The Formation of Consumer Brand Preferences

Bart J. Bronnenberg¹,² and Jean-Pierre Dubé³,⁴

¹CentER, Tilburg University, 5000 LE Tilburg, Netherlands; email: bart.bronnenberg@tilburguniversity.edu
²Centre for Economic Policy Research, London EC1V 0DX, United Kingdom
³Booth School of Business, University of Chicago, Chicago, Illinois 60637; email: jdube@chicagobooth.edu
⁴National Bureau of Economic Research, Cambridge, Massachusetts 02138

Keywords
brand, brand capital, preferences, advertising, marketing, market structure

Abstract
Brands and brand capital have long been theorized to play an important role in the formation of the industrial market structure of consumer goods industries. We summarize several striking empirical regularities in the concentration, magnitude, and persistence of brand market shares in consumer goods categories. We then survey the theoretical and empirical literatures on the formation of brand preferences and the ways in which brand preferences contribute to our understanding of these empirical regularities. We also review the literature on how brand capital creates strategic advantages to firms that own established brands.
1. INTRODUCTION

The economics literature has long recognized the importance of consumer brands in the formation of the industrial market structure of consumer goods industries. Braithwaite (1928, p. 31) describes the rapid growth in the number of consumer brands during the early twentieth century as follows:

A growing number of goods which were formerly sold in bulk are now sold as branded goods and are advertised nationally. Tobacco, groceries, drugs, sweets, paint are a few of the more obvious examples. Such commodities can be put in distinctive packages, branded and labelled in such a way that their individuality can easily be established. With goods sold by the yard or ready-made clothes this is less easily accomplished, but even with such things producers are now attempting to make their brand or trade mark a distinctive feature of the commodity. By these means they are able to tie up their advertising with their own products, and to ensure that, when they incur advertisement costs, the increased demand will be directed, not to the commodity as a whole, but to their particular make of the commodity. Thus they are able to secure for themselves, if their advertising is successful, a partial monopoly based on reputation which it is not possible to secure when goods are sold in bulk.

Bain (1956, p. 216) argues that “the advantage to established sellers accruing from buyer preferences for their products as opposed to potential-entrant products is on the average larger and more frequent in occurrence at large values than any other barrier to entry.” In the latter part of the twentieth century, the degree of concentration in consumer goods industries grew at a much faster pace than that of other industries in the United States (Caves & Porter 1980). By the end of the twentieth century, most consumer goods industries were dominated by a small number of brands commanding the largest share of sales (Bronnenberg et al. 2007). Most striking, many of the dominant consumer brands in 1923 were still the dominant brands in their respective categories in 1983, more than half a century later (Advert. Age 1983), although the findings are predominantly in food categories.1

Surprisingly, consumers are routinely found to be unable to distinguish between leading national brands in blind taste tests (Husband & Godfrey 1934; Thumin 1962; Allison & Uhl 1964, p. 36) in spite of self-reported strong preferences for specific national brands. In the famous Pepsi Challenge promotional campaign during the 1970s, subjects exhibited a more than 50% chance of choosing Pepsi over Coca Cola (Green & Wei 2013) even though Coca Cola was the dominant cola brand and one of the world’s most valuable consumer brands at the time.2

In this article, we review the literature studying the formation of consumers’ brand preferences. Our goals consist of summarizing several stylized facts regarding brands and the industrial market structure of consumer goods industries and surveying the theoretical and empirical analyses of different mechanisms through which consumers develop a brand preference. The practice of branding spans a wide array of products, ranging from physical goods to services and even events, and a wide array of buyer contexts, ranging from households to enterprises. However, we focus most of our discussion on consumer preferences for physical consumer goods and, in particular, for nondurables sold in supermarket, drug, convenience, and mass-merchandise retail store formats.

Our review is structured as follows. In Section 2, we begin with a brief overview of consumer brands and their marketing function. We then use a new and very large database summarizing

---

1Golder (2000) shows that these results were obtained by a selective focus predominantly on food industries and that the findings weaken when extended to the database’s full set of 100 consumer goods categories.

2Coca-Cola continued to dominate consumer brands and was listed as the world’s second most valuable brand in 1993 according to Kelly (1994) and the most valuable brand from 2000 to 2013 according to Interbrand (see Gambrell 2013).
the sales and marketing of consumer brands to replicate several empirical regularities in the market structure of branded consumer goods industries. Many branded consumer goods industries exhibit substantial geographic dispersion in market shares. This dispersion is found to be persistent, in some cases over more than a century, and may tie back to the original order of entry among the current surviving brands (e.g., Bronnenberg et al. 2009). At the individual consumer level, brand tastes are also found to be highly persistent and to evolve slowly over a consumer’s lifetime.

In Section 3, we review various mechanisms that may potentially contribute to the persistence of brand tastes and of brand market structure. On the demand side, there is a long-standing literature studying brand choice inertia and brand loyalty as a psychological switching cost. However, the empirical magnitudes of estimated loyalty are typically insufficient to explain the persistence of brand market shares across decades. A separate literature has discussed the role of consumer knowledge and brand information in brand choices. Our discussion of this literature distinguishes between knowledge about search characteristics, or brand attributes that can be determined prior to purchase, and experiential characteristics, or brand attributes that are only learned after purchase and consumption. Empirically, consumers have limited information about both search and experiential characteristics. Empirical estimates of search costs associated with gathering search attributes are typically found to be high, and accordingly, consumers may only consider a small subset of the available variety when making purchases. The literature has been mixed regarding the rate at which consumers learn about experiential characteristics through purchase and consumption. The evidence suggests that information barriers could create persistent advantages to established brands.

On the supply side, firms’ incentives may not favor the endogenous supply of information in equilibrium, even when brand information disclosure is costless. In fact, in some settings, firms may endogenously seek to obfuscate information through measures that increase consumers’ search costs. These supply-side incentives would likely exacerbate consumers’ lack of information, sustaining the advantages of familiar and established brands. Firms with established brands may also be able to leverage consumers’ brand familiarity to create advantages in new product launches that extend the established brand’s name. To keep the article focused, we do not cover the extensive research on consumer brand perceptions from the psychology perspective. For a recent overview of that literature, readers are referred to, for instance, the work of Muthukrishnan (2015) or Schmitt (2012).

In Section 4, we discuss the proliferation of brands and various underlying factors that may influence the overall supply of brand variety. We also discuss how firms with established brands and brand capital may have an advantage when launching new products that extend an existing brand name into new markets. Empirically, many of the new consumer product variants launched later in the twentieth century were extensions of established brands. Section 5 concludes.

2. CONSUMER BRANDS AND MARKET STRUCTURE

In this section, we define the concept of a consumer brand and discuss several stylized facts regarding the market structure of consumer packaged goods (CPG) brands in the United States. CPG categories consist of consumable goods like food, drinks, household cleaning products, and health and beauty products. These goods are typically replaced at a regular frequency, in contrast with durables like automobiles and furniture. The CPG sector provides a useful case study of consumer brands. The industry is very large, with annual revenues reaching $8 trillion globally (Hirose et al. 2015) and $2.1 trillion in the United States (http://www.gmaonline.org/about).
2.1. What Is a Brand?

According to Wikipedia, the term brand derives from the Old Norse “brandr,” meaning “to burn.” The use of the term evolved in Middle English to refer to the practice of permanently marking cattle and livestock with a hot iron. The practice of branding consumer goods with a name or logo is, however, much older and has been part of economic exchange since the invention of papyrus, in early Egypt, and paper, during the Western Zhou Dynasty of China, needed to make posting signs (Landa 2006). The identity of the brand differentiated the product from others of the same category and enabled buyers to appraise its origins and value before buying. Preserved loaves of bread recovered from the ruins of Pompeii carried markings, made with so-called bread stamps, to signify the origin of their maker and their quality, providing a demonstration of the use of branding as early as AD 79 (Creative Bloq 2015). The branding of bread was also used in medieval Europe to enforce regulation of its quality.

In 1931, the American Marketing Association formally defined a brand as a “word, letter, group of words or letters composing a name, a design, or a combination of these which identifies the goods or services of one seller or group of sellers and/or distinguishes them from those of competitors” (Comm. Defin. 1935, p. 149). But a brand is more than the product it identifies “because it can have dimensions that differentiate it in some way from other products designed to satisfy the same need. These differences may be rational and tangible—related to product performance of the brand—or more symbolic, emotional, and intangible—related to what the brand represents” (Keller 2012, p. 5).

Brands are considered to be highly valuable assets. According to industry expert Interbrand, the top 100 most valuable global brands represented a joint value in excess of $1.7 trillion in 2015. Of these, CPG brands represented 21 of the world’s 100 most valuable brands, with a combined value of $259 billion (see Interbrand’s 2015 ranking at http://interbrand.com/best-brands/best-global-brands/2015/ranking/). The brand value or commercial value of a brand’s underlying trademark to the firm that owns the mark can be a large and important intangible asset. Although it is widely believed that brands create important barriers to entry and help sustain supranormal profits (e.g., Bain 1956; Demsetz 1982; Schmalensee 1982, 1983), the measurement of brand value is challenging in practice. The commercial value to a firm of owning the trademark to a brand is ultimately defined relative to a counterfactual (Goldfarb et al. 2008): What is the net present value of a firm’s factual equilibrium profits compared to what would have been the firm’s net present value of equilibrium profits but for the brand? 

2.2. Brands and Geography

Previous work has documented the large differences in the industrial structure of consumer goods industries across countries (Sutton 1991, Adams 2006). However, there has been a long debate in the marketing literature regarding the potential differences within a large country like the United States. The marketing literature has routinely used the term national brand to refer to widely distributed manufacturer’s or producer’s brands. The use of this term has been discouraged since at least 1935, unless the brand “is used in advertising and selling over a considerable portion of

---

3Borkovsky et al. (2017) develop an equilibrium framework with which to evaluate this counterfactual in the context of a dynamic oligopoly with firms competing on both prices and advertising investments into brand goodwill stocks. A limitation of these model-based approaches is that they infer the brand equity using a residual approach that loads all the unexplained variation in consumer brand choices (i.e., net of observed prices and marketing variables) into an all-encompassing brand intercept. Because this intercept will also account for unobserved (to the researcher) characteristics of the product, it may overstate the actual brand equity.
Table 1  Description of brand sales and shares

<table>
<thead>
<tr>
<th>Concentration</th>
<th>Observations</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equivalent units</td>
<td>14,494,228</td>
<td>470,755</td>
<td>3,740,882</td>
<td>0.01</td>
<td>827,000,000</td>
</tr>
<tr>
<td>Equivalent units share</td>
<td>14,494,228</td>
<td>0.26</td>
<td>0.25</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Dollar sales</td>
<td>14,494,228</td>
<td>47,369</td>
<td>163,987</td>
<td>0.00</td>
<td>20,800,000</td>
</tr>
<tr>
<td>Dollar sales share</td>
<td>14,494,228</td>
<td>0.27</td>
<td>0.25</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Number of brands</td>
<td>14,494,228</td>
<td>240</td>
<td>464</td>
<td>1</td>
<td>7,467</td>
</tr>
</tbody>
</table>

Equivalent units are used to ensure that brand sales in a category are measured in comparable units (e.g., milliliters for beverages, ounces for foods, counts for sachets).

the country irrespective of the nature of its sponsorship” (Comm. Defin. 1935, p. 149). However, the term national brand may even be misleading for a brand with national distribution. According to Coutant (1934, p. 125), “There is no such thing as a national market” and, although “certain groups of these cities have similar characteristics... it is fallacious to regard them altogether as a unified big city market.” Coutant (1934) relies on anecdotal evidence for regional differences in the preferences and perceptions of given brands.

Bronnenberg et al. (2007, 2011) provide a detailed description of the geographic patterns in CPG brand shares across the United States using a large, longitudinal Nielsen database spanning the top two brands in each of 31 food categories for the 39 months between 1993 and 1995. The categories are highly concentrated, with an average one-firm concentration ratio of 41% within a geographic area. For the average category studied, they also find that 92% of the pooled variation in market shares across time, markets, and brands is explained by the interaction of persistent market and brand effects, with no category under 63%. Consistent with other work (e.g., Bronnenberg & Mahajan 2001), they find spatial covariance in the market shares, but most categories exhibit geographic independence after about 500 miles. The geographic variance has several striking features. Bronnenberg et al. (2007) find considerable regional dispersion in the identity of the brand leader within a category. They also find that the market shares and brand leadership positions are persistent over the three-year span of the data.

We now extend the findings of Bronnenberg et al. (2007) to the entire US CPG industry. We use the Nielsen-Booth extracts of the Nielsen Retail Measurement System (RMS) data available from the Kilts Center for Marketing at the University of Chicago Booth School of Business. The data span about 35,000 stores, with formats including supermarket, drug, mass, and convenience, located across 76 Scantrack markets.4 The data track weekly retail sales and prices at the level of the individual Universal Product Code and span 1,088 product categories (designated by Nielsen’s module codes) from 2006 to 2014. These categories span 10 broad departments: alcoholic beverages, dairy, deli, dry grocery, fresh produce, frozen foods, general merchandise, health and beauty, nonfood grocery, and packaged meat. Although these categories exclude nongrocery consumer goods such as electronics and apparel, the RMS database collectively represents almost US$2 trillion in sales over the markets and time periods covered. We collapse the data to the brand, market, and month level, yielding 231,580 unique brands and 108 months. For each of the categories, we then retain the top two brands based on overall national dollar sales.5 Table 1 summarizes the data used for our analysis.

4A Nielsen Scantrack definition corresponds approximately to a large metropolitan area.
5For 19 of these categories, there is only one brand available in a given geographic market. Therefore, Table 3 reports on 1,069 categories.
Table 2  Category concentration (by category and market)

<table>
<thead>
<tr>
<th>Concentration</th>
<th>Observations</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>80,636</td>
<td>0.46</td>
<td>0.24</td>
<td>0.03</td>
<td>1</td>
</tr>
<tr>
<td>C2</td>
<td>80,636</td>
<td>0.65</td>
<td>0.24</td>
<td>0.06</td>
<td>1</td>
</tr>
<tr>
<td>C3</td>
<td>80,636</td>
<td>0.74</td>
<td>0.22</td>
<td>0.08</td>
<td>1</td>
</tr>
<tr>
<td>C4</td>
<td>80,636</td>
<td>0.79</td>
<td>0.20</td>
<td>0.10</td>
<td>1</td>
</tr>
</tbody>
</table>

C1, C2, C3, and C4 represent the one-, two-, three-, and four-firm concentration ratios, respectively. Sales within a category are measured using equivalent units (e.g., milliliters for beverages, ounces for foods, counts for sachets). Concentration is measured using the sample equivalent unit sales for a category (i.e., total across time within a market).

In general, the sales in these CPG product categories are very concentrated, as summarized in Table 2. On average across categories and Scantrack markets, the top brand commands 46% of equivalent unit sales, which is consistent with the analysis of food categories by Bronnenberg et al. (2011). The average four-firm concentration ratio is 79%. The 25th percentile of one-firm concentration ratios is almost 30%, and the 25th percentile of four-firm concentration ratios is 66%. Although this is not reported in the tables, there is little systematic difference in the concentration patterns across the 10 departments. Only the alcoholic beverages department stands out with an average one- and four-firm concentration ratio of only 40% and 71%, respectively.

In Table 3, we report the results of an analysis of variance that pools the top two national brands in each of 1,069 categories across the 108 months and 76 geographic markets. We decompose the brand shares into market, brand, and month fixed effects, along with an interaction between market and brand. Table 3 reports the $R^2$ for each of these components. Our findings broadly replicate the analysis of food categories by Bronnenberg et al. (2007). We do find larger time effects than Bronnenberg et al. (2007), explaining, on average, 11% of the variance in a category’s shares. However, market and brand effects are much larger on average than these time effects. Moreover, the interaction between market and brand explains 68% of the variance on average and is much larger than the sum of the individual brand and market components.

In Table 4, we report the average $R^2$ values across all the categories within a department. With the exception of the fresh produce department, we observe that the main effect of brands is smaller and that of markets is larger for the food and beverage department than for nonfood departments like general merchandise and health and beauty. The relatively small role of brand effects confirms the contentious notion of a national brand, at least in food categories. Interestingly, the nonfood categories also have much higher time effects that explain almost 20% of the share variance. In contrast, time explains less than 10% of the share variation in the food categories, except for fresh produce. These findings suggest that the persistence in market structures and the differences in market shares of individual brands across regions of the United States may be more pronounced in food categories than in nonfood categories.

Table 3  Analysis of variance of top two consumer packaged goods brand shares by category

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market</td>
<td>1,069</td>
<td>0.18</td>
<td>0.17</td>
<td>0.00</td>
<td>0.95</td>
</tr>
<tr>
<td>Brand</td>
<td>1,069</td>
<td>0.33</td>
<td>0.32</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Month</td>
<td>1,069</td>
<td>0.11</td>
<td>0.14</td>
<td>0.00</td>
<td>0.82</td>
</tr>
<tr>
<td>Market $\times$ brand</td>
<td>1,069</td>
<td>0.68</td>
<td>0.26</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

$R^2$ values correspond to a category-specific analysis.
Pooling all the top-selling national brands in each of the categories, the average geographic range in market shares is 41 percentage points. Thus, a brand commanding a 50% share in one US city might only have a 9% share in another US city. If we focus on food departments, the average range is well over 50%. We also observe variation in the identity of the market leader. The top-selling national brand dominates, on average, 82% of the geographic markets.

### 2.3. Persistence in Market Shares

The small time effects documented by Bronnenberg et al. (2007) and generalized in Section 2.2 are suggestive of persistence in brand shares within a market. Previous work has documented persistence in market shares over horizons typically spanning less than a decade. In a meta-analysis of 44 studies that apply univariate time-series models to 213 brand share series, Dekimpe & Hanssens (1995) find that almost all of the studied series exhibit market share stationarity. A growing body of evidence suggests that the persistence in brand market shares spans much more than a decade. A survey-based study in the business press found that, for 25 large consumer products categories, 20 of the top brands in 1923 were still dominant in 1983, more than half a century later (Advert. Age 1983). All 25 brands were still ranked among the top five in the category in 1983. Golder (2000) extends this analysis to 100 categories and uses more reliable 1997 market share data. He finds that only 23% of the dominant firms in 1923 remained dominant in 1997, although nearly 50% remained in the top five. The national data used in these studies might mask some of the important regional differences in market shares and leadership positions summarized in Section 2.2.

In contrast to this previous work, Bronnenberg et al. (2012) account for geographic differences in market shares across US regions and find much stronger evidence of longer-term persistence in market shares. They study the long-term persistence in CPG market shares across geographic areas in the United States using Nielsen’s Homescan Panel. They combine these data with the Consolidated Consumer Analysis database compiled by aggregating consumer surveys conducted by newspapers nationwide from 1948 to 1968. For 27 CPG product categories, the authors observe the market shares for the top two brands by geographic region for the periods 2006–2008 and 1948–1968. A pooled regression of the historic share level on the current share level fails to reject an intercept of zero and slope of one. The magnitudes of the point estimates lead the authors to conclude “that the best predictor of a past purchase share given the data we observe is the present purchase share” (Bronnenberg et al. 2012, p. 2492).
The empirical literature on pioneering advantage also documents persistence in brand market shares, albeit with a focus on the persistence of dominance dating back to the initial entrants into a new product category. Early work found strong evidence of a persistent market share advantage for first entrants (Robinson & Fornell 1985, Urban et al. 1986, Lambkin 1988, Robinson 1988, Parry & Bass 1990, Kerin et al. 1992, Brown & Lattin 1994). A similar persistence of dominance dating back to 1933 is reported for consumer brands in the United Kingdom (Keller 2012, p. 21). Kalyanaram et al. (1995) provide a thorough survey of the literature, along with empirical generalizations regarding the negative correlation between historic order of entry and current market share. The evidence for the effect of pioneering advantage on market shares has been under debate. Some of the debate has revolved around the accuracy of the definition of a market pioneer (Golder & Tellis 1993). The key consistent finding is the persistence in market shares and the advantages to early movers (even if not for the first entrant) that survive in the long term.

On the econometric side, a potentially serious concern with the interpretation of these sources of persistence regards the econometric identification of an early mover effect. The literature has typically relied on pure time-series analysis. These data cannot empirically distinguish between state dependence (the early mover effect) and the plausible heterogeneity between firms and their brands. For instance, if market pioneers systematically exhibit greater managerial skill or launch better brands or products this heterogeneity could spuriously identify a pioneering advantage effect.

Bronnenberg et al. (2009) resolve this econometric problem by pooling within-market time-series data for the 50 largest Nielsen Scantrack markets. Like Golder & Tellis (1993), they obtain their historic rollout information through an extensive search of historic documents and archives. They assemble six CPG category case studies for which they observe 39 months of monthly market share data for the leading national brands in each of the 50 markets. These data are matched with the exact year during which each of these leading brands entered each of the 50 markets. Their key identifying assumption is that the timing of entry into each market is exogenous. Most of the brands studied originated during the late 1800s, long before marketing and distributional technology existed to coordinate a national launch.

For the six categories studied, Bronnenberg et al. (2009) find that the historic order of entry (often a century before their study) amongst survivor brands in a geographic market predicts the current rank order of market shares in that market. These findings are visualized in Figure 1, which plots the geographic distribution of brand shares for the ground coffee category across US cities. Historic order of entry in a geographic market also predicts the current rank order of brands’ perceived quality levels, as measured by the marketing and communications company Young & Rubicam’s 2004 Brand Asset Valuator survey. For 49 of the top two national brands in 34 CPG categories, Bronnenberg et al. (2009) are able to identify the city of origin (although not the complete roll-out history). They find a strong correlation between a brand’s share in a given market and the Euclidean distance to its market of origin. In particular, a brand’s share is, on average, 20 percentage points higher in the market of origin than in a distant market more than 2,500 miles away. This finding is consistent with the historic diffusion of brands launched in the late nineteenth and early twentieth centuries, with entry in more distant markets occurring relatively later.

6One exception is provided by Brown & Lattin (1994), who use a cross section of markets with no within-market variation. Unfortunately, in their data, the first entrant is the same in 37 of the 40 studied markets.

7For instance, Bartels (1976) and Tedlow (1990) provide detailed discussions of how entrepreneurs in the late nineteenth century gradually rolled out their new consumer brands across the United States.
In summary, the current market structure of CPG brands in terms of their market shares is highly persistent over a horizon spanning at least half a century. Moreover, the historic roll-out patterns for brands seem to be associated with the observed share levels among surviving market leaders today, even in categories for which roll-out began during the late nineteenth century.

3. PERSISTENCE IN BRAND PREFERENCES

3.1. Overview

Many factors might potentially explain the persistent differences in the market shares across markets and the order-of-entry effects discussed in Section 2.3. On the supply side, Bronnenberg et al. (2009) explore a number of potential mechanisms, including location of plants and even potential relationships with retailers. However, none of these factors were found to explain much of the geographic component in brand shares. We now explore theories of the formation of consumer brand preferences and empirical evidence that might help researchers understand the underlying mechanisms contributing to the observed persistence.

On the demand side, consumer psychologists have studied how a consumer develops a brand preference through positive associations between the brand and the consumption benefits of the underlying product. Such associative learning could arise, for instance, through signals whereby the consumer learns that the brand predicts a positive consumption experience. Alternatively, under evaluative conditioning, the consumer forms a positive preference for a brand through repeated co-occurrences with positive stimuli, like good mood (affect) or a popular celebrity. In the same vein, a consumer may learn about a brand through their memory of positive experiences with similar products. We refer the interested reader to Van Osselaer (2008) for a survey of the consumer psychology literature on consumer learning processes.

Most of the literature surveyed in this review treats the consumer’s perceived value of a brand as a product characteristic. As the perception of the brand changes, the consumer’s expected utility for the branded product changes accordingly. In contrast, Becker & Murphy (1993) theorize that brands and marketing activities related to branding constitute consumption goods that are
complementary to the consumption of the physical good being branded. In this regard, the consumption benefit of the brand enhances the consumption benefit of the good and vice versa. A limited set of empirical studies has attempted to test this paradigm directly. To the best of our knowledge, Kamenica et al. (2013) provide the only direct evidence. They conduct randomized clinical trials to test whether the treatment effect of direct-to-consumer advertising has a causal effect on a subject’s physiological reaction to a drug. In particular, a branded antihistamine was found to be more effective when subjects were exposed to that brand’s advertising as opposed to a competitor brand’s advertising. This complementarity offers one explanation for the inconsistencies in laboratory studies of consumer preferences in conditions where brands are removed (blind taste tests) versus conditions where brand labels are present (e.g., Allison & Uhl 1964).

To formalize our discussion of the empirical literature, consider the following model of the conditional indirect utility for a consumer choosing brand \(j\) from among \(J\) alternatives at time \(t\):

\[
U_{jt} = \mu_j(X_t; \alpha) + F_j(H_t; \gamma), \quad j = 1, \ldots, J,
\]

where \(\mu_j(X_t; \alpha)\) is the relevant function to brand \(j\) of the products’ characteristics, including current marketing conditions like prices, \(X_t\). \(F_j(H_t; \gamma)\) is the relevant function to brand \(j\) of a household’s shopping history (or past brand experiences), summarized in the state vector \(H_t\). Finally, \(\Theta = (\alpha, \gamma)'\) are structural parameters (tastes). In the discussion that follows, it is helpful to distinguish between the role of current marketing and product characteristics, \(\mu_j(X_t)\), and the consumer’s purchase history and past experiences, \(F_j(H_t)\). In extremis, one might assume that all that matters are the initial brand influences a consumer experiences during childhood, in which case \(\mu_j(X_t) = 0\). Alternatively, one might assume that all that matters are the current marketing conditions, in which case \(F_j(H_t) = 0\).

Theorists have analyzed various mechanisms through which current willingness to pay for brands reflects past brand experiences. The function \(F_j(H_t)\) acts like a brand capital stock that could stem from various sources, including habit formation (e.g., Pollak 1970, Becker & Murphy 1988), switching costs (e.g., Farrell & Klemperer 2007), advertising and branding goodwill (e.g., Schmalensee 1983, Doraszelski & Markovich 2007), evolving quality beliefs through learning (Schmalensee 1982), and peer influence (e.g., Ellison & Fudenberg 1995). The function \(\mu_j(X_t)\) captures current marketing conditions that could include prices, in-store promotions like discounts and displays, advertising, and even shelf space allocation. Our discussion below focuses mainly on sources of brand capital stock.

### 3.2. Persistent Brand Preferences Over a Consumer’s Lifetime

Lenin is credited with the quotation, “Give me four years to teach the children and the seed I have sown will never be uprooted” (quoted in Albig 1939, p. 132). This strong view is shared by some of the literature on consumer brand preferences: “If Tide laundry detergent is the family favorite, this preference is easily passed on to the next generation. The same can be said for brands

---

8An indirect test of Becker & Murphy’s (1993) theory exploits the Slutsky symmetry condition by testing whether a shift in demand for the physical good increases the consumption of the brand’s advertising. Tuchman et al. (2015) use data that match household-level time-shifted television viewing on digital video recorders with in-store shopping behavior. They find that in-store promotions that increase a household’s consumption of a brand cause an increase in the household’s propensity to watch (i.e., not skip) that same brand’s commercials.

9The findings indicating an effect of shelf space on consumer attention to brand and purchase have been mixed, with small effects documented in the field experiments of Drèze et al. (1994) and much larger effects in the eye-tracking experiments of Chandon et al. (2009). Similarly, Ataman et al. (2008) find that the success of a newly launched brand in an existing CPG category is highly correlated with the breadth of its distribution.
of toothpaste, running shoes, golf clubs, preferred restaurants, and favorite stores” (Berkman et al. 1997, pp. 422–23). Returning to our model in Equation 1, under this view, a consumer’s preferences are entirely driven by $F_j(H_0)$, where $H_0$ represents their initial experiences in life, including how their parents shopped.

Guest (1955) documents early evidence that brand preferences develop early in a child’s life and persist into adulthood. He conducted a brand preference survey for 813 school children in early 1941 and was able to repeat the survey for 20% of the original respondents 12 years later, in the spring of 1953. Across 16 product categories, he finds a 39% rate of agreement between the stated preferred brands in the two waves. A subsequent survey in 1961 of the same 20% of the original respondents generated a rate of agreement of 26% (Guest 1964). The evidence suggests that events early in a consumer’s lifetime may have lasting effects. However, “obviously, one cannot simply assume that what is learned during childhood somehow ‘transfers intact’ to adult life” (Ward 1974, p. 4).

The literature on consumer socialization research has studied mechanisms through which adult brand preferences are formed early in life (Moschis & Moore 1979), especially through intergenerational transfer and parental influence (Ward 1974, Moschis 1985, Carlson et al. 1990, Childers & Rao 1992, Moore et al. 2002) and peer influence (Reisman & Roseborough 1955, Peter & Olson 1996). Anderson et al. (2015) document a strong correlation in the automobile brand preferences of parents and their adult children. Sudhir & Tewari (2015) use a 20-year survey panel of individual Chinese consumers and find that growing up in a region that experienced rapid economic growth during one’s adolescence is correlated with consumption of nontraditional aspirational goods and brands during adulthood. Returning to our model in Equation 1, this evidence suggests that a consumer’s preferences at a given moment in time will reflect a combination of their current influences, $\mu_j(X_t)$, including family, peers, and local marketing, as well as their past brand purchases and experiences, $F_j(H_t)$.

With the exception of the surveys used by, e.g., Guest (1955) and Sudhir & Tewari (2015), researchers typically do not have access to individual-level brand choice histories that span a consumer’s lifetime. Bronnenberg et al. (2012) exploit the geographic heterogeneity in brand popularity discussed in Section 2.2. They match the Nielsen-Kilts Homescan CPG purchase panel data with a Panelviews survey of the same households. From the survey, they determine the primary shopper’s state of birth, state of current residence, age at which they moved to another state, and current age for nearly 40,000 US households. They then test whether and the extent to which a household’s current shopping behavior is associated with the primary shopper’s state of birth. Variation in the current age of a consumer and the historic timing of their move creates variation in how much time a consumer spent in their birth state and current state. Accordingly, Bronnenberg et al. (2012) observe how quickly a consumer’s brand preferences converge toward those of the typical nonmigrant in the current state of residence. The authors study an individual’s purchase shares for the two top brands across 238 product categories.

Bronnenberg et al. (2012) document two striking regularities. First, nearly 60% of the difference in brand shares between the state of origin and current state of residence appears to be eliminated almost immediately after moving. Evidence is provided both within and between households. Second, the remaining 40% of the gap closes very slowly, requiring more than 20 years to close even half the gap. Even migrants who moved before the age of five exhibit a small persistent gap.

10 These aspirational goods consist primarily of Western brands consumed socially.

11 Mela et al. (1997) also find that consumer brand preferences are responsive to shorter-term changes in point-of-purchase marketing efforts like discounts.
The authors conclude that, “since the stock of past experiences has remained constant across the move, while the supply-side environment has changed, we infer that approximately 40 percent of the geographic variation in market shares is attributable to persistent brand preferences, with the rest driven by contemporaneous supply-side variables” (Bronnenberg et al. 2012, p. 2473).

Returning to our model in Equation 1, these findings imply that approximately 40% of consumers’ expected conditional indirect utility derives from \( F_j(H_t) \) and 60% from \( \mu_j(X_t) \). These findings are consistent with the long-term habit formation for food studied by Atkin (2013).

Collapsing their analysis by age cohorts, Bronnenberg et al. (2012, p. 2481) find that “migrants who moved during childhood have relative shares close to those of nonmigrants in their current states, while those who move later look closer to nonmigrants in their birth states.” This evidence is consistent with a theory of brand capital stock whereby older migrants who have accumulated more brand capital would exhibit more inertia. The authors use the data to estimate a simple model of demand with habit formation occurring through a brand capital stock (Pollak 1970, Becker & Murphy 1988). Consumer choices are influenced by current marketing conditions and a brand capital stock reflecting historic brand consumption experiences.

To interpret these results, Bronnenberg et al. (2012) use a simple model of habit formation in which myopic consumers’ brand choices each period depend on both contemporaneous marketing and their stock of past consumption experiences. The key identifying assumptions are that “migration status is orthogonal to stable determinants of brand preferences” and that “a brand’s past market share in a given market is equal in expectation to the share today” (Bronnenberg et al. 2012, p. 2474). The key result is that the impact of past consumption experiences depreciates at a rate of 2.5% per year. Several policy simulations indicate that this persistence in preferences is large enough to rationalize some of the persistent order-of-entry effects discussed in Section 2.3.

### 3.3. Brand Choice Inertia, Switching Costs, and Loyalty

Switching costs represent another highly studied source of persistence in consumer demand. Klemperer (2005, p. 125) explains: “A product exhibits classic switching costs if a buyer will purchase it repeatedly and find it costly to switch from one seller to another.” It is unlikely that consumers literally face a financial switching cost from switching between typical CPG brands, unlike, for instance, the software costs associated with changing one’s operating system on a computer. However, consumer psychologists have studied psychological switching costs as a source of repeat-purchase behavior [see, for instance, the work of Mittelstaedt (1969) and the survey of psychological theories of repeat buying by Muthukrishnan (2015)]. In the CPG industry, analysis of household purchase histories routinely finds that brand switching rates are, on average, quite low (Dekimpe et al. 1997) and that, “in a typical grocery product category, where availability is often in the hundreds or even thousands of options, households on average spend about half of their total category expenditure over one year on one single product” (Hansen & Singh 2015, p. 1). This finding is particularly striking in CPG categories, where consumers often cannot identify

---

12The evidence is, however, also consistent with the literature on the aging consumer. Studies in (neuro)psychology on the relation between age and consumption report that cognition and memory decline with age, with working memory being more affected by age than long-term memory (Carpenter & Yoon 2011). Processing of (new) information also declines with age and leads to aging consumers tending to choose from a smaller number of considered options (John & Cole 1986) and engaging in less product comparisons (Lambert-Pandraud et al. 2005). These factors contribute to a decline in the flexibility of the purchasing patterns of the aging consumer (see also Drolet et al. 2010).

13These assumptions are tested using within-household data before and after moving and historic market share data from the 1940s to the 1960s.
their preferred CPG brands in blind taste tests (e.g., Thumin 1962, Allison & Uhl 1964). Similar patterns of high repeat-purchase behavior have been documented in other settings where goods are relatively undifferentiated other than by the brand names: health plans (Handel 2013), auto insurance (Honka 2014), broadband service (Liu et al. 2010), mortgage services at retail banks (Allen et al. 2016), and mobile phone services.15

Interestingly, Klemperer (2005) associates the origins of the literature on switching costs with Selten’s (1965) model of demand inertia. One of the oldest and most widely studied empirical topics in the marketing literature is the analysis of brand loyalty as a source of demand inertia. In the early 1950s, Advertising Age sponsored a research relationship between the Chicago Tribune and the University of Chicago to study brand loyalty patterns using a novel, household-level diary purchase panel (Brown 1953). Loyalty was defined based on patterns of inertia (or repeat buying) in the observed household brand choice sequences. The initial findings revealed a large number of households that exhibited long spells of repeat buying of the same brand over time. Nevertheless, many households routinely switched between brands over time, in contrast with the assumption of an infinite switching cost often assumed in the theoretical microeconomics literature (e.g., Beggs & Klemperer 1992) and in some of the psychology literature (e.g., Berkman et al. 1997, pp. 422–23). A brand choice theory with infinite switching costs is consistent with the view, discussed in Section 3.2, that brand preferences are fully developed during childhood and reflect parental brand preferences.

With the advent of CPG shopping diary panels, a literature emerged studying patterns of inertia in brand choices and testing the order of the stochastic brand choice process. The basic idea consisted of testing whether an observed household choice sequence looked like a zero-order Markov process rather than some other higher-order Markov process. Rejection of the null hypothesis of zero-order choice behavior was interpreted as evidence of brand loyalty. Two classic studies are those of Frank (1962) and Massy (1966), who apply a nonparametric binomial runs test to individual choice histories to determine whether observed choices are zero order. Bass et al. (1984) provide an excellent survey of the subsequent literature and generalizations of the findings. The findings in the literature are highly mixed regarding the fraction of households for which one can reject the null hypothesis of zero-order behavior. Many studies interpret a failure to reject the null as evidence against brand loyalty. However, the tests themselves are known to have relatively low statistical power, especially for choice histories with fewer than 20 observations (Massy et al. 1970, Bass et al. 1984). Because most choice histories are relatively short (typically one or two years), it is unlikely that a pure within-household analysis will provide reliable evidence for loyalty and inertia. In addition, these tests are binary and cannot correctly account for the choice among many brands. Finally, the tests fail to control for point-of-purchase causal factors like prices and other marketing variables.

A more recent literature has adopted a more structural approach that models the multinomial nature of brand choices with controls for prices and other marketing variables (Jeuland 1979, Guadagni & Little 1983, Jones & Landwehr 1988, Roy et al. 1996, Keane 1997, Seetharaman et al. 1999, Shum 2004, Osborne 2008, Dubé et al. 2010). The added structure of the choice model improves statistical power, and many authors have documented statistically and economically significant degrees of inertia. However, the recent literature has debated the interpretation of

14In December 2004, 95.7% of the digital subscriber line (DSL) market share was held by phone companies even though many third-party DSL ISPs existed, some with award-winning service (see, for instance, https://ecfsapi.fcc.gov/file/6518051597.pdf).

15Jackson (2013) reports that switching rates for US mobile phone services are 1.1% at AT&T and 0.91% at Verizon.

www.annualreviews.org • Formation of Consumer Brand Preferences
the inertia. One interpretation, termed structural state dependence (Heckman 1981), involves a causal link between past and current brand choices, i.e., true brand loyalty. Alternatively, inertia could arise due to serially correlated unobserved sources of between-consumer heterogeneity, a phenomenon termed spurious state dependence (Heckman 1981) and recognized as a confound to the empirical finding of brand loyalty since at least the work of Massy (1966).

Consider the following reformulation of the model of the conditional indirect utility for a consumer $h$ choosing brand $j$ from among $J$ alternatives on trip $t$:

$$u_{jt}^h = \lambda_j^h + \alpha^h X_{jt} + \gamma^h I\{s_{ht} = j\} + \epsilon_{jt}^h, \quad j = 1, \ldots, J,$$

where $X_{jt}$ are the marketing conditions for brand $j$ (e.g., its price); $s_{ht} \in \{1, \ldots, J\}$ is consumer $h$’s loyalty state, defined as the previously purchased brand; and $\Theta^h = (\lambda_1^h, \ldots, \lambda_J^h, \alpha^h, \gamma^h)'$ are structural parameters to be estimated. The term $\epsilon_{jt}^h$ is a random utility disturbance for each shopping trip. For instance, if we assume $\epsilon_{jt}^h$ are independent and identically distributed draws from a Type I extreme value distribution, we have the familiar conditional logit model of demand. Equation 2 is a very simple but commonly used specification that allows the observed choice process to be either first-order or zero-order Markov. A test for zero-order choice behavior would consist of the null

$$H_0: E(\gamma^h) = 0,$$

where $E(\gamma^h) > 0$ implies inertia and $E(\gamma^h) < 0$ implies variety seeking. In practice, the researcher could specify a more general specification that both relaxes the linearity and allows for a higher-order choice process that depends on a more general function of the choice history rather than focusing on the previous choice.\(^\text{16}\)

In practice, unless the researcher adequately controls for heterogeneity (observed and unobserved) between consumers, a rejection of the null in Equation 3 could merely constitute spurious state dependence. Keane (1997) and Dubé et al. (2010) apply rigorous controls for persistent, unobserved heterogeneity in $\Theta^h$ and serial dependence in $\epsilon_{jt}^h$.\(^\text{17}\) The authors find statistically and economically significant levels of structural state dependence, although the magnitudes decline substantially after controlling for heterogeneity. Dubé et al. (2010) estimate median levels of $\gamma$ that are equivalent to 21% of the magnitude of average prices in the refrigerated orange juice category. These magnitudes more than double in analogous models without controls for unobserved heterogeneity. Nevertheless, the empirical magnitudes of switching costs are sufficiently small to rule out perfect lock-in (or infinite switching costs), which is often assumed in the theoretical literature.

The results suggest that many households form short-term brand buying habits that would create persistence in their observed choices over time. Dubé et al. (2010) test several potential underlying mechanisms for this state dependence, including loyalty, price search, and learning. They conclude that loyalty is most consistent with their data and that $\gamma$ can be interpreted as a deep structural parameter.

The loyalty term $\gamma$ in Equation 2 is analogous to the typical microeconomic formulations of switching costs (Klemperer 1987, Beggs & Klemperer 1992). Dubé et al. (2009) explore the implications of their estimated switching costs for consumer brand prices and find that, contrary

\(^\text{16}\)Keane (1997) and Guadagni & Little (1983) use a non-Markov specification in which loyalty is modeled as an exponentially smoothed weighted average of a consumer’s entire observed choice history.

\(^\text{17}\)When the researcher does not observe consumers’ initial choices, an initial conditions bias can also arise from the endogeneity in consumers’ initial observed (by the researcher) states. Handel (2013) avoids this problem in his analysis of health plan choices. He exploits an intervention by an employer that changed the set of available health plans and forced employees to make a new choice from this changed menu.
to the conventional wisdom (e.g., Farrell & Klemperer 2007), the presence of such switching costs can toughen price competition and lead to lower equilibrium price levels.

The evidence for brand loyalty is nevertheless insufficient to understand the decades-long patterns of brand share persistence discussed in Section 2.3. Consider again the model of conditional indirect utility in Equation 2. The model includes brand-specific intercepts, \( \lambda_j \), to control for persistent intrinsic preferences for the brand itself. Even after controlling for inertia, \( \gamma \), Keane (1997) and Dubé et al. (2010) find very large brand intercepts. Keane (1997) and Seetharaman (2004) find large brand-specific heterogeneity components even after allowing for more sophisticated non-Markovian choice processes in which loyalty depends on a consumer’s entire observed choice history. Because most brand choice studies use approximately one or two years of observed choices, the brand intercepts, \( \lambda_j \), are typically treated as nuisance parameters in this literature. However, they represent persistent sources of brand preferences that are not accounted for by loyalty (or brand buying habits) over the typical time horizons studied. Moreover, these persistent brand tastes, \( \lambda_j \), are much more predictive of brand choices than the inertia created through the loyalty term. In economic terms, the psychological switching costs create short-term inertia in brand choices but are much smaller than the persistent source of brand tastes in the intercepts of most brand choice models of demand. Using the posterior mean hyperparameter values from the five-component mixture estimates of demand for margarine used by Dubé et al. (2010), the importance weights for loyalty, price, and brand are 6.4%, 53.6%, and 40%, respectively.18 Part of \( \lambda_j \) could reflect unobserved (by the researcher) product characteristics in \( \mu_j(X_i) \), but \( \lambda_j \) could also include some of the slowly evolving components of tastes due to lifetime brand purchase and consumption experiences, \( F_j(H_t) \). In the latter scenario, brand habits may still arise, but over a much longer time horizon than the ones typically studied in the brand loyalty literature.

### 3.4. Advertising and Brand Goodwill

The literature on persuasive advertising also has a tradition of modeling the persistent effects of past advertising as a goodwill stock (see Bagwell 2007 for a survey). Such advertising investments have long been considered to constitute important barriers to entry (Bain 1956, Sutton 1991). To the extent that brand preference reflects past advertising, the finding of a stock of persistent brand capital, \( F_j(H) \), could be picking up past exposure to advertising and branding. An extensive empirical literature has attempted to estimate the long-run effects of advertising on demand using aggregate data (see, e.g., Leone 1995). A detailed discussion about the econometric challenges of identifying consistent advertising effects is beyond the scope of this review.

Under sequential entry, when advertising investments in the goodwill stock are fixed and sunk, an early entrant faces investment incentives to deter future entry. In some cases, an early entrant may invest in a large goodwill stock to soften the effect of future competition (Schmalensee 1983, Fudenberg & Tirole 1984). A related literature has studied the entry-deterring effects of large fixed investments in a sustainable product quality advantage to deter later entrants or to soften future competition postentry (Lane 1980, Moorthy 1988, Sutton 1991, Lehmann-Grube 1997,

---

18 Following the convention in the literature on conjoint analysis, an importance weight approximately describes the percentage of utility deriving from a given component. The model in Equation 2 has three components to utility: brand, marketing variables, and loyalty with respective part-worths (or marginal utilities)

\[
PW_{brand}(brand = j) = \lambda_j - \min(0, (\lambda_j - 0.5)I_{j = 1}),
\]

\[
PW_{marketing}(X_{jt} = x) = a[x - \min(c)],\]

\[
PW_{loyalty}(\gamma) = \gamma.
\]

We can then assign an importance weight to each of these components, scaled to sum to one, as follows:

\[
IW_{brand} = \frac{\max(PW_{brand})}{\max(PW_{brand}) + \max(PW_{marketing}) + \max(PW_{loyalty})},
\]

\[
IW_{loyalty} = \frac{\max(PW_{loyalty})}{\max(PW_{brand}) + \max(PW_{marketing}) + \max(PW_{loyalty})}.
\]

www.annualreviews.org • Formation of Consumer Brand Preferences 367
Bronnenberg (2008). More recently, Doraszelski & Markovich (2007) show how initial advantages to one firm can persist in the long term using a more realistic model with an infinite horizon and investment by firms in competing, depreciating advertising stocks. These predictions for early mover effects in an advertising game with sunk investments in a goodwill stock are consistent with the empirical findings of an order-of-entry effect on brand market shares, as discussed in Section 2.3. Clark (2007) exploits an interesting natural experiment in which advertising to children was banned in Quebec but not in other Canadian provinces. He finds that the market share advantage of older, established children’s breakfast cereal brands relative to newer brands is larger in Quebec than in other provinces that permit newer brands to advertise. A similar brand share differential in Quebec does not exist for established adult brands that are not subject to the advertising ban.

3.5. Brands and Consumer Information

An important historical motivation for the emergence of brands and trademarks is that they “assured the buyer or trader of the quality of the merchandise” (Landa 2006, p. XX) in a time in which information about the maker or seller of the merchandise was often impossible to obtain directly. Resolving this consumer uncertainty about product quality has often been cited as one of the main roles of brands, and economists have long speculated about the consequences of consumer reliance on brands as a proxy for objective product quality. As discussed by Bronnenberg et al. (2015, p. 1670),

“Braithwaite (1928) writes that advertisements ‘exaggerate the uses and merits’ of national brands, citing aspirin and soap flakes as examples. Simons (1948) advocates government regulation of advertising to help mitigate ‘the uninformed consumer’s rational disposition to ‘play safe’ by buying recognized, national brands’ (1948, 247). Scherer (1970) discusses premium prices for national-brand drugs and bleach, and writes that ‘it is hard to avoid concluding that if the housewife-consumer were informed about the merits of alternative products by some medium more objective than advertising and other image-enhancing devices, her readiness to pay price premiums as large as those observed here would be attenuated’ (1970, 329–332).”

Nelson (1970, p. 311) concludes that “limitations of consumer information about quality have profound effects upon the market structure of consumer goods.”

3.5.1. Consumer knowledge and shopping expertise. We begin by describing consumers’ knowledge about products and the association of this knowledge with market structure. In some instances, a brand may indeed convey reliable information about a product. For instance, Marquardt & McGann (1975) find a positive correlation between advertising and Consumer Reports product ratings. However, striking levels of price dispersion have been documented in markets with relatively homogeneous goods differentiated mainly by brand names that command a high price premium: mutual funds (Hortacsu & Syverson 2004), fund managers (Hastings et al. 2013), online book vendors (Brynjolfsson & Smith 2000), twin automobiles (Sullivan 1998), health insurance (Handel 2013), and pharmaceuticals (Hurwitz & Caves 1988). Particularly striking is the prevalence of consumer spending on nationally branded goods when a cheaper and comparable store brand (or private label) is available. In an analysis of nearly 38,000 stores across over 100 chains, Bronnenberg et al. (2015, p. 1670) “find that consumers would spend $44 billion less per year on consumer packaged goods (CPG) if they switched from a national brand to a store brand alternative whenever possible.” In the United States, the private label sector is still surprisingly
underdeveloped relative to other Western countries, representing only 18% of 2014 CPG expenditures, versus a 16.5% weighted global average and rates exceeding 40% in the United Kingdom, Switzerland, and Germany (Nielsen Co. 2014). In several CPG categories, Erdem et al. (2004) find evidence for more consumer uncertainty about the quality of private labels in the United States than in the European countries they analyze.

The literature on consumer behavior routinely finds that consumers shop with limited information. In-store interviews find that most consumers rarely engage in price comparisons and are unable to recall product prices even for goods that they have just put in their baskets (Dickson & Sawyer 1990). Observing consumers in the detergent aisle of a supermarket, Hoyer (1984) documents very limited price search: Only 8% of consumers inspected a single shelf tag, and only 3% inspected more than a single tag. Consumers also appear to lack brand information. In branding experiments, consumers routinely fail to identify their preferred brands in blind taste tests (Husband & Godfrey 1934, Thumin 1962, Allison & Uhl 1964, p. 36).

Bronnenberg et al. (2015) provide empirical evidence for systematic differences in the shopping behavior of experts and nonexperts. They match a CPG shopping panel for almost 90,000 households with survey-based information about each panelist’s professional occupation and product knowledge for selective product categories. After controlling for income and socioeconomic status, consumers with a health-related occupation are considerably more likely to buy unbranded (i.e., store brand) health-related CPG products than consumers in nonhealth occupations. For instance, pharmacists purchase more than 91% unbranded over-the-counter pain medications, whereas the overall average market share for these products is only 65%. There is also a high association between the purchase of unbranded pain medications and the consumer’s knowledge of the active ingredients in typical over-the-counter pain medications. Similarly, consumers with an occupation related to food preparation or food production are considerably more likely to buy unbranded CPG pantry staples, like sugar, salt, and flour. This link between occupation and preference for private labels is domain specific: Health professionals are not more likely to buy private label food staples than other households with comparable socioeconomic status who do not have a food-related occupation.

Generic alternatives to several of the CPG brands studied by Bronnenberg et al. (2015) have been available for decades, suggesting that consumers remain persistently uninformed. Several market frictions can potentially limit consumers’ endogenous gathering of information, or search. Nelson (1970) makes a distinction between search characteristics, which can be determined prior to purchase, and experience characteristics, which are determined after purchase through trial and consumption. We next discuss various mechanisms on the supply and demand sides that can influence the amount of information consumers obtain for both search and experience characteristics of products.

3.5.2. The supply of product quality information. Theorists have studied firms’ incentives to supply product information through branding and brand advertising. Advertising can directly convey information regarding search characteristics, including prices, product features, and availability. Advertising can also indirectly convey information regarding experience characteristics like product quality. The money burning theory of advertising postulates that, in equilibrium, advertising can indirectly convey information about product quality if firms with high-quality products optimally invest more in advertising than firms with low-quality products (Nelson 1974, 1970; Scherer 1978).
Kihlstrom & Riordan 1984, Milgrom & Roberts 1986; see also Bagwell 2007 for a comprehensive survey of the literature on advertising and product quality). Nelson (1974) observes a higher advertising-to-sales ratio for experience goods than for search goods. In an empirical analysis of consumer exposure to advertising before and after the first trial of a newly launched yogurt product, Ackerberg (2001) finds a significant advertising effect only for inexperienced users who had not yet purchased the new product. Similarly, Sovinsky Goeree (2008) finds that informative advertising affects the likelihood that a computer brand enters consumers’ consideration sets, although she models the consideration set in a reduced form.

These theories of informative advertising assume a rational consumer. A more recent behavioral economics literature has studied the incentives for firms to withhold information when selling to boundedly rational consumers. For instance, Gabaix & Laibson (2006) consider a market where firms sell to a mix of expert and nonexpert (myopic) consumers. In this case, sellers optimally withhold information (a process known as shrouding), for instance, regarding add-on prices, to exploit the myopic nonexpert consumers. Interestingly, firms would not engage in educational advertising or marketing to inform consumers about competitors’ add-on practices even if such advertising was costless. In fact, McDevit (2014) studies cases where firms that advertise provide relatively low quality. For instance, firms with brand names starting with the letter A or a number tend to charge higher prices and receive more service complaints. Similarly, firms that purchase advertisements in Yellow Pages or on Google are found to provide lower-quality service. McDevit (2014) rationalizes these patterns with a model in which, in equilibrium, uninformed consumers with relatively low willingness to search will settle for easily found firms (i.e., at start of an alphabetical list or with prominently positioned advertisements).

3.5.3. Consumer learning about the quality of brands. On the demand side, consumers may also gather information on experience characteristics through their endogenous purchase and consumption decisions. The rate at which consumers learn through trial and consumption can be critical for whether they will experiment with new products and learn the objective qualities of substitute brands. Even in a model with consumer learning about product quality where both brands may be valued equally ex post, if the purchase frequency is too low, then a consumer will still prefer to continue buying the known pioneering brand and will not become informed about the quality of newly launched substitute products (Schmalensee 1982). Consumer learning rates can therefore affect the industrial market structure, creating a sustainable advantage to pioneers.

An empirical literature has studied consumers’ learning rates through incoming word of mouth (e.g., Roberts & Urban 1988) and through purchase and experimentation (e.g., Erdem & Keane 1996, Ackerberg 2003, Crawford & Shum 2005). Using diary shopping panels, Demsetz (1962) documents suggestive evidence that consumers do eventually learn about products and that the advantages to branded goods might erode relatively quickly over the course of a few years. However, more recent evidence on product learning rates has been mixed. Consumer psychologists have studied how blocking can prevent consumers from learning about objective product characteristics. If a consumer initially learns to use the brand name to predict an outcome (e.g., taste quality or headache relief), “subsequent learning of the importance of another characteristic (e.g., a grape varietal or an active ingredient) may be blocked” (Van Osselaer 2008, p. 703). In contrast, econometric evidence from the analysis of brand choice panel data using structural models of demand with normal Bayesian learning about product quality has generated relatively fast estimated rates of learning that may only require a few purchase or consumption occasions (e.g., Erdem & Keane 1996, Ackerberg 2003, Crawford & Shum 2005).

One concern with the structural evidence for consumer learning is the potential initial conditions bias associated with consumers’ initial beliefs at the start of the data sample. Most studies
do not observe consumer choices from the initial trial and therefore need to make untestable assumptions about initial product quality beliefs. Typically, researchers assume homogeneous prior beliefs across consumers at the start of the sample period, which could falsely attribute learning to unobserved heterogeneity across consumers in their initial product knowledge at the start of the sample period.\textsuperscript{20} To resolve this bias, Shin et al. (2012) match consumer shopping panel data for toothpaste with a household-level survey on product beliefs conducted immediately before the start of the panel period. Once the survey data are used to calibrate initial beliefs, the authors find considerably slower rates of learning.\textsuperscript{21} These findings suggest that product information could persist as a barrier to entry, granting an advantage to pioneering brands, as in the results of Schmalensee (1982). In the context of durable consumer goods, Erdem et al. (2005) tackle a related identification problem that arises when product quality learning rates and declining prices (e.g., price skimming) co-move over time. They supplement their purchase panel data with survey data on price and quality expectations to resolve the separate identification of learning about product quality and expectations of declining future prices. These survey data also allow them to relax the standard rational expectations assumption typically used in the empirical literature on dynamic discrete choice.

3.5.4. Consumer search for brand information. A separate empirical literature has studied how consumers gather information about search characteristics prior to purchase, in the spirit of the theoretical literature on price search and match-value search (Stigler 1961, Weitzman 1979). Previous work has found that consumers engage in relatively limited search prior to purchase (Beatty & Smith 1987, Moorthy et al. 1997). Several authors have documented an inverted U–shaped relationship between a consumer’s experience in a product category and the extent to which they engage in information search prior to purchase (Punj & Staelin 1983, Moorthy et al. 1997), reflecting the trade-offs between information and prior knowledge. Recent empirical literature has used structural approaches to study the interdependence of consumer choices and consumer search behavior when aspects of product quality are uncertain prior to purchase (see, e.g., Kim et al. 2010, Koulayev 2014, Jeziorski & Segal 2015, Moraga-Gonzalez et al. 2015).\textsuperscript{22} These studies find relatively high estimated search costs and, consequently, very limited consumer consideration sets. In a case study of digital camcorders on Amazon.com, Kim et al. (2010) find that consumers typically consider a very small fraction of the total set of available products and that established brands, like Sony, have a much higher presence in these consideration sets. Similarly, Honka (2014) finds that consumers obtain price quotes from, on average, only 2.96 auto insurance providers in a given year (including their current provider) even though consumers in her data collectively choose among 17 providers and there are at least 141 providers in total in the United States. Even among those who switch, the average consumer obtains only 3.51 price quotes (including that of their current provider). Geico, the largest-share provider and a heavy advertiser, is in over half of the consumers’ consideration sets. As in the results of Schmalensee (1982), these information advantages to established brands could create barriers to entry for newer products.

From an econometric perspective, failure to account for the search process and the consumer’s considered set of brands can also bias the estimates of consumers’ brand preferences. High

\textsuperscript{20}One exception is provided by Ackerberg (2003), who resolves this problem by focusing on learning about one recently launched brand. He is then able to study consumer choices before versus after the initial trial.

\textsuperscript{21}Other work has also found that self-reported survey measures of uncertainty about brand quality and reliability are correlated with actual brand purchase choices and with brand purchase intentions (e.g., Erdem & Swait 1998, Erdem et al. 2002).

\textsuperscript{22}Sovinsky Goeree (2008) does not formally study the underlying search process but allows for informative advertising to influence consumers’ consideration sets for computers.
repeat-purchase rates for brands could be falsely attributed to brand preferences when search costs are not accounted for (see, e.g., Hortacsu & Syverson 2004, Sovinsky Goeree 2008, Honka 2014). This bias in estimates of brand preference can emerge even in contexts where the researcher assumes consumers are fully informed about product qualities and engage only in price search. Identifying search costs separately from preferences typically requires data on the search process or the consideration set (see, e.g., de los Santos et al. 2012, Honka 2014, Koulayev 2014). Alternatively, Hortacsu & Syverson (2004) supplement demand-side moments with supply-side moments based on equilibrium prices under a specific search conduct, and Moraga-Gonzalez et al. (2015) use exogenous shifters of search costs.

Theorists have studied firms’ incentives to influence the magnitude of consumer search costs through information disclosure. Zettelmeyer (1995) finds that firms may endogenously choose not to facilitate consumer search through information disclosure, even if it is costless to do so. Theorists have also studied mechanisms through which firms can further reduce the discovery of information on search characteristics by engaging in obfuscation strategies that increase consumer search costs and, thus, the prices paid by consumers in equilibrium (e.g., Wolinsky 1987, Gabaix & Laibson 2006, Ellison & Wolitzky 2012, Piccione & Spiegler 2012). Such obfuscation strategies have been documented as a supply-side response to new information technologies, such as price comparison websites, that facilitate consumer search (Ellison & Fisher Ellison 2009).

3.5.5. Brand information interventions. We next discuss potential information policy interventions that could be used to inform consumers about objective product quality. Several studies have successfully documented an effect of the provision of calorie information on caloric consumption (Bollinger et al. 2011) and of hygiene-related information on restaurant choices (Jin & Leslie 2003). However, attempts to provide information to overcome the brand versus (unbranded) private label preference gap have been mostly unsuccessful. In a study of product information disclosure for relatively homogeneous goods, Cox et al. (1983) find that few subjects switched to a cheaper private label alternatives even after being informed of their objective similarity to the more expensive branded alternatives. In a cross-store field experiment, Carrera & Villas-Boas (2015) find that posting objective information about the comparability of branded and private label headache medicines does not lead to a discernible increase in demand for the private labels, which sell at lower prices. A psychological signal learning explanation may partly explain the difficulties in overcoming brand preference. As explained by Van Osselaer (2008, p. 703), “if consumers first learn to predict an outcome (e.g., taste quality or headache relief) based on a brand name (e.g., of a wine or headache medicine), subsequent learning of the importance of another characteristic (e.g., a grape varietal or an active ingredient) may be blocked.”

Bronnenberg et al. (2016) conduct in-store blind taste tests in which consumers compare a private label food product to the leading national brand in its category. The advantage of the taste test is that the consumer experiences the subjective consumption experience without the confounding signal from the brand itself. Surprisingly, over 75% of the participants chose the private label in the blind taste test even though only 40% predicted that they would do so. By matching the taste test data with actual loyalty card shopping panel data, the authors use quasi-experimental methods to test for a persistent effect from the information conveyed in the taste test. Although they observe a large short-term effect—the private label demand increases by almost 20 percentage points during the week after the tests—the blind taste test effect weakens considerably over time. By five months after the test, the treatment effect is only about 1.5 percentage points, which

23The authors do find a positive effect of information about the number of peers that buy the unbranded alternative.
may be suggestive of forgetting or may reflect the neutralizing effect of national brand marketing efforts.

3.5.6. Summary. In summary, consumers often make brand purchase decisions with limited information about the availability of products, their prices, and their objective qualities. The extant literature has demonstrated that such limited information can favor established brands. In some cases, consumers may be willing to pay large brand price premia even when comparable products are available at lower prices. On the supply side, firms’ incentives to supply objective product information may be limited. Moreover, evidence from various field studies suggests that the ability to treat consumers with objective product information requires more than mere disclosure.

4. BRAND CAPITAL AND VARIETY

An extensive theoretical literature has analyzed the factors governing the equilibrium supply of product variety within a product category (e.g., Chamberlin 1933, Spence 1976, Dixit & Stiglitz 1977). In this section, we discuss the related literature on the strategic implications of brand capital and firms’ incentives to launch new brands into markets.

4.1. Brand Proliferation

The Federal Trade Commission’s complaint (FTC v. Kellogg 1972) against the four largest ready-to-eat cereal manufacturers in 1972 stimulated academic interest in the concept of brand proliferation and its potentially anticompetitive effects. Between 1950 and 1972, the six largest manufacturers launched over 80 brands (Schmalensee 1978). Between 1974 and 1980, the top four firms in the ready-to-eat breakfast cereal industry launched 33 new brands (Raubitschek 1988). Surprisingly, 42% of these brands failed completely and 21% failed to achieve a nontrivial market share. Between 1985 and 1992, 78 new cereal brands were launched by the largest manufacturers. However, only one of these brands succeeded in becoming an established brand (Hitsch 2006).

Theorists have debated the underlying mechanisms leading to the proliferation of brands. Schmalensee (1978) argues that cereal brand proliferation was the outcome of a multiperiod game in which incumbents crowded the product space to deter future entry. If the launch of new brands requires large, exogenous fixed and sunk development (e.g., advertising) costs, incumbents can preempt future entry by crowding the product space. Such new brand launch costs can vary from $50 million to $100 million (Aaker 1990). This theory has subsequently been criticized on the theoretical grounds that such preemptive product proliferation is only credible in the presence of high exit costs (Judd 1985), a questionable assumption for most CPG categories.

Sutton (1991) also questions the validity of the assumption of exogenous sunk costs for the launch of new consumer brands. When firms can strategically choose the magnitude of the sunk investment to build vertical differentiation (e.g., a higher-quality brand image), one would not expect a competitive escalation in the launch of new brands. Rather, one would expect an escalation in advertising spending as firms compete to build better (not more) brands. Using a broad, cross-industry approach, Sutton (1991) and Bronnenberg et al. (2011) provide empirical evidence of a competitive escalation in the level of CPG advertising expenditures and not in the number of advertised CPG brands in larger markets. Consistent with the theory of exogenous sunk costs, Bronnenberg et al. (2011) do, however, observe an escalation in the number of nonadvertised local (or fringe) CPG brands.

Demand uncertainty provides an alternative mechanism for brand proliferation along with an explanation for the low success rate of new launches. Raubitschek (1988) proposes a simple two-stage model in which firms first choose how many products to launch, under demand uncertainty,
and then compete in the product market. The equilibrium number of products launched by a firm increases with the probability of product success. More recently, Hitsch (2006) studies the high failure rate of new cereal brands. He considers a model in which a firm faces demand uncertainty and sequentially learns the product’s profitability by observing sales each period. In particular, he formalizes the product launch and exit problem as a real option problem. Numerical simulations calibrated with cereal demand estimates indicate that the firm’s value of reducing demand uncertainty is sufficiently high that, as uncertainty rises, firms should be more likely to launch new products even when the expected profits are negative.

4.2. Umbrella Branding

The extant literature on branding has also studied yet another important form of brand capital: brand quality reputation. Many of the new brand launches in consumer industries are simply extensions of an existing brand, such as Coca-Cola’s launch of Diet Coke in the diet cola category in 1982. A firm with an established brand can leverage the strong brand reputation through brand extensions in new product categories. Aaker (1990) enumerates several risks associated with the extension of an established brand to a new product, a practice also known as umbrella branding. Any failures or negative associations with the extension could harm the original brand’s reputation.\(^\text{24}\) Nevertheless, according to Aaker (1990), 40% of the new brands launched in US supermarkets between 1977 and 1984 were brand extensions. Among 7,000 new products launched in supermarkets during the 1970s, only 93 grossed over $15 million and two-thirds of these were brand extensions.

A theoretical literature on brand reputation has emerged that studies the practice of brand extension as a signaling equilibrium. The basic premise involves consumer uncertainty about the quality of new products. Consumers rely on brand reputation as a signal of product quality. By extending its established brand name to new products, a firm can signal high quality.\(^\text{25}\) However, brand extensions to low-quality products will damage the future reputation of the brand and, thus, the future profitability of the firm. In the Wernerfelt (1988) model, firms face a cost to extend the brand, and consumers are uncertain about the quality of both the established branded product and the new product. When the firm extends the brand to a new, low-quality product, consumers form less favorable beliefs about the quality of the established product, reducing their future demand for the established product. In the Choi (1998) model, a multiproduct monopolist considers branding a future stream of new product opportunities with the same name as its high-quality established product. A low-quality brand extension damages the reputational capital of the brand, thereby reducing the positive signal for all future brand extensions.\(^\text{26}\) In the Cabral (2000) model, a firm’s product qualities are correlated, and consumer learning about new products with the same brand name will feed back on their willingness to pay for older products. In all of these

\(^{24}\)Consumer psychologists have found mixed evidence on such spillovers. For instance, a poor experience with a new brand extension may be attributed to the extension component and not to the original brand, limiting feedback in the original category (Van Osselaer 2008).

\(^{25}\)Here, too, the evidence from consumer psychology has been mixed. For instance, if a well-known product that carries only a family brand name (e.g., Godiva chocolate) adds a sub-branded product (e.g., L’Amour by Godiva chocolate truffles) with the same outcomes (e.g., identical levels of quality) to its product portfolio, learning about or forming associations with the sub-brand name could be blocked. If the sub-brand is inferior, it may protect the family brand by absorbing any negative brand equity (Van Osselaer 2008).

\(^{26}\)Specifically, the firm uses price as a signal. Reputational capital reduces the extent to which price distortions are required to signal high quality.
models, there exists a separating equilibrium in which only firms with high-quality products use brand extensions.27

The empirical evidence in the field for such spillovers in consumer quality beliefs is limited. Erdem & Winer (1999) fit a structural model of demand to consumer purchase panel data for toothbrushes and toothpaste. The parameter estimates imply correlation in how consumers perceive a brand across categories. Using the same data, Erdem (1998) fits a structural model of demand with normal Bayesian learning about product qualities in the two categories. Her parameter estimates imply that consumers’ prior beliefs about brand qualities are correlated between the two categories, which would allow for learning spillovers. Erdem & Sun (2002) extend the model to allow for learning effects from marketing decisions like advertising and promotion. The parameter estimates imply not only that advertising and promotion reduce uncertainty about product quality but also that these effects can spill over across the two categories.

The reputational cost from extending a brand to a low-quality new product also potentially creates an implicit exit cost if the new product fails, damaging the reputation of the brand and any future profit opportunities from the brand, including the sales of established products. Thomas (1996) conjectures that this exit cost creates a credible entry-deterring motive for brand extensions. The empirical evidence is mixed. In case studies of the US beer, coffee, and soft drink categories, Thomas (1995) finds that firms with established brand leaders are typically the first to enter new submarkets. However, in a comprehensive analysis of 95 brands across 11 CPG categories, Sullivan (1992) finds that new brands typically enter into new product markets earlier than brand extensions. However, brand extensions that enter later are more likely to succeed in the long run and typically exhibit above-average market shares after controlling for order of entry and advertising.

The literature on umbrella branding reveals that firms with established brands and brand capital may be able to leverage the corresponding reputational benefits into new product markets. Therefore, brand capital may create barriers to entry for new firms not only within an existing product category but also in related new product categories, for which consumers may rely on the reputation of a previously established brand.

5. CONCLUSION

The empirical regularities in the market structures of nondurable consumer goods industries reveal a persistent and central role for established brands. Most nondurable consumer goods industries are highly concentrated, with a small number of brands driving most of the market share. The identity of the dominant brands varies from country to country and even, within the United States, from city to city. Surprisingly, this dominance persists for decades and, in some cases, dates back to the late nineteenth and early twentieth centuries, when the categories themselves were still in their formative stages. The geographic variation in brand shares across US cities today is explained, at least to a large extent, by the historic order of entry among surviving brands.

In this review, we have surveyed an extensive literature in economics and marketing that studies various mechanisms through which established brands can figure prominently in the consideration sets of consumers, generating advantages in both awareness and willingness to pay. Brand experiences early in life have persistent effects on a consumer’s brand choice behavior throughout their lifetime. Some of this persistence may reflect market frictions that prevent a consumer from becoming informed about competing brands and new products. Surprisingly, even in established

27Moorthy (2012) critiques the off-equilibrium beliefs typically used in the models. A refinement allowing for more realistic off-equilibrium beliefs diminishes the signaling capability of a brand extension.
categories with mature brands, consumers appear to shop with very limited information about the broad set of prices and product qualities for the available products. Empirical estimates of the costs of acquiring product information prior to purchase are found to be commensurately high. Consequently, consumers often consider only a small subset of the available products. Learning rates from trial and experimentation after purchase are also found to be slow, further reducing the ability for consumers to become informed.

On the supply side, firms may lack incentives to facilitate the consumer search and learning processes through information disclosure, even when it is costless. In some settings, firms may endogenously obfuscate information to soften price competition. Even if such information disclosure was in the interest of firms, empirical evidence indicates the difficulties in conveying brand quality information to consumers in practice through informational interventions.

The body of literature and the collection of empirical evidence support the longstanding notion that established brands constitute important barriers to entry (e.g., Bain 1956, Demsetz 1982, Schmalensee 1982). Consumers’ brand capital stocks (consumption experiences and branding goodwill) combined with slow rates of learning and the costly acquisition of product information also suggest an early mover advantage that can lead to persistent brand dominance in a product category. The brand capital stocks of established brands can also be leveraged to create strategic advantages in newly emerging markets through brand extensions and umbrella branding.

Most of the literature surveyed in this review uses a microeconomic focus on the role of brands in specific markets and on consumer behavior within those markets. Product differentiation and imperfect competition have been considered in macroeconomic growth models, for instance, to understand new-Keynesian patterns of nominal price rigidity (e.g., Matsuyama 1995) and R&D and technological innovation through the emergence of new products (e.g., Grossman & Helpman 1991, chapter 3). Product differentiation and imperfect competition have also been incorporated into models of trade to study the impact of international trade on intra-industry firm composition and aggregate industry productivity (e.g., Melitz 2003). Given the empirical findings discussed in this review, it could be interesting to consider the role of persistent brand preference and brand loyalty, along with the corresponding brand capital stock, in macroeconomic settings of trade and growth. An interesting first step in this direction was taken by Ravn et al. (2006), who show that consumer habit formation and brand loyalty can help explain countercyclical patterns in markups.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We are grateful to Tulin Erdem, Elisabeth Honka, Carl Mela, Sridhar Moorthy, Robert Sanders, Brad Shapiro, Andrey Simonov, and Stijn Van Osselaer for comments and suggestions.

LITERATURE CITED

Ackerberg DA. 2001. Empirically distinguishing informative and prestige effects of advertising. RAND J. Econ. 32(2):316–33
Becker GS, Murphy KM. 1993. A simple theory of advertising as a good or bad. Q. J. Econ. 108(4):941–64


FTC (Fed. Trade Comm.) v. Kellogg et al., Docket No. 8883 (1972)
Honka E. 2014. Quantifying search and switching costs in the U.S. auto insurance industry. RAND J. Econ. 45:847–84
Jackson L. 2013. Top ten US wireless trends in 2013 for investors to bank on. 24/7 Wall Street, May 22
Kelly KJ. 1994. Here are the top 20 worldwide brands for 1993 as ranked by Financial World: (based on 1993 revenues) (chart) BIG BLUES FOR IBM COCA-COLA SHOWS THAT TOP-BRAND FIZZ. *Advertising Age*, July 11


