children during weekends. Finally ρ, the correlation between the unobserved shock to new brands ξ_{it} and the residual from the shopping channel selection, is negative. Large shocks to the

[17] For a market with J products, each product's own price elasticity and $J-1$ cross-price elasticities can be computed for each of the settings. The figures for the cross-price elasticity online and in-store are averages.

instantaneous utility of time (which drive the agent to shop online) are associated with negative shocks to the taste for new brands. This is consistent with online trips being more likely to occur in instances when utility of time is higher.

Because in my model utility derived from consumption of a brand is a function of whether the agent has bought it before, ultimately I am estimating a form of state dependence. It is well known (Heckman 1981; Dube, Hitsch, and Rossi 2010) that unobserved heterogeneity can be confounded with state dependence, but this caveat should not apply in the present context. I am interested in estimating the *difference* between the level of state dependence online and in-store, whereas the aforementioned bias would affect the *level* of state dependence. Unless we allow consumers to have different preferences for brands on the two distribution channels (i.e., liking Rice Krispies more online than in-store), the estimates of the cross-channel gap in state dependence should not be affected by unobserved heterogeneity. Nevertheless, I directly address this concern by estimating the demand model including random coefficients for β_2^{online} and β_2^{store} (Table 9, column 3). As expected, the level of state dependence changes with respect to the baseline model, but the gap between state dependence online and in-store is not altered significantly.

VI. Counterfactual Exercises

In this section, I present some counterfactual experiments that help in understanding the practical implications of the estimates of the structural model. In each counterfactual, each agent's initial shopping history of is based on purchases in the first three months of data. After each simulated trip, the customer's shopping list is updated to include any new brand she could have picked. The parameters used for the simulations are displayed in Table 9, column 2. The results are averages over 10,000 simulations.

A. *Shopping Cost and Brand Exploration*

The estimates of the structural model pointed to the fact that features reducing the cost of shopping online are a major hurdle to online brand exploration. Furthermore, these features are the most relevant to both the players in the industry and to policymakers. In fact, their very existence depends on a choice about the design of the Web site which is completely under the control of the retailer, much unlike instantaneous shocks to the utility of time, or the intrinsic uncertainty over quality of new items.[18]

How much additional brand exploration would take place online if the past shopping history list feature were not available? Recall that I interpret the difference in the estimate of β_2^{online} and β_2^{store} as the effect of this list. Therefore, simulating a world in which the list does not exist implies running the model setting β_2^{online} at the same level as β_2^{store}.[19] Although brand exploration online is still lower than in the store, the

[18] Lewis (2011) argues that extensive use of text and pictures, as measured in megabytes allocated to the description of an item, can reduce information asymmetries in the case of a buyer unable to personally inspect a car. In this spirit, some design changes also could have the effect of reducing the gap in uncertainty over new brands between the online and the traditional channel.

[19] Changing the design of the Web site could affect agents' selection of the shopping channel. This is not captured in the counterfactuals, as the selection of the shopping channel is not simulated. Performing this extra step is

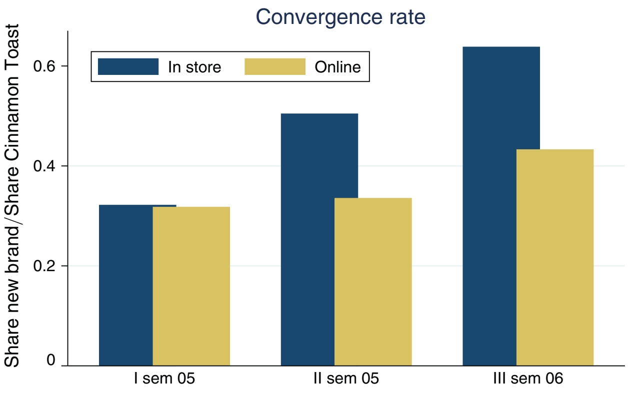

Figure 2. Simulation of the Entry of a New Brand

Notes: The bars represent the ratio between the market share of a new entrant with the same price and characteristics as Cinnamon Toast Crunch and Cinnamon Toast Crunch's market shares in the baseline. Figures are averages of 10,000 simulations.

total number of brand trials in online purchases over the two-years period increases by 23 percent in this simulation.

B. *Barriers to Entry*

Low online brand exploration may mean higher entry barriers for new brands on the Internet distribution channel. If consumers often use features such as the "shopping history" to save on time spent browsing, then new entrants—which by definition do not show up on such lists—will struggle to capture Internet customers. There are challenges to measure the relevance of this effect using sales data. First, I do not observe entry of any major cereal brand during the period under analysis. Furthermore, even if I did, the results would be hard to interpret. Finding that the new cereal sells more in-store than online would not constitute compelling evidence because it might be a natural result of the traditional channel being more popular.

So, I instead illustrate the presence of entry barriers using a counterfactual exercise. I remove Cinnamon Toast Crunch from the initial shopping histories of all consumers in the sample and then observe the evolution of its market share in-store and online *compared* to the same figures in the baseline. This is equivalent to observing the performance of a new brand with the same price and characteristics as Cinnamon Toast Crunch but lacking any installed base.[20] The original market share of Cinnamon Toast Crunch serves as a normalization factor, accounting for the performance the brand should achieve, given its characteristics. Therefore, any

unfeasible because nothing in the data identifies the impact of changes in design on taste for the online channel. In fact, the grocer's Web site maintained the same structure throughout the sample period. The changes considered in the counterfactuals probably would reduce the appeal of the Internet service, because they limit its potential to reduce the cost of shopping. Therefore, the reported increases in the amount of online exploration can be seen as upper bounds.

[20] To be precise, the exercise simulates the entry of such a brand in a world where Cinnamon Toast Crunch did not exist.

deviation of the counterfactual market share from the original share can be interpreted as attributable to the existence of entry barriers.

Figure 2 shows the ratio of the counterfactual to the baseline market share over time on the two channels (online and in-store). The new brand performs similarly over the two channels in the first semester. However, this new entrant eventually recovers more than 60 percent of its "potential" market share in-store. On the other hand, its online market share after 18 months is still less than half of what it was in the baseline case. I interpret this as evidence that entry barriers faced by the entrant are higher online than in-store.

C. *Recommendation Systems*

Contextual advertising and customer recommendations are popular forms of advertising and are customer-loyalty enhancing for online businesses. Context ads are individually targeted promotions, based on the content of the page the consumer is browsing. For example, they are regularly displayed along with the results from a search on a search engine. Recommendation systems offer suggestions to a customer based on her previous shopping history. The firm uses purchases by other customers with a similar shopping history to forecast the goods or offers that may be of interest to her. This system has been made popular by Amazon.com and Netflix. Inclusion of recommendations or context ads in the design of the Web site can make it easy for consumers to notice brands that they have not tried before, and thus to encourage exploration.

I simulate the effect of introducing a recommendation system on the grocer's Web site and then assess its impact on brand competition. Once again, I remove Cinnamon Toast Crunch from the brand history of every household. However, in this simulation, I set, only for this brand, β_2^{online} at the same level as β_2^{store}. This means that the effort required of a household to purchase the new entrant is the same as the effort needed to purchase another brand that is already part of the household's shopping history list. This would be the case if the brand appeared in a recommendation box that pops up when the household is browsing its shopping history list.

I find that the evolution of brick-and-mortar market share for the new entrant resembles that of the previous counterfactual. However, there is a marked difference in its performance in the online channel. The new cereal now reaches a higher market share than that of by Cinnamon Toast Crunch in the baseline case within six months of its introduction. This implies that if a recommendation system were in place, it would take just a short time for a new entrant to make up for its lack of customer base. The introduction of customer recommendations thus completely changes the bottom line for the competitive effect of online shopping. In this scenario, the new entrant is more visible and performs better online. In fact, by the end of the simulation period, the online market share is 50 percent higher than its benchmark.

VII. Conclusions

In this paper, I compare brand exploration in brick-and-mortar stores versus Internet shopping. I document that brand exploration is more prominent in-store and I quantify the role of three different explanations for this result. My estimates suggest that Web site features that reduce the time spent shopping (such as "favorites lists") play a major role in limiting online brand exploration. The fact that the Internet allows for such reductions in the cost of shopping is especially relevant in certain contexts, and leads customers to more repetitive behavior, lower search cost notwithstanding. The counterfactual exercises confirm that this result implies higher barriers to entry online. However, the introduction of pop-up ads or customer recommendations can foster brand trial on the Internet, making it easier for new entrants to succeed online.

REFERENCES

Ackerberg, Daniel, Keisuke Hirano, and Quazi Shahriar. 2006. "The Buy-it-now Option, Risk Aversion, and Impatience in an Empirical Model of eBay Bidding." http://www.econ.ucla.edu/ackerber/ahs_2006_11_271.pdf.

Bakos, J. Yannis. 1997. "Reducing Buyer Search Costs: Implications for Electronic Marketplaces." *Management Science* 43 (12): 1676–92.

Bayus, Barry L. 1992. "Brand Loyalty and Marketing Strategy: An Application to Home Appliances." *Marketing Science* 11 (1): 21–38.

Bhatnagar, Amit, Sanjog Misra, and H. Raghav Rao. 2000. "On Risk, Convenience, and Internet Shopping Behavior." *Communications of the ACM* 43 (11): 98–105.

Brown, Jeffrey R., and Austan Goolsbee. 2002. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy* 110 (3): 481–507.

Brynjolfsson, Erik, Astrid A. Dick, and Michael D. Smith. 2010. "A Nearly Perfect Market?: Differentiation vs. Price in Consumer Choice." *Quantitative Marketing and Economics* 8 (1): 1–33.

Brynjolfsson, Erik, Yu Jeffrey Hu, and Duncan Simester. 2007. "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on Concentration of Product Sales." Massachusetts Institute of Technology (MIT) Center for Digital Business Working Paper 276. http://digital.mit.edu/research/papers/2007.11_Brynjolfsson_Hu_Simester_Goodbye%20Pareto%20Principle_276.pdf.

Brynjolfsson, Erik, and Michael D. Smith. 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science* 46 (4): 563–85.

Burke, Raymond R., Bari A. Harlam, Barbara E. Kahn, and Leonard M. Lodish. 1992. "Comparing Dynamic Consumer Choice in Real and Computer-Simulated Environments." *Journal of Consumer Research* 19 (1): 71–82.

Chen, Pei-Yu, and Lorin M. Hitt. 2002. "Measuring Switching Costs and the Determinants of Customer Retention in Internet-Enabled Businesses: A Study of the Online Brokerage Industry." *Information Systems Research* 13 (3): 255–74.

Chiou, Lesley. 2009. "Empirical Analysis of Competition between Wal-Mart and Other Retail Channels." *Journal of Economics and Management Strategy* 18 (2): 285–322.

Chu, Junhong, Pradeep Chintagunta, and Javier Cebollada. 2008. "A Comparison of Within-Household Price Sensitivity across Online and Offline Channels." *Marketing Science* 27 (2): 283–99.

Clemons, Eric K., Il-Horn Hann, and Lorin M. Hitt. 2002. "Price Dispersion and Differentiation in Online Travel: An Empirical Investigation." *Management Science* 48 (4): 534–49.

Crawford, Gregory S., and Matthew Shum. 2005. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica* 73 (4): 1137–73.

Danaher, Peter J., Isaac W. Wilson, and Robert A. Davis. 2003. "A Comparison of Online and Offline Consumer Brand Loyalty." *Marketing Science* 22 (4): 461–76.

Degeratu, Alexandru M., Arvind Rangaswamy, and Jianan Wu. 2000. "Consumer Choice Behavior in Online and Traditional Supermarkets: The Effects of Brand Name, Price, and Other Search Attributes." *International Journal of Research in Marketing* 17 (1): 55–78.

Dubé, Jean-Pierre, Günter J. Hitsch, and Peter E. Rossi. 2009. "Do Switching Costs Make Markets Less Competitive?" *Journal of Marketing Research* 46 (4): 435–45.

Dubé, Jean-Pierre, Günter J. Hitsch, and Peter E. Rossi. 2010. "State Dependence and Alternative Explanations for Consumer Inertia." *RAND Journal of Economics* 41 (3): 417–45.

Einav, Liran, Ephraim Leibtag, and Aviv Nevo. 2010. "Recording Discrepancies in Nielsen Homescan Data: Are They Present and Do They Matter?" *Quantitative Marketing and Economics* 8 (2): 207–39.

Ellison, Glenn, and Sara Fisher Ellison. 2009. "Search, Obfuscation, and Price Elasticities on the Internet." *Econometrica* 77 (2): 427–52.

Erdem, Tülin, and Michael P. Keane. 1996. "Decision-Making Under Uncertainty: Capturing Dynamic Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science* 15 (1): 1–20.

Geweke, John. 1992. "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments." In *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M Smith, 169–94. New York: Oxford University Press.

Gicheva, Dora, Justine Hastings, and Sofia Villas-Boas. 2010. "Investigating Income Effects in Scanner Data: Do Gasoline Prices Affect Grocery Purchases?" *American Economic Review* 100 (2): 480–84.

Goettler, Ronald L., and Karen Clay. 2011. "Tariff Choice with Consumer Learning and Switching Costs." *Journal of Marketing Research* 48 (4): 633–52.

Hann, Il-Horn, and Christian Terwiesch. 2003. "Measuring the Frictional Costs of Online Transactions: The Case of a Name-Your-Own-Price Channel." *Management Science* 49 (11): 1563–79.

Heckman, James J. 1981. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process." In *Structural Analysis of Discrete Data with Econometric Applications*, edited by Charles F. Manski and Daniel McFadden, 179–97. Cambridge, MA: MIT Press.

Hong, Han, and Matthew Shum. 2006. "Using Price Distributions to Estimate Search Costs." *RAND Journal of Economics* 37 (2): 257–75.

Jin, Ginger Zhe, and Andrew Kato. 2007. "Dividing Online and Offline: A Case Study." *Review of Economic Studies* 74 (3): 981–1004.

Johnson, Eric J., Steven Bellman, and Gerald L. Lohse. 2003. "Cognitive Lock-In and the Power Law of Practice." *Journal of Marketing* 67 (2): 62–75.

Klemperer, Paul. 1995. "Competition When Consumers Have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade." *Review of Economic Studies* 62 (4): 515–39.

Lewis, Gregory. 2011. "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors." *American Economic Review* 101 (4): 1535–46.

Mazar, Nina, Andreas Herrmann, and Eric J. Johnson. 2007. "Preferences and Choice Behavior Online: A Cognitive Cost Approach to Understanding Product Class Differences." Unpublished.

Milkman, Katherine L., and John Beshears. 2009. "Mental Accounting and Small Windfalls: Evidence from an Online Grocer." *Journal of Economic Behavior and Organization* 71 (2): 384–94.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica* 69 (2): 307–42.

Pozzi, Andrea. 2012. "Shopping Cost and Brand Exploration in Online Grocery: Dataset." *American Economic Journal: Microeconomics*. http://dx.doi.org/10.1257/mic.4.3.96.

Scherer, F. M. 1979. "The Welfare Economics of Product Variety: An Application to the Ready-to-Eat Cereals Industry." *Journal of Industrial Economics* 28 (2): 113–34.

Schmalensee, Richard. 1974. "Brand Loyalty and Barriers to Entry." *Southern Economic Journal* 40 (4): 579–88.

Schmalensee, Richard. 1978. "Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry." *Bell Journal of Economics* 9 (2): 305–27.

Shin, Sangwoo, Sanjog Misra, and Dan Horsky. 2012. "Disentangling Preferences and Learning in Brand Choice Models." *Marketing Science* 31 (1): 115–37.

Shum, Matthew. 2004. "Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast-Cereals Market." *Journal of Economics and Management Strategy* 13 (2): 241–72.

Tanner, Martin A., and Wing Hung Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* 82 (398): 528–40.

Train, Kenneth. 2001. "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit." http://elsa.berkeley.edu/~train/compare.pdf.