Repetitive Buying Behavior
An Empirical Investigation on the Role of Personal and Product Characteristics

Karsten Hansen Romana Khan Vishal Singh

1 Karsten Hansen is Associate Professor of Marketing at the Rady School of Management, UCSD; Romana Khan is Assistant Professor of Marketing at Ozyegin University, and Vishal Singh is Associate Professor of Marketing at the Stern School of Business, New York University.
Repetitive Buying Behavior: An Empirical Investigation on the Role of Personal and Product Characteristics*

July 1, 2014

Abstract

This article provides a large scale empirical study on the extent of repetitive versus variety-seeking behavior in our daily shopping choices, and how this varies with personal and context characteristics. We utilize a unique database that records detailed grocery shopping histories for a nationwide sample of over 100,000 households observed over several years. Across a wide variety of product categories, we find that consumers display high levels of repetitive buying and narrow choice sets, even over an extended period of several years. This behavior is moderated by household and product specific factors. Repetitive buying behavior is found to increase with age, income, education, conservative ideology, and for males; and decrease with the frequency of purchase. These relationships vary with category characteristics that measure the stability of the choice context and the utilitarian versus hedonic nature of the product. Keywords: habits, automaticity, variety-seeking, ideology, ageing, gender

*Note: The figures in this paper make extensive use of color and should be viewed digitally or printed on a color printer. The authors would like to thank the seminar participants at NYU Social Psychology, Marketing Department at the Kellogg School of Management (Northwestern), and the Behavioral Economics and Decision Research (BEDR) group at Cornell University for useful comments and suggestions. They would also like to acknowledge the support of the James M. Kilts Center for Marketing at The University of Chicago Booth School of Business and the Nielsen Company for providing the data used in this paper.
1 Introduction

From the moment we wake up to the point of sleep, our daily lives are consumed with choices and decisions—what clothes to wear to work, what to eat for breakfast, which mode of transport to use, what to read from a potentially infinite number of daily news sources, and so forth. Although on many occasions such actions require conscious cognitive deliberation (dressing for a special meeting, or driving to work rather than public transport), the majority of our daily judgments and behaviors are guided by implicit cognition that is spontaneous, effortless, and unconscious (Bargh (1994), Moors and De Houwer (2006), Ashby and Crossley (2012)). Consider for instance a casual trip to your neighborhood grocery store. A typical supermarket in the US offers more than 40,000 unique products that are organized into hundreds of product categories (e.g., breakfast cereal, laundry detergent, and paper towels). Within each category are hundreds of options differentiated by attributes such as brand name, price, size, flavor, and temporary promotions. For a Martian or a new immigrant to the Western world, a choice amongst such a large array of options can be rather daunting, or even paralyzing — the psychological costs of what is often referred to in the literature as “choice overload” (Schwartz (2004), Iyengar and Lepper (2000), Botti and Iyengar (2006)). To avoid making cognitively hard choices, many consumers may instead rely on well-rehearsed shopping protocols and heuristics such as “repeat buy the favorite product” or “buy whatever brand is on promotion”. At the same, an array of options allows us the freedom to indulge our need for variety – poetically identified as the “very spice of life, that gives it all its flavour” (Cowper (1785)).

Suppose we reviewed our shopping history over several years for a variety of frequently purchased grocery products such as cereal, detergent, toothpaste, or soup. What would that product “portfolio” look like? Would our purchases be concentrated on a small set or even a single product within a category? Or would they be diversified across a large set that varies on attributes such as brand name, price, flavor, or size? Would behavior differ based on the characteristics of the product? For example, while our choices of utilitarian products such as detergent could be highly concentrated, we might seek variety in hedonic consumption of ice-cream and desserts. Would this tendency toward repetitive buying versus variety-seeking relate to our personality traits such as extraversion and openness to experience, or even to demographics such as socioeconomic status, gender, or age? In this article
we seek to address these issues by analyzing a total of 123 million decisions made by over 100,000 households in a large number of product categories.

The research questions posed in this paper are of interest from several perspectives. First, given the high purchase frequency and low ticket value of most grocery products, consumers are unlikely to spend significant cognitive resources in evaluating options on every purchase occasion. This in turn suggests a routinized, habitual type of shopping behavior characterized by high levels of repeat buying (Howard and Sheth (1969)). An extensive literature in psychology and cognitive science documents that repetitive behaviors (and thoughts) in a stable context can take the form of automated reflexive responses to external or internal triggers, i.e., take the form of habits (Verplanken & Aarts (1999), Wood & Neal (2007), Danner et. al (2008), Graybiel (2008), Neal, Wood, & Lally (2011)). Empirical evidence of such habitual behavior exists on casual daily activities, as well as important domains such as diet, exercise, and energy consumption (Cohen and Farley (2008), Marechal (2009), Orbell and Verplanken (2010)). Since grocery shopping is a frequent activity (averaging 2.2 shopping trips per week for a typical US household (FMI, (2011)) and occurs in the stable context of a small set (often single) of neighborhood grocery stores, we might expect shopping behavior to display a high level of repetitive buying. Taste preferences and brand loyalty for a particular product (Kahn, Kalwani, and Morrison 1986) are also likely to contribute to repetitive purchase patterns when people choose to consume their most preferred option.

Other factors may however mitigate the tendency toward repetitive choice outcomes. An extensive literature documents an intrinsic consumer motivation to seek variety even in the presence of the favorite option (McAlister (1982), Read and Lowenstein (1995), Ratner, Kahn, and Kahneman (1999)). This desire for variety could be driven by a number of factors such as uncertainty of preferences (Pessemier (1978), Kahn and Lehmann (1991), Kahn (1995)), signaling uniqueness and open mindedness (Ariely & Levav (2000)), or simply satiation with existing options and a desire for novelty (McAlister (1982), Inman (2001)). In addition, the impact of satiation may be more apparent with higher product usage or purchase frequency, leading consumers to desire more variety (Galak, Kruger and Loewenstein (2011)). Deviation from the default or favorite option could also be triggered by external factors (Van Trijp, Hoyer and Inman (1996)). Although grocery shopping shares many features
of mundane habitual activities considered in previous work, it occurs in a stimulus-rich environment of advertising and frequent point-of-purchase promotions. Furthermore, the consumer packaged goods (CPG) industry is characterized by a high level of new product introductions that include new brands as well as modifications (e.g., new flavors) of existing products. Such intrinsic and/or environmental factors could in turn lead to a product portfolio characterized by a wide array of product choices over time.

Although theoretical arguments exist for both, repetitive versus variety-seeking behavior is likely to vary across product type and across consumers (Van Trijp, Hoyer and Inman (1996). As noted above, people may be more habitual for utilitarian products such as detergent and garbage bags, while seeking variety in hedonic consumption of ice-cream and desserts (Kahn and Lehmann (1991)). Similarly there could be systematic differences across consumers based on their personality traits and/or demographics such as gender and education. Consider for instance age. A large stream of literature in psychology and neuroscience examines the impact of aging on cognition, emotions, and decision making (for reviews, see Ballesteros et al. (2009), Yoon et al. (2009), Charles and Carstensen (2010), Salthouse (2012)). This literature finds an ubiquitous age-related cognitive decline in domains such as processing speed, fluid reasoning, episodic memory, and concentration. Older adults are also found to focus more on emotion regulation, drawn less to new information, and generally tend to have smaller social networks and spheres of interests (Carstensen (2006)). Given these findings, we might expect aspects of cognitive aging to persist beyond the social and personal domains such that older consumers also show higher brand affinity and a narrower product portfolio in their grocery purchases. Similarly, the tendency toward repetitive behavior may vary across consumers based on differences in personality traits such as openness to new experiences and desire for novelty/change.

To explore these issues, we analyze a unique panel database that records the details of grocery shopping histories for a nationwide sample of over 100,000 households. The data are observed for six years (2004-09) and include a wide array of frequently purchased products typically found in supermarkets. For each household on each shopping occasion, we observe the details of the product

---

¹For instance, in 2011 the Consumer Packaged Goods (CPG) industry spent approximately $470 billion in couponing activities alone (NCH, Coupon Facts 2011). These include coupons targeted at the point of checkout; for example, a Coke buyer may get a coupon from Pepsi on the back of the checkout receipt.
chosen, price paid, and indicator of any promotional activities. We utilize the household transaction histories to compute measures of purchase concentration in a wide variety of food and non-food products. The data includes detailed demographics on households such as income, education, age and household size. Although we do not have measures on personality traits, our strategy is to create theoretically relevant proxies using external data. For example, extensive research in political ideology shows systematic differences across liberals and conservatives on traits such as openness to change and desire for novelty (Jost et. al. (2003), Jost (2006)). This in turn suggests that higher level of “conservativeness” may manifest in higher levels of brand loyalty and purchase concentration. In the empirical application, we use the household’s residential location (zip codes and county) to create measures of conservativeness using data on political voting and religiosity.

The empirical analysis presented in the paper is related to several lines of research. Besides the experimental work discussed above, a stream of research in Economics and Marketing addresses a similar question about the degree to which consumers display “inertia” or “variety seeking” in purchase behavior (e.g., Keane (1997), Erdem and Sun (2001), Seetharaman (2004)). The focus in this line of research is on sequential purchase decisions, and the extent to which past purchases versus inherent preferences are the drivers of choice. The general conclusion is that household’s exhibit inertia in brand choices over time (Keane (1997), Seetharaman, Ainslie and Chintagunta (1999)) driven by the psychological costs of switching (Dube, Hitsch and Rossi (2010)). Given the primary objectives of understanding the underlying decision processes, these applications typically analyze the behavior of a small sample of households (200-800), in a few (often single) product categories, over limited time periods (usually 1-2 years). Our objectives in this paper are different. In this paper we do not seek to disentangle and interpret the individual elements of the purchase decision process (i.e., the decision “inputs” such as brand preferences, price sensitivities etc.), but focus instead on the concentration in decision outcomes (i.e., “outputs”). How concentrated are choices when aggregated over hundreds of purchase occasions for a single consumer? How does this concentration vary by choice contexts (e.g., different types of product categories) and across different demographic segments? Across heavy and light users of a product? This article addresses these questions by analyzing the purchase histories of a large sample of households across a broad range of food and non-food products. We seek the answers
to these questions not only out of idle curiosity, but also out of the firm belief that broad empirical
generalizations of consumer behavior is lacking in marketing. How much variety do consumers actually
seek in markets? We aim to provide at least some evidence on this issue in this paper.

The outline of the paper is as follows. Section 2 provides the context of the study and gives a
detailed account of the data used in the paper and discusses measures of repetitive buying or choice
congestion. In Section 3 we describe our methodology and presents the main findings. Section 4
concludes the paper.

\section{Context of the Study}

\subsection{Data Description}

The cornerstone of our empirical strategy is an extensive consumer panel database provided by AC-
Nielsen.\footnote{All data used in the paper can be obtained for academic use from Chicago Booth Kilts Research Center (http://research.chicagoboost.edu/kilts)} Nielsen provides households in its panel with in-home optical scanners that are used to
record details of grocery purchases from all retail outlets. The data used in this paper contains the
purchase histories of approximately 100,000 demographically balanced households spanning 50 US
states. The data are observed from 2004-09 and households stay in the panel for an average of 3
years. For approximately 18,000 households, we observe purchase histories for the entire six year
period.

The database contains information on each store visit including date of purchase, identification of
the retail outlet, and the total amount spent on the shopping occasion. Furthermore, the database
records detailed information on every product in the shopping basket such as the product description
(brand name, size, fat content, etc.), number of units purchased, price paid, and indicators for any
promotion or coupon usage. We also observe a large number of household demographics that are
discussed in detail below. Over the past several decades, these data have become one of the most
important tools for understanding consumer buying behavior in the CPG industry and has spawned
a vast academic literature in marketing.
2.2 Purchase Concentration Measures

The primary objective of the paper is to analyze the degree to which product purchases are concentrated on a few options. The first natural question that arises is how to define a “product”. A typical supermarket in the US contains approximately 40,000 unique items that are organized into several hundred product categories (e.g., Yogurt, Breakfast Cereal, Paper towels). Each category in turn contains several brands and variants within brands based on packaging, size, flavor, fat content, and so forth. For example, the Yogurt category contains store brand (or private label) as well as several national brands such as Dannon, Yoplait, Stonyfield, and Chobani. Within Dannon, there are a total of 314 unique products or UPCs (Universal Product Code) over the 6-year period that vary by packaging, size, flavor, and fat content. Naturally the degree to which consumers display variety seeking or repetitive buying behavior will crucially depend on the level of aggregation. For example, a brand loyal customer of Dannon may have high product concentration when measured at the brand level, but may display high level of variety seeking in other attributes such as flavor. Our approach in the paper is to create concentration measures at the UPC level, which is the most granular unit of analysis possible.

Having defined the unit of analysis, we next need to create measures of purchase concentration. Several metrics exist for measuring concentration. For example, one could simply count the total number of unique products purchased over a given time span (say a year). This metric is often used in the marketing area as a proxy for “consideration set”, i.e., a measure of the total number of products considered by a consumer. An obvious problem with this metric is that it ignores quantity/expenditure. For the results reported here we use a commonly used measure, Concentration Ratio (CR). For each household, the CR1 in a category is defined as the percentage of total category expenditure (or total volume) allocated to the household’s top choice (in terms of expenditure or volume) in the category. CR2 is the percentage of the household’s total expenditure (or volume) allocated to the household’s top two choices in the category. The more concentrated the choices, the higher the value of CR; the more varied the choices, the lower the value of CR. Note that CR can be computed based on dollar expenditures as well as quantities (for example, ounces of Yogurt consumed). These measures are highly correlated, and in what follows we report results based on CR1-UPC expenditure, which is the
proportion of total money spent in the category that is allocated to the top UPC.\footnote{We have also computed other concentration measures such as the Herfindahl-Hirschman Index (HHI), originally used in Economics to measure concentration of firms in an industry, and Shannon Entropy used in Information Theory to measure the uncertainty of a random variable. All concentration measures were highly correlated and the choice of specific metric used does not alter any of the results reported in the paper.}

For the empirical analysis, we selected the top 100 product categories in the data based on total volume. These categories include a wide variety of products such as food (e.g., cereal, soup), non-food (e.g., detergent, toilet papers), condiments (e.g., ketchup, salad dressing), health and beauty (e.g., shampoo, toothpaste), and over-the-counter medicines (e.g., headache remedies, multi-vitamins). Since purchase concentration by default is very high when only one or two purchases are made, we exclude households who have made less than 5 purchases during the sample period. Households with at least 5 purchases in a category are labeled as “category users” and included in the final sample. Thus for each household in the data, we create concentration measures for every product category in which the household is a user.\footnote{Keeping only observations where households make at least 10 purchases does not alter the results in any meaningful way.} Our final sample size is 4,416,917 purchase concentration observations spread out across 100 categories and 100,439 households.

**Purchase Concentration: Summary Statistics** Average concentration measures across all households and product categories are displayed in Fig.\footnote{These summary statistics are computed using equal weights for all households, i.e., simple averages. We also computed the same measures using sample weights that reweigh the sample to make it comparable to the US Census. We got almost identical results: The average CR1 and CR2 measures across all categories were 40% and 58%.}. In the top panel we provide overall averages as well as measures for households that are in the data for one and six years. For comparison we provide averages for both the top UPC (CR1) and the top two UPCs (CR2). The average purchase concentration is high in most categories. For example, the average CR1 for households observed over one year is between 40% and 70% for the majority of categories. For about half of the 100 categories for households observed for one year, the average concentration ratio is over two-thirds for the top two products. Even for households whose purchase history is 6 years long, we see about 50 categories where the average CR2 measure is above 50%.

The bottom left panel provides the histogram of these 100 averages for both CR1 and CR2. The average CR1 and CR2 across all 100 categories are 39% and 58% respectively.\footnote{These summary statistics are computed using equal weights for all households, i.e., simple averages. We also computed the same measures using sample weights that reweigh the sample to make it comparable to the US Census. We got almost identical results: The average CR1 and CR2 measures across all categories were 40% and 58%.} This means that the average household in the sample on average across categories spends almost 40% of all category
expenditure on one single product (defined as a brand, size, flavor etc. combination) and almost 60% on the top two products. The bottom right graph in Fig.4 shows household level variation of CR1 in three products. This chart shows that the overall category averages in the left panel mask substantial heterogeneity across consumers: Each of the three categories have a right tail of highly concentrated households and a left tail of households exhibiting highly diffuse purchase behavior.

In evaluating these concentration numbers, a few institutional details are worth considering. First, the analysis at UPC level implies that minor substitutions, for example, the purchase of two different sizes of Diet Coke, would lower the CR1 metric. In addition, in the CPG industry minor variations of products (e.g., slight change in package design) is often accompanied by change in the UPC code. Thus the concentration ratios reported above provide a lower bound on the degree of purchase concentration in this context.

2.3 Correlates of Purchase Concentration

The discussion above shows that consumer purchases for grocery products tend to be concentrated on a few options even over long time horizons. At the same time, concentration varies significantly–both across product categories as well as consumers within categories. Can this variation be related to household and product characteristics? For example, can we explain why some households are more concentrated than others purely based on the characteristics of that households, e.g., age and income? We next turn to this question.

2.3.1 Household Characteristics

The database contains a large set of demographics such as Income, Education, and Household Size. Note that certain variables in the data are measured at household level (for example, Income), while others are reported separately for Male and Female household heads (for example, Education). For these variables, we assume females to be the primary shopper and use female demographic variable in the estimation (e.g. education level for the female head). Summary statistics of the sample demographics are reported in Fig.2. We briefly describe each demographic variable along with our priors on how it may impact purchase concentration.
Figure 1: Purchase Concentration Summary Statistics. Note: The top row shows the average concentration ratio for the households top UPC choice and top two UPC choice for 100 product categories. The bottom left figure shows the distribution across 100 categories of the average concentration ratio for the top UPC (CR1) and top two UPC (CR2). The bottom right figure shows the distribution across households of CR1-UPC for three product categories.
Figure 2: Distribution of Demographics, 100,439 Households.
**Age:** Biological aging inevitably impacts all aspects of our physiological and psychological life. A large stream of literature in psychology and neuroscience examines the impact of ageing on cognition, emotions, and decision making (Ballesteros et al. (2009), Yoon et al. (2009), Charles and Carstensen (2010), Salthouse (2012)). Although empirical evidence on aging and variety seeking behavior is limited, consumer surveys show that older consumers tend to prefer older established brands (Lambert-Pandraud and Laurent (2010)). On a broader scope, older people are observed to have smaller social networks and spheres of interest, are drawn less to novelty, have reduced sensation-seeking (Zuckerman (1979)), and place greater importance on emotionally close and familiar contacts (Carstensen, Charles, and Fung (2003), Carstensen (2006)). Given these findings, we expect purchase concentration to increase with age.

**Socioeconomics:** The data contains two measures of households’ socioeconomic status: Income and Education. Although these variables tend to be correlated, the data has sufficient variation to identify the impact of each.\(^6\) While there are differences in decision-making based on socioeconomic characteristics (de Bruin, Parker and Fischhoff (2007)), the impact of these variables on the degree to which consumers may display variety seeking or habitual behavior is unclear. Higher education is likely to be associated with greater awareness of product alternatives, while higher income is likely associated with lower financial restrictions, and greater opportunities to explore newer options. Higher education is also associated with higher optimal stimulation levels and exploratory behavior (McAllister and Pessemier (1982), Raju (1980)). Together this suggests that households with higher socioeconomic status may seek more variety in their choices. At the same time, the CPG industry is characterized by high levels of promotional activities such as coupons, price cuts, and in-store displays. Since lower income consumers tend to be more price sensitive, they are more likely to be swayed by such promotions and choose products based on the best available deal. Consumers in higher socioeconomic brackets can afford the luxury of purchasing their favorite option even in the presence of price discounts by rival products.

**Household Composition: Size & Gender:** Households in the Nielsen panel differ in terms of both size (number of members) as well as composition: Joint households (both female and male

\(^6\)See online appendix for joint distribution of Income and Education.
household present), single female (no male household head) and single male (no female household head). In our data, we have approximately 7,000 single males and over 15,000 single females. Note that single male and female are defined as single head households with no other members (i.e., living alone). All multi-unit households are classified based on the size of household (2-member, 3-member, etc.). Larger households should display lower levels of purchase concentration to cater to preferences of different members.

Given the large sample size, we can isolate the behavior of males and females, which to the best of our knowledge has not been studied using purchase history data. Although a large body of research exists on documenting gender differences in various cognitive and psychological domains (Eagly et al. (2012)), there is little empirical evidence on gender differences in variety seeking behavior. Research on sensation-seeking finds that optimal stimulation levels are higher for men than for women (Zuckerman (1979)) but whether this manifests in gender differences in variety-seeking behavior in the grocery context is uncertain.

**Ideology:** Besides consumer demographics, the tendency of habitual or variety seeking behavior may be related to personality traits such as extroversion and openness to new experiences (Raju (1980)). Although our database contains detailed information on household demographics, we have no measures on such personality traits. Our strategy instead is to combine household residential location with measures of political voting and religiosity to capture the degree of “conservativeness”. Extensive research in political ideology shows systematic differences across liberals and conservatives on traits such as openness to experiences, conscientiousness, and desire for novelty (Jost et. al. (2003), Jost (2006)). These personality traits, if captured by measures of religiosity and voting behavior, could manifest in higher levels of product loyalty and purchase concentration. To explore this question, we matched the households in our data with county level measures of voting behavior (using 2012 Presidential election) and religiosity. For religiosity, we use data provided by the Association of Religion Data Archives (ARDA) on religious adherence at the county level (2010 report). Fig. shows the geographical distribution of these measures. Although there appears to be large geographical variation, it is important to note that over 60% of households reside in approximately 200 counties (indicated by large bubbles in the maps). Given the aggregated nature and hence limited variability
in these variables, it remains an empirical question whether they are related to any aspect of habitual behavior.

Figure 3: Geographical Distribution of Religiosity and Political Beliefs at County Level.

2.3.2 Category Characteristics

As seen from the bottom panel of Fig.1, purchase concentration varies significantly across categories. Although products sold in supermarkets share many characteristics (e.g., relatively small ticket items with high purchase frequency), there are also systematic differences that are likely to impact behavior. For example, consumer purchase behavior may differ based on whether a product fulfills functional or
experiential needs. Relative to say fashion or restaurant dining, most grocery products can broadly be classified as utilitarian necessities. Yet, purchases in categories such as ice-cream, chocolates, candies, and soda can be driven by experiential or affective motives (Holbrook and Hirschman (1982)). Choices in such “hedonic” product categories may display higher levels of variety seeking behavior compared with purchases in utilitarian categories. In Fig. 4(a), we show the difference in the concentration ratio for selected hedonic and utilitarian categories. Based on the summary measures, it does appear that behavior varies systematically across this broad classification of products. For example, hedonic categories tend to have lower purchase concentration than more utilitarian product categories.

Another approach to classifying grocery products is to utilize observed market outcomes. For instance, there are large differences across categories in factors such as assortment size, frequency of new product introductions, and the intensity of promotional activities. In the real marketplace, all these factors can impact the choice context, and hence behavior. However, variation in these observed category attributes is unlikely to be random. In particular, these attributes represent outcomes that are driven by firms acting strategically based on underlying consumer behavior. For example, in products where consumers seek variety, manufacturers have an incentive to offer higher levels of variety via larger assortments and frequent product introductions. Likewise, retailers have an incentive in promoting products that serve specific objectives (e.g., drive traffic to the store). The patterns observed in panel b of Fig. 4 do suggest that these variables are likely outcomes of strategic firm behavior. For example, the average assortment sizes and average number of new product introductions are biggest for hedonic products which also tend to be product categories with the lowest purchase concentration. The empirical approach discussed below is motivated by such causality concerns.

2.3.3 Household & Category Interactions

As seen from Fig. 4(c), we might also expect interactions between consumer and product attributes. Consider, for example, purchase frequency. This is a household/category specific variable that measures the degree to which consumers make repeated choices in a particular category. To operationalize this, we first computed the annual purchase frequency for each household within a category. For use in

Note that our classification of products is simply to illustrate the potential factors that may explain differences across categories. In the results section we provide estimates for all 100 products analyzed in the paper.
Figure 4: Summary Measures for Examples of Hedonic/Utilitarian Categories and Food/Non-Food Categories.
the analysis described in the next section we then discretized this measure by allocation into quintile categories. The expected impact of purchase frequency on the degree of purchase concentration is unclear. On the one hand, high levels of repetition of a mundane activity can lead to “automaticity” in behavior (Wood and Neal (2007)). At the same time, higher purchase frequency and intensity of use can lead to satiation (McAlister (1982), Galak, Kruger and Loewenstein (2011)) and trigger a desire for variety. Furthermore, a frequent user of a product is more likely to encounter price promotion interventions by firms.

Fig. 4(c) illustrates that the impact of higher purchase frequency (and other demographics) may vary across product types. Some of these interactions can be of interest from a theoretical perspective (e.g., impact of Age, Purchase frequency or Gender across products), while others simply serve as controls. For example, it should be no surprise that larger households display less concentration in hedonic products (to serve the taste needs of different members), while there is little need for variety in larger households for cleaning products.

3 Empirical Approach & Results

3.1 Model

The discussion above shows large variation in concentration measures across categories, and across consumers within product categories. In addition we might expect interactions between household characteristics (e.g., Age, Income) and product type. To allow for such possibilities, we estimate category-level regression models that allow the impact of consumer demographics to vary across products. The dependent variable in the regression model is the CR1 measure for each household in each of the 100 product categories. Across all households and categories, our sample comprises of approximately 4.5 million observations.\(^8\) We specify – for each category – a linear model relating demographic effects, purchase frequency, household location and observed sample length, to purchase

---

\(^8\)Note that purchase frequency is highly correlated with total usage of the product. Hence this variable proxies both frequency of purchase decisions as well heavy vs. light usage.

\(^9\)Note that these 4.5 million observations are aggregate summaries of the individual household by category by store by date decisions of which there 123 million.
where $cr_{hc}$ is the observed concentration measure (CR1-UPC) for household $h$ in category $c$. The model is written using “effects notation”. For example, $\alpha_{j[h;c]}^{\text{market}}$ is the effect for the market in which household $h$ is located and $\alpha_{j[h;c]}^{\text{age}}$ is age effect for the age category that $h$ belongs to. An alternative representation of the model would be that of a standard linear regression model where all explanatory variables enter in discrete levels. Table 1 shows all levels of all effects/variables used in the model. Note that model contains both category ($\theta$) and market effects. There are 76 markets in the data spanning 50 states (e.g., Dallas, Texas, and Syracuse, New York). The market effects in the data controls for any market idiosyncrasies (e.g., assortment, quality, number of stores etc.).

For estimation one could simply use 100 linear regressions – one for each category. Instead we use a Bayesian estimation procedure that incorporates exchangeability across categories and demographics. Rather than treat each of the 100 categories as individual entities (as standard linear regression does), this procedure pools information across categories and across levels of each demographic variable. This provides a boost in estimation precision for some of the smaller demographic/product category cells. For example, to estimate the effect on concentration in the Paper Napkins category of a household head belonging to the 18-34 Age category, the procedure uses information not only from that product/age category but also from other age categories and other product categories. Specifically, we assign a

10 Given the richness of the data, one could take a number of different approaches to study the the questions of interest. We have tried various model specifications including simple variance decompositions and logit transforming the dependent variable. One can also incorporate interactions between product category attributes (e.g., assortment size) and effects sizes in a multilevel hierarchical framework. We do not include category attributes in the actual model due to the reverse-causality reasons discussed in Section 2.3.2. Instead of incorporating category attributes directly in the data analysis, we explore the link between these and the estimated effects sizes in a post-hoc type analysis, see below. Overall, the different model specifications converge on the main findings related to demographic effects and variation across products. Details on other model specifications are available from the authors.

11 To be clear on the index notation: If household number 90 in the sample resides in market 27 and belongs to age category 3 then $j[90] = 27$ for the market effect and $j[90] = 3$ for the age effect.

### Description of Effect Levels

<table>
<thead>
<tr>
<th>Purchase Intensity</th>
<th>Bottom, 2nd, 3rd, 4th Top Quintile</th>
<th>Hypothesized Impact of Increase on Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household Composition</td>
<td>Gender: Single Male, Single Female 2 people, . . . , 5 people or more</td>
<td>Uncertain</td>
</tr>
<tr>
<td>Age, Household Head</td>
<td>18-34, 35-44, 45-54, 55-64, 65 or more.</td>
<td>Higher concentration</td>
</tr>
<tr>
<td>Household Income</td>
<td>$\leq$ $20K, 20K-$35K, $35K-$50K, $50-$70K, $70K-$100K, $100K+</td>
<td>Uncertain</td>
</tr>
<tr>
<td>Education</td>
<td>High School, College, Post-College</td>
<td>Uncertain</td>
</tr>
<tr>
<td>County Religiosity</td>
<td>1st &amp; 2nd decile, 3rd-8th decile, 9th &amp; 10th decile</td>
<td>Higher concentration</td>
</tr>
<tr>
<td>County Rep. Voting</td>
<td>1st &amp; 2nd decile, 3rd-8th decile, 9th decile &amp; 10th decile</td>
<td>Higher concentration</td>
</tr>
</tbody>
</table>

### Controls

- Category Effect: Category 1, Category 2, . . . , Category 100.
- AC Nielsen Scan Market: Market 1, Market 2, . . . , Market 76.
- Sample Length: 1 Year, 2 Years, . . . , 6 Years.

#### Table 1: Summary of Effects and Hypothesized Impact on Concentration.

Prior distribution for each of the demographic effects in the model above. For example, for the age effects we assume

\[
\alpha_{jc}^{\text{age}} \sim N(\pi_j^{\text{age}}, 1/\tau_j^{\text{age}}), \quad c = 1, \ldots, 100; \ j = 1, \ldots, 5; \quad (2)
\]

\[
\pi_j^{\text{age}} \sim N(0, 1/\psi_j^{\text{age}}), \quad j = 1, \ldots, 5. \quad (3)
\]

Here $\pi_j$ is the average age effect of age group $j$ across all categories (or main effect of age group $j$). The second distribution specifies the variation among these averages for the 5 age groups. The higher the precision parameter $\tau_j^{\text{age}}$ is, the stronger is the pooling of individual category estimates towards the common mean. All other effects in the model are specified in an analogous fashion. Complete technical details of the model and estimation procedure are described in the appendix.

#### 3.2 Results

The model described above is quite flexible in allowing category specific parameters for all household characteristics. This flexibility comes at the cost of generating an extremely large set of parameter...
estimates. For example, with 5 levels of the Age variable and 100 product categories, there are 500 age specific parameters alone. Summarizing the information for such a large number of parameters can be daunting. Our strategy is to begin by providing the main demographic effects, followed by analyzing how these effects may vary across products. Fig. provides the average estimates of the impact of the demographic variables across all 100 product categories. Recall that the dependent variable is CR1 (% of total dollar allocated to a single UPC). Hence larger (positive) values imply more concentrated purchases. In Fig. we summarize the number of categories for which the difference between the highest and lowest levels in each demographic group (e.g. youngest versus oldest, no college versus post graduate degree) is positive (implying higher purchase concentration), negative (lower purchase concentration or more variety seeking) or uncertain. Finally, Fig. shows the effect sizes (i.e., old minus young, rich minus poor, etc.) for all 100 product categories for age, income and gender.

**Demographic Main Effects**

In Fig. we find that on average across all categories purchase concentration is increasing monotonically with age. The effects sizes for the difference between the youngest and oldest – shown in Fig. & – indicate that in about two-thirds of all product categories, older consumers have more concentrated purchases than the young. This suggests that age related affective factors such as preference towards smaller social circles and spheres of interest may persist beyond the social and personal domains: Older consumers seem also to show higher brand affinity and narrower product portfolios. Note that while the average age effect across categories is small in size, there is substantial heterogeneity in effects size. For example, there are categories where the average concentration ratio for old households is 5 to 6 percentage points bigger than for young households (holding all other factors fixed). For reference we note that the average standard deviation of CR1-UPC across categories is 18%, so the age effects size max out at about 1/3 of the standard deviation of the dependent measure.

Purchase concentration is also increasing monotonically with both socioeconomic variables. Higher income and education is associated with more concentrated purchase behavior in the majority of the product categories. Again we see the importance of allowing for interactions between product categories. 

\[^{13}\text{All estimates are “effects-coded”, i.e., normalized to have mean zero within a categorial group. For example, a negative estimate for a certain age group means that that age group has lower concentration compared to the average.}\]
Figure 5: Average Effects across 100 Product Categories. Note: For each variable (e.g., Age) the graph shows the mean effect across categories for each level of the variable, ordered from lowest to highest (e.g., youngest to oldest) – see Table 1 for definition of levels. Vertical bars indicate a 95% probability interval defined by the 2.5% and 97.5% posterior quantiles.
Figure 6: Counts of Positive, Negative and Uncertain Sign Effects. Note: The plot shows the number of categories for which the probability of the corresponding effect being positive is 95% or above (i.e., an increase in the corresponding variable almost certainly implies higher concentration), the number of categories for which the probability of the corresponding effect being negative is 95% or above (i.e., an increase in the corresponding variable almost certainly implies lower concentration) and the number of categories where these probabilities are below 95% (uncertain).
Figure 7: Age, Income and Gender Effects for all 100 Product Categories.

category and effects. For example, while richer households exhibit more concentrated behavior in most categories, there are categories where the sign of the effect flips. This highlights the potential for erroneous conclusions when estimating either only an average effect of income or estimating the effect for only a few product categories.

As expected, purchase concentration is declining monotonically with household size. More interestingly, we find large differences across gender, with single males displaying significantly higher levels of concentration than single females. This effect is emphasized dramatically through the category level effect sizes of the difference between single males and females shown in Fig. 7, where we find that males have higher levels of purchase concentration than females in almost all of the product categories.

As seen in Fig. 7, higher level of religious adherence and conservative ideology (as measured by Republican voting) is associated with higher purchase concentration. Fig. 7 shows that there are five times as many products where higher religiosity is associated with higher purchase concentration compared with categories where higher religious adherence leads to more variety. The impact of Republican voting is somewhat muted with a large number of product categories showing uncertain
effects. As noted before, our measures of political voting and religiosity are derived at a county level and not observed for individual households. This in turn limits the amount of observed variation in these measures. In addition, our model specification already includes market effects that controls for location specific differences. Despite these factors, our results do suggest that higher levels of religiosity is associated with higher repetitive buying and more concentrated product purchases. This corresponds with the psychological traits associated with a conservative ideology such as preference for tradition and status quo, uncertainty avoidance, and skepticism towards new experiences (Jost et al 2003, Jost 2006).

**Hedonic versus. Utilitarian**

In Fig.4, we saw the average product concentration (across all households) is systematically higher in utilitarian categories, suggesting that consumers seek higher variety in hedonic products. In Fig.8, we provide the regression estimates for various demographics in hedonic versus. utilitarian categories. The patterns appear quite different across product groups. For example, purchase concentration is monotonically increasing with both age and income for all hedonic products, while the pattern is less consistent for utilitarian products; the estimated age effects curves are much “flatter” for utilitarian products. For household composition, we find that purchase concentration is monotonically decreasing with family size for hedonic products, while there is less variation for utilitarian products. This is to be expected since taste preferences of family members may differ in hedonic products such as potato chips, while the same brand of detergent can serve the needs for all household members. More interestingly, and as noted above, single males display consistently higher levels of purchase concentration than females across all product categories.

**Purchase Frequency**

Finally, we summarize the impact of purchase frequency in Fig.9. The top left panel shows the average effect across all categories. The level of concentration falls monotonically with the frequency of purchase: For the average category, households in the top purchase frequency quintile have concentration ratios that are on average over 7 percentage points lower than households in the lowest quintile.
Figure 8: Estimates of Impact of Age, Income, Household Composition and Gender for 10 Hedonic and 10 Utilitarian Categories.
However, as seen in top right panel, the impact of higher purchase frequency varies systematically based on the type of product. In particular, frequent purchases in hedonic product categories leads to a sharp decline in concentration, while the impact is significantly smaller for utilitarian products. For some utilitarian product categories we even see a “U”-shaped effects curve. Recall from Fig.4 that the average purchase frequency for ‘hedonic’ non-food products is in fact lower than utilitarian products. Yet the impact of higher purchase frequency is significantly more pronounced in these products. The bottom panel of Fig.4 provides some clues on why this may be the case. This panel provides the correlation between the effect size (difference in parameter estimates for the highest versus. lowest quintiles in purchase frequency) and product attributes. The impact of purchase frequency is most pronounced in product categories characterized by high assortments, frequent product introductions, and high levels of promotional activities. These firm activities have two consequences: (a) they increase the likelihood that the product currently in use by a consumer is replaced by a newer version, and (b) new product introductions are often accompanied by high levels of advertising and promotion creating a “disruptive stimuli” to break existing purchase cycles. On the other hand, utilitarian products provide a more stable decision context for purchase habits to develop – even for heavy users.

4 Discussion

This article provided a large scale empirical investigation into the buying behavior of US consumers. Using a unique panel data on shopping histories of over 100,000 households, we examined the degree to which consumers display repetitive versus varied behavior for frequently purchased household products. Our finding of extremely high levels of purchase concentration, even over long time horizons of several years, is related to the literature on habits that shows repetition as a central feature of daily lives (Neal, Wood and Quinn (2005)). Such high levels of purchase concentration in grocery products is noteworthy as the CPG industry is characterized by large assortments, frequent new product introductions, and high levels of promotional activity, often directed explicitly to induce brand switching. Repetitive buying behavior in such environments may well be a strategy to avoid the negative consequences of “choice overload” characterizing this industry (Schwartz (2004)).
Figure 9: Purchase Frequency Effects
We find that this tendency toward repetitive behavior is stronger for males versus females, and increasing with age, income, education, and conservativeness. These demographic moderators relate to several lines of research in psychology, cognitive science, and economics. For example, understanding the cognitive and behavioral concomitants to biological aging has become important with the aging global population. With prolonged lifespan and growing number of years spent in retirement, decision making of older adults has been studied in several important domains such as the choice of Medicare drug prescription plans (Hanoch, Rice, Cummings, and Wood (2009)), stock investments (Korniotis and Kumar (2011)), and age related differences in home equity, mortgage, credit card debt and auto loans (Agarwal et al. (2009)). Our paper adds to this growing literature by providing empirical evidence that aspects of aging also manifest in mundane daily choices. As another example, consider the empirical patterns observed with religiosity and republican voting. Our motivation for incorporating these variables was to proxy conservative ideology, which in turn might capture aspects of psychological traits such as openness to change and desire for novelty (Jost et al (2003), Jost (2006), Khan et al. (2012)). The empirical evidence, albeit weak, does show that conservative ideology (at least religiosity) is associated with higher brand/product affinity.

There are of course several limitations of our approach from both methodological and theoretical perspectives. First, our measure of purchase concentration relies on simple metrics such as concentration ratios. The extent to which such metrics capture theoretical constructs of variety seeking or habitual behavior is subject to debate. We are also limited by the set of demographic predictors collected by the data provider (ACNielsen). Hence we are unable to design our own survey to elicit more meaningful psychological traits using standard scales. These methodological limitations also suggest caution in the theoretical conclusions that can be drawn from such an analysis. For example, our analysis is limited to one mundane activity of grocery shopping behavior and the impact of various demographic moderators may be quite different in other contexts. Notwithstanding these limitations, the analysis presented in the paper is unique in that it captures behavior for a large sample of representative households in the real market place. Furthermore, to the best of our knowledge, the long time frame and breadth of products analyzed is unique in the literature. Our hope is that the empirical regularities provided in the paper can foster similar research investigations in other
domains.
References


