The Economics of Brands and Branding\textsuperscript{1}

(Preliminary draft. Please do not quote or cite.)

Bart J. Bronnenberg, Tilburg University and CEPR

Jean-Pierre Dubé, Chicago Booth and NBER

Sridhar Moorthy, Rotman School of Management, University of Toronto

April 22, 2019

\textsuperscript{1}Bronnenberg: Tilburg School of Economics and Management, Tilburg University, Warandelaan 2, 5037AB Tilburg, The Netherlands (e-mail: bart.bronnenberg@tilburguniversity.edu); Dubé: Booth School of Business, University of Chicago, 5807 S Woodlawn Ave, Chicago, IL 60637, USA (e-mail: jdube@chicagobooth.edu); Moorthy: Rotman School of Management, University of Toronto, 105 St. George St, Toronto, ON M5S 3E6, Canada (e-mail: moorthy@rotman.utoronto.ca). Dubé acknowledges the support of the Kilts Center for Marketing and Moorthy acknowledges the support of the Social Sciences and Humanities Research Council of Canada. The authors thank Tülin Erdem, Pedro Gardete, Avi Goldfarb, Brett Hollenbeck, Carl Mela, Helena Pedrotti, Martin Peitz, Sudhir Voleti, and two anonymous reviewers for comments and suggestions.
Abstract

Consumer goods industries tend to be dominated by a small number of nationally-branded goods. In many cases, brands are one of the only sources of differentiation between otherwise physically undifferentiated products. In this chapter we survey the literature on the role of consumer branding and brand preferences in driving consumer demand. We also discuss the incentives for firms to invest in and build brands on the supply side. Finally, we discuss the measurement of the economic value of brands as intangible assets.

Jel Codes: D11, D12, L66, M3
# Contents

1 Introduction  

2 Brand equity and consumer demand  
   2.1 Consumer brand equity as a product characteristic  
   2.2 Brand awareness, consideration and consumer search  
   2.2.1 The consumer psychology view on awareness, consideration and brand choice  
   2.2.2 Integrating awareness and consideration into the demand model  
   2.2.3 An econometric specification  
   2.2.4 Consideration and brand valuation  

3 Consumer brand loyalty  
   3.1 A general model of brand loyalty  
   3.2 Evidence of brand choice inertia  
   3.3 Brand choice inertia, switching costs and loyalty  
   3.4 Learning from experience  
   3.5 Brand advertising goodwill  

4 Brand value to firms  
   4.1 Brands and market structure  
   4.2 Measuring brand value  
   4.2.1 Reduced-form approaches using price and revenue premia  
   4.2.2 Structural models  

5 Branding and firm strategy  
   5.1 Brand as a product characteristic  
   5.2 Brands and reputation  
   5.3 Branding as a signal  
   5.4 Umbrella branding  
   5.4.1 Empirical evidence  
   5.4.2 Umbrella branding and reputation  
   5.4.3 Umbrella branding and product quality signaling  
   5.5 Brand loyalty and equilibrium pricing  
   5.6 Brand loyalty and early-mover advantage  

6 Conclusions
1 Introduction

The economics literature has long puzzled over the concept of brand preference and consumer willingness to pay a price premium for a product differentiated by little other than its brand. In blind taste tests, consumers are often unable to distinguish between their preferred brands and other competing products (Husband and Godfrey, 1934; Thumin, 1962; Allison and Uhl, 1964, p. 336). Nevertheless, branding and brand advertising are perceived to be important investments in sustainable market power:

“A well-known soap-flake which is a branded article costs £150,000 per year to advertise. The price of two unadvertised soap-flakes is considerably less (one of them by more than 50 per cent) than that of the advertised product. Chemically there is absolutely no difference between the advertised product and the two unadvertised soap-flakes. Advertisement alone maintains the fiction that this soap-flake is something superfine. If the advertisement were stopped, the product would become merely one of a number of soap-flakes and would have to be sold at ordinary soap-flake prices. Yet the success of the undertaking, from the producer’s point of view, may be seen from the fact that this product brings in half a million net profit per year.” (Braithwaite, 1928, p. 30)

Brands have also long been recognized as invaluable assets to firms that create barriers to entry and contribute to supranormal economic profits:

“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.” (Bain, 1956, p. 216)

The conceptual meaning of a brand has evolved over time. According to the Oxford English Dictionary, the word “brand” originated in the 10th century. During the 1600s, the term brand was used in the American colonies to designate “mark of ownership impressed on cattle” (Kerschner and Geraghty, 2006, p. 21). Since the nineteenth century, the term brand has taken on a commercial role as a “a trademark, whether made by burning or otherwise” on items ranging from wine and liquor to timber and metals (Murray, 1887, p. 1055). Current marketing practice interprets the brand as “a name, symbol, design, or mark that enhances the value of a product beyond its functional purpose” where the added value of these enhancements to the basic product are often broadly termed “brand equity” (Farquar, 1989, p. 24). On the demand side, this added value can comprise consumption benefits such as image and information benefits such as quality reputation. On the supply side, industry experts associate very high economic value with the commercial rights to leading brands, with reported valuations in the billions of US dollars.\(^1\)\(^2\)

This chapter discusses the economics of brands and branding to understand their impact on the formation of industrial market structures in consumer goods industries. We review the academic literature analyzing the underlying economic mechanisms through which consumers form brand preferences, on the demand side, and the economic incentives for firms to invest in the creation and maintenance of brands, on the supply side. Our discussion builds on earlier surveys of marketing science models of brand equity (e.g., Erdem and Swait, 2014). However, we refer readers seeking

\(^1\)For instance, according to Forbes magazine, the 100 most valuable brands in 2017 represent a global value of US$ 1.95 trillion.

\(^2\)The broad use of the term “brand equity” in reference to both consumer and firm benefits creates confusion. In some of the literature, the added value of brand enhancements to consumers are termed “brand equity” whereas the add value of brand enhancements to firms are termed “brand value”(see for instance Goldfarb, Lu, and Moorthy, 2008).

We have organized this chapter around the following topics. Section 2 discusses two principle roles of brands in affecting demand. First, we discuss how brands affect preferences, and incorporate brand preferences into a neoclassical “characteristics” model of demand. Here, we discuss how consumer brand equity is estimated from consumer choice data. Second, we discuss the role of brands in generating awareness, directing attention and consumer search, and determining the composition of the consideration sets from which brand choices are made. In Section 3, we focus on the formation of consumer brand preferences over time and the emergence of “brand loyalty.” Section 4 discusses brand value estimation from the firm’s point of view. In Section 5 we discuss the strategic considerations for firms to create brand value through reputations, the investment in brand capital, and potentially extending the use of a brand name across products marketed under a common brand umbrella. Finally, Section 6 concludes.

2 Brand equity and consumer demand

2.1 Consumer brand equity as a product characteristic

In this subsection, we focus on characteristics models of demand and the role of brand as a quality-enhancing product feature. The incorporation of product quality into the modeling of consumer preferences represented a turning point in the consumption literature, allowing for a more granular analysis of product-level demand as opposed to commodity-group-level demand (Houthakker, 1953). The role of quality was formalized into a “characteristics approach.” The product is defined as a bundle of characteristics. Consumers have identical perceptions of the objectively measured
characteristics comprising a product, and have potentially heterogeneous and subjective preferences for these characteristics (Lancaster, 1971; Baumol, 1967; Rosen, 1974).

Early work in characteristics models of demand focused purely on objective attributes and did not consider brand. Unlike objective product characteristics, consumer brand preferences (or “brand equity”) typically comprises intangible, psychological factors and benefits. For instance, Keller (1993, p.3)’s conceptual model of brand equity starts with a consumer’s brand knowledge, or “brand node in her memory to which a variety of associations are linked.” These associations in memory include the consumer’s brand awareness and her perceptions of the brand, or “brand image.” But, the psychological mechanism through which brand equity affects a consumer’s utility from a product presents a challenge for the neoclassical economic model. Economists have historically shied away from the psychological foundations of preferences:

“The economist has little to say about the formation of wants; this is the province of the psychologist. The economist’s task is to trace the consequences of a given set of wants.” Friedman (1962, p. 13)

Not surprisingly, early micro-econometric work took a simplified view of the brand as a mark that merely identifies a specific product and links it to a supplier. In his hedonic specification, Rosen (1974, p. 36) explained: “The terms ´product,´ ´model,´ ´brand,´ and ´design´ are used interchangeably to designate commodities of given quality or specification.” Accordingly, Rosen (1974, p. 37) assumed: “If two brands offer the same bundle, but sell for different prices, consumers only consider the less expensive one, and the identity of sellers is irrelevant to their purchase decisions.” So the traditional characteristics approach assumes the consumer derives no preference from the brand itself other than through the objective product characteristics. Brand choice (i.e. “demand”)
is therefore governed entirely by the brand’s objective characteristics.

A micro-econometric demand specification that excludes brand preferences would have limited predictive power in many product markets. According to the standard characteristics model, “two brands which have approximately the same attribute values should have approximately the same market shares,” (Srinivasan, 1979, p.12) a prediction that is frequently rejected by actual market share data (see, for example, the brand share analysis in Bronnenberg, Dhar, and Dubé, 2007). Blind taste tests with experienced consumers also reveal a strong role for brand. In comparisons of blinded and unblinded taste tests that hold all the attributes of popular national brands fixed except the brand labeling on the packaging, experienced consumers routinely exhibit different preference orderings (Husband and Godfrey, 1934; Thumin, 1962; Allison and Uhl, 1964). In Allison and Uhl (1964)’s study, subjects—males who drank beer at least three times a week—tasted six bottles of beer over a week, first blind, with no brand identifiers, and then non-blind, with all the brand identifiers present. In the blind tasting, the six bottles of beer were actually three different brands with “taste differences discernible to expert taste testers.” In the non-blind tasting, the six bottles were actually six different brands—the three that they had originally tasted blind, plus three additional brands. After each tasting, subjects were asked to evaluate the beers, overall, and on particular product attributes such as “after-taste,” “aroma,” and “carbonation.” In the blind tasting, subjects generally rated all the beers to be about the same quality—including the brand that they drank most often. However, unblinded, subjects rated each of the original three beer brands higher and the increases in evaluation varied across brands. Subjects generally rated “their” brands as significantly better than the others even though they could not distinguish them in the blind test. Allison and Uhl conclude: “Participants, in general, did not appear to be able to discern the taste differences among the various beer brands, but apparently labels, and their associations,
did influence their evaluations.³

Ratchford (1975) was an early study that acknowledged the close connection between the characteristics approach in economics and the multi-attribute psychometric approaches (e.g., Green and Srinivasan, 1978; Wilkie and Pessemier, 1974) used in consumer psychology to describe and measure brand preferences and brand attitudes. The lab-based nature of psychometric and stated-preference measures limited their broad applicability to the analysis of consumer purchase data in the field. A parallel literature using stated-preference data, or conjoint analysis,⁴ instead defined the brand equity as a residual which can be measured as a separate brand fixed effect in addition to the other objective product characteristics (Green and Wind, 1975; Srinivasan, 1979). With the advent of consumer shopping panel data, this same approach to brand equity was incorporated into empirical brand choice models (Guadagni and Little, 1983) derived from random utility theory. We now explore such quantitative models of brand choice.

More formally, we consider the following discrete choice or “brand choice” formulation of demand. Suppose consumers have unit-elastic demands for \( j = 1, \ldots, J \) perfectly substitutable branded goods in a category.⁵ We also allow for a \( J + 1^{st} \) “outside good” which we interpret for now as the non-purchase choice. Assume the consumer derives the following choice-specific

---

³These findings would later inspire the famous “Pepsi Challenge” campaign during the 1970s in which subjects exhibited a more than 50% chance of choosing Pepsi over Coca Cola in a blind taste test (http://www.businessinsider.com/pepsi-challenge-business-insider-2013-5).

⁴The conjoint approach to preference estimation defines consumers’ product preference by conjoining her tastes for the product’s underlying attributes, much like the “characteristics approach.”

⁵Following Rosen (1974), we make the discrete choice assumption for ease of presentation.
utilities (i.e., conditional indirect utilities):

\[ v_j = U(\psi_j, y - p_j) + \varepsilon_j, \quad j = 1, \ldots, J \]  

(1)

\[ v_{J+1} = U(0, y) + \varepsilon_{J+1} \]

where \( \varepsilon_j \) is a random utility component for product \( j \), \( y \) is the consumer’s budget and \( p_j \) is the price of product \( j \). It is straightforward to include additional controls for point-of-sale marketing variables, such as in-store merchandizing like displays, in the model. The key object of interest in our discussion of brand preference is \( \psi_j \), the consumer’s total perceived value from brand \( j \) (Guadagni and Little, 1983; Kamakura and Russell, 1989; Louviere and Johnson, 1988; Kamakura and Russell, 1993). In principle, the sign and magnitude of \( \psi_j \) can vary across customers so that branding can lead to both horizontal and vertical sources of differentiation. In section 5, we discuss how firms endogenously make branding decisions on the supply side.

The brand choice literature has proposed various methods to extract the intrinsic perceived brand value from \( \psi_j \). Kamakura and Russell (1993) propose a framework to reconcile the gap between the psychological components of brand preference and the objective product attributes. They use a hierarchical structure that decomposes total brand value as follows

\[ \psi_j = x_j \beta + \gamma_j \]  

(2)

where \( x_j \) are the objectively measured product attributes, \( \beta \) is a vector of corresponding attribute tastes and \( \gamma_j \) is an intrinsic utility for the intangible and psychological components of brand \( j \). For the remainder of our discussion, we will refer to \( \gamma_j \) as the intrinsic value of brand \( j \) in reference to
the added benefits beyond the usual consumption benefits associated with the objective attributes, \( x_j \). This decomposition reveals a potential identification problem if the attributes of a given brand \( j \) do not vary over time or across consumer. In this case, the marginal utilities of all the attributes, \( \beta \), and the perceptual features of brand \( j \), \( \gamma_j \), are not separately identified. Kamakura and Russell (1993) impose additional parameter restrictions to resolve the problem in an application to consumer purchase data. Using stated-preference data, such as a conjoint experiment, circumvents the problem by randomizing the attributes \( x_j \). Alternatively, when more granular, individual product- or so called stock-keeping-unit (SKU)-level data are available, the researcher can exploit the fact that a common brand name may be applied across multiple SKUs with different objective attributes such as pack size, packaging format, and flavor (e.g., Fader and Hardie, 1996). In section 5, we discuss how firms can create brand differentiation through reputation even when products are otherwise undifferentiated (i.e., \( x_j = x_k, \forall j, k \) for all objective attributes).

The total intrinsic brand value in equation 2 can be augmented to include subjective and perceptual aspects of the brand, such as biases in consumer perceptions of the objective attributes (Park and Srinivasan, 1994) and image associations. Typically these psychological attributes are elicited through consumer surveys (see Keller and Lehmann 2006 for a discussion).

The estimated residual, \( \hat{\gamma}_j \), is typically interpreted as the brand equity or brand-related value. We can then derive a natural micro-econometric measure of the total economic value to the consumer associated with the brand equity of product \( j \) using the classic Hicksian compensating differential (Hicks, 1939). The Hicksian compensating differential consists of the monetary transfer to the consumer to make her indifferent between the factual choice set in which brand \( j \) offers brand equity and a counter-factual choice set in which brand \( j \) no longer offers brand equity, all else equal. Researchers often use the term willingness-to-pay (WTP) since the compensating dif-
ferential is equivalent to the maximum dollar amount a customer should objectively be willing to pay to retain brand $j$’s equity in the choice set, all else equal.\footnote{The terminology WTP dates back at least to Trajtenberg (1989) and is used throughout the literature on the value of new products and the value of product features.} Suppose we assume quasi-linear utility, $U(\psi_j, y - p_j) = \psi_j + \theta (y - p_j)$, where $\theta$ is the marginal utility of income (see Chapter I of this volume for more discussion). When the the random utility shocks are also assumed $\epsilon \sim \text{i.i.d. EV}(0, 1)$, we get the multinomial logit demand system and the willingness-to-pay for brand $j$’s equity is:

$$WTP^\text{brand}_j = \int \frac{\ln \left( 1 + \sum_{k=1}^J \exp \left( U(\cdot; \beta, \gamma_k) \right) \right)}{\theta} - \frac{\ln \left( 1 + \sum_{k=1}^J \exp \left( U(\cdot; \beta, \gamma_k \neq j, \gamma_j = 0) \right) \right)}{\theta} \, dF(\Theta)$$

(3)

where $F(\Theta)$ is the distribution reflecting the researcher’s statistical uncertainty over all the model parameters, $\Theta$.\footnote{Many applications also allow the equilibrium prices to adjust in response to the demand shift associated with the removal of brand $j$’s equity, $\gamma_j = 0$.} Swait, Erdem, Louviere, and Dubelaar (1993) propose a similar measure, termed “Equalization Price,” which measures the compensating differential without taking into account the role of the random utility shocks, $\epsilon$. Since the estimation of the brand intercepts typically requires a normalization, the exact interpretation of $WTP^\text{brand}_j$ depends on the definition of the base choice against which the brand intercepts are measured, typically the “no purchase” choice which is assumed to offer no brand equity. A more comprehensive set of survey-based, perceptual measures such as brand attitudes, consumer opinions, perceived fit and reliability can also be incorporated into the analysis (e.g., Swait, Erdem, Louviere, and Dubelaar, 1993).

In practice, some researchers use a simpler monetary measure of the brand equity based on the
equivalent price reduction (e.g., Louviere and Johnson, 1988; Sonnier, Ainslie, and Otter, 2007):

\[ BE_j = \frac{\gamma_j}{\theta}. \]  (4)

Holding all else constant, this price reduction ensures that the consumer has the same expected probability of buying brand \( j \) in the counterfactual scenario where \( \gamma_j = 0 \), i.e., where the intrinsic utility for the intangible, psychological components of brand \( j \) are absent. In practice, researchers typically plug point estimates of \( \gamma \) and \( \theta \) in 4. Formally, one ought to use the correct expected incremental utility that takes into account the statistical uncertainty in the estimates:

\[ BE_j = \int \frac{\gamma_j}{\theta} dF(\Theta). \]  (5)

where, as before, \( F(\Theta) \) is the distribution reflecting the researcher’s statistical uncertainty over the model parameters, \( \Theta \). It is straightforward to show that (5) is identical to the willingness-to-pay for brand \( j \)'s equity only in the extreme case where consumer utility is deterministic (i.e. there is no random utility component), brand \( j \) is the only available product and the consumer is forced to purchase it.\(^8\) Another advantage of using \( WTP_j^{brand} \) as in (3) versus (5) to measure brand equity is that the former will vary depending on how we combine brand \( j \) with other product features and prices.

In many demand studies, the intrinsic brand value, \( \gamma_j \), is treated as a nuisance parameter that controls for all the intangible aspects of a product that are either difficult or impossible to measure objectively. In this regard, the brand intercepts improve the predictive power of the model. The

---

\(^8\)Once we include random utility shocks into the model, \( BE_j \) is no-longer a welfare measure.
non-parametric manner in which $\gamma_j$ controls for brand preference is, however, both a blessing and a curse. Brand value research often interprets the estimated residual, $\hat{\gamma}_j$, as marketing-based component of brand equity (e.g., Park and Srinivasan, 1994), in contrast with the product-based component captured by the objective product attributes, $x_j$. An obvious limitation of this approach is that any omitted product characteristics will be loaded into $\hat{\gamma}_j$. So brand equity measures like (4) and (3) should probably be interpreted as noisy measures of the marketing-based component of brand equity. An additional limitation is that the model treats perceived brand equity as a static feature of the product without providing any insight into the formation of brand preferences. In section 5.2, we discuss an alternative informative theory of brand equity that assumes there is no intrinsic brand preference. Rather, the brand name conveys an informative signal used by the consumer to infer product quality through the brand’s marketing and/or reputation.\textsuperscript{9} In section 3, we extend our discussion to a dynamic setting in which the consumer’s brand equity evolves over time through her past consumption and marketing experiences. A final concern is that the measures above fail to account for the supply side. If demand for a branded good is fundamentally altered or if a branded good is excluded from the market, then equilibrium prices would likely re-adjust on the supply side (along with other marketing decisions). In section 4, we discuss firms’ branding strategies.

The “characteristics approach” to consumer brand value described above is the most common approach to deriving and measuring the economic value for a brand. Becker and Murphy (1993)\textsuperscript{9}

\textsuperscript{9}As noted by Nelson (1974); Sutton (1991), firms spend large amounts of money in seemingly uninformative advertising in categories that are essentially commodities. While Nelson (1974) has argued that this sort of advertising might signal product quality indirectly, the empirical evidence for his hypothesis is mixed at best. Caves and Greene (1996a, p. 50), after examining 196 product categories, conclude that “[t]hese results suggest that quality-signalling is not the function of most advertising of consumer goods”; Bagwell (2007a, p. 1748)’s review is only slightly more circumspect: “... the studies described here do not offer strong support for the hypothesis of a systematic positive relationship between advertising and product quality.”
proposed an alternative “complementary goods” theory of brand value whereby the market good and its brand are both complementary consumption goods in the sense of Hicks and Allen (1934). In this framework, brand equity would need to be defined through the complementary effect of the consumption of the brand/branding and the consumption of the corresponding physical good. To the best of our knowledge, Kamenica, Naclerio, and Malani (2013) provide the only direct evidence for this theory of brands.\textsuperscript{10} 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.

Tests for brand effects as complementary goods (Becker and Murphy, 1993) and the specification of a demand system with complementary goods (e.g., Song and Chintagunta, 2007) are beyond the scope of this chapter.

### 2.2 Brand awareness, consideration and consumer search

#### 2.2.1 The consumer psychology view on awareness, consideration and brand choice

The Lancasterian model described above takes the extreme view that the consumer has complete information about the set of available brands and their attributes at the time of purchase.

In the psychology literature on consumer behavior, Lynch and Srull (1982) refer to this scenario as “stimulus-based choice.” At the opposite extreme, the consumer uses a pure “memory-based choice” (Bettman, 1979), whereby all relevant choice information must be recalled from memory.

\textsuperscript{10} An indirect test of Becker and Murphy (1993)’s 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, Nair, and Gardete (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.
As explained in Alba, Hutchinson, and Lynch (1991), in practice most brand purchase contexts will require at least some reliance on recalled information. Even when the purchase environment (e.g. the shelf display in a store) contains all the relevant brand and attribute information, the complexity of the task, the ease with which certain brands are noticed relative to others, and the consumer’s time cost or effort can all lead to reliance on recalled information.

In the brand choice literature, researchers studying brand choice under incomplete information distinguish between limited awareness and consideration. A consumer’s brand awareness comprises the set of brands recalled from her memory. This set may be much broader than the subset of brands the consumer evaluates more seriously for choice (Campbell, 1969), the so-called “evoked set” (Howard and Sheth, 1969), or “consideration set” (Wright and Barbour, 1977). However, the concept of awareness precedes that of consideration, i.e., any brand associations that facilitate recall will, in turn, influence a brand’s inclusion in the consideration set (Keller, 1993).

**Awareness** The extent of awareness for a given brand in the market has been studied in a number of ways. Laurent, Kapferer, and Roussel (1995) define the unaided awareness for a brand as the fraction of households who spontaneously recall a specific brand when asked about choice options in a category. A related measure, top-of-mind awareness for a brand, indicates the fraction of households who spontaneously recall that brand as the first one when prompted. Aided awareness measures the fraction of households that recognize a specific brand name from a given list of brands in a category. Most studies find that a consumer’s brand awareness within a product category is quite limited. Laurent, Kapferer, and Roussel (1995) report that unaided brand awareness in a product category is 15-20%, even for those brand names recognized by 75% of consumers once prompted. In addition to being limited, the literature also reports that unaided brand awareness
varies over time within households (Day and Pratt, 1971; Draganska and Klapper, 2011).

The relevance of awareness in this chapter on branding stems from research showing that a consumer’s ability to recall a specific product from memory is affected by the corresponding brand name. For instance, preferred brands tend to be recalled earlier than non-preferred brands (Axelrod, 1968; Nedungadi and Hutchison, 1985). Further, as consumers accumulate more knowledge about a product category, they tend to structure their memory around brands (Bettman and Park, 1980), largely due to the fact that for consumer goods, most experiences are brand-based (e.g., advertising, in-store merchandizing and consumption experience). Even factors as simple as lack of name recognition can block a brand from being recalled and, subsequently, from entering a consumer’s consideration set (see the discussion in, e.g., Alba, Hutchinson, and Lynch, 1991).

The branding literature has viewed brand awareness as a necessary but insufficient condition for brand consideration and choice, at least since Axelrod (1968). In a study of the German ground coffee market, Draganska and Klapper (2011) report that even in a heavily advertised category like coffee, the typical consumer spontaneously recalls only three brands from the total available set consisting of five major national brands and many fringe brands. Furthermore, Draganska and Klapper (2011) report that the set of recalled brands varies across respondents and accounts for a large part of the heterogeneity in choices (we deliberately avoid using the term “heterogeneity in preferences”).

Consumer psychologists assign a distinct role to brand awareness versus brand preferences on brand choices. For instance, in lab experiments that manipulate the level of brand awareness and product quality, Hoyer and Brown (1990) find that subjects picked the “familiar” brand 77% of the time, even though the familiar brand was frequently not the one with the highest quality. Surprisingly, subjects were more likely to choose the high-quality alternative when none of the
brands was “familiar.” Nedungadi (1990) also finds that choice outcomes can be affected by factors that affect brand recall but not brand preference. Consumer expertise likely moderates these effects. For instance, Heilman, Bowman, and Wright (2000) find that first-time consumers in a product category are more likely to purchase more familiar brands than experienced consumers.

**Consideration**  A separate literature has explicitly studied consumers’ brand consideration sets for choice. Even though a consumer may be aware of a number of brands, she may only consider a subset of them on any given purchase occasion (Narayana and Markin, 1975; Bettman and Park, 1980; Ratchford, 1980; Shugan, 1980). For an overview of early literature on empirical consideration sets, see the discussion in Shocker, Ben-Akiva, Boccara, and Nedungadi (1991). In an empirical study of brand choices, Hauser (1978) found that consumers’ consideration sets explained 78% of the variance in their brand choices; only 22% was explained by preferences within consideration sets. Empirical researchers have typically found that consideration sets in brand choice settings range in size from only 2 to 8 alternatives (Bronnenberg, Kim, and Mela, 2016; Hauser and Wernerfelt, 1990; Honka, 2014; Moorthy, Ratchford, and Talukdar, 1997; Newman and Staelin, 1972; Punj and Staelin, 1983; Ratchford, Talukdar, and Lee, 2007; Urban, 1975). These limited consideration sets are consistent with the psychological theory that individuals’ ability to evaluate choices may be cognitively limited to a maximum of about seven (Miller, 1956).

In sum, consumer psychologists make a distinction between brand awareness, which is recalled from memory, and brand consideration, which reflects the consumer’s deliberation process of narrowing down the set of options before making a brand choice. The literature has further documented strikingly limited degrees of awareness and consideration. An interesting direction for future research might consist of testing the extent to which the limited varieties purchased by
households in most categories of consumer goods reflects a lack of awareness of all the available brands. One recent study of over 32 million US shoppers found that over a 52-week period ending in June 2013, even the most frequent shoppers purchased only 260 of the 35,372 stock-keeping units available in supermarkets, about 0.7%. Across category, the amount varied from as low as 0.2% in Health & Beauty to as high as 1.7% in Dairy (Catalina Media, 2013). Moreover, awareness has been shown to influence brand choices independently of brand preferences. The literature has not yet studied whether consumers rationally plan their awareness, by informing themselves strategically about brands or, alternatively, whether this awareness set is exogenous to consumer decision making.

A recent empirical literature has used data on consumers’ consideration sets to show that the assumption that consumers consider all the available brands in a market will likely result in biased estimates of brand preferences. In practice, if consumers are more likely to consider branded goods, which typically charge higher prices, a naive model of full consideration may generate a downward bias in the estimated price sensitivity (Honka, 2014). Similarly, a naive model that ignores the consideration stage may generate an upward bias in the degree of estimated preference heterogeneity (Dong, Morozov, Seiler, and Hou, 2017).

2.2.2 Integrating awareness and consideration into the demand model

We now formalize the notions of awareness and consideration into our economic model of consumer demand. We build on the Lancasterian framework from section 2.1 that assumed consumers were fully aware of all available brands and considered each variant for choice. Throughout this section, we maintain the assumption that all products in a category are perfect substitutes and, hence, that consumers will make a pure discrete choice purchase decision.
We assume that at the time of purchase from a commodity group, the consumer recalls the brand alternatives in the set $S_a \subseteq S$, where $S$ is the full set of available products; but is unaware of brand alternatives in its complement $\setminus S_a = S - S_a$. The consumer is uninformed about the availability or the characteristics of products in the complement of $S_a$ and does not take this uncertainty into account when making a decision. In this section, we take a static view of the consumer that treats $S_a$ as exogenous.\footnote{In section 3 below, we also offer a dynamic view of the consumer that considers her purchase history. In the multi-period setting, awareness can form endogenously through a consumer’s past purchase experiences.} In section 2.2.3 below, we also allow for a consumer’s awareness set, $S_a$, to be influenced by the endogenous branding and marketing activities of firms, on the supply side.

Conditional on her awareness set $S_a$, we assume the consumer’s purchase decision is the outcome of a two-stage sequential process: (1) the search and evaluation stage, and (2) the choice stage.\footnote{A broad literature has documented evidence that consumers use such two-step “consider-then-choose” decision-making (e.g., Payne, 1976, 1982).} During the first stage, or search and evaluation stage, the consumer forms her consideration set, $S_c \subseteq S_a$ by evaluating ex ante which products to include in the consideration set so as to maximize her expected consumption utility net of search and evaluation costs (Shugan, 1980; Roberts, 1989; Roberts and Lattin, 1991).\footnote{While consumers generally use cost/benefit decision rules, they may rely on simpler heuristic approaches in situations with more complex decision tasks (e.g., Payne, 1976, 1982).}

Formally,

$$S_c = \arg \max_{S_c \subseteq S_a} \left[ \mathbb{E} \left( \max_{j \in S_c} (v_j) \right) - C(S_c) \right], \quad (6)$$

where $v_j$ is the indirect utility from consuming alternative $j$, and $C(S_c)$ is a product evaluation or search cost associated with assembling the consideration set $S_c$. Unlike the traditional Lancasterian model in which branding affected choices through preferences for the branded goods, we now consider the possibility that branding plays another, complementary, role. Let the incremental cost
of gathering information for brand $j$ be denoted by $\Delta C_j$. We allow the cost of gathering and/or interpreting information $\Delta C_j$ to depend on branding. This effect of branding could reflect explicit factors at the point of purchase, such as shelf placement, that can aid consumers in processing information. It could also reflect branding efforts outside the store, like advertising, which may affect the consumer’s ability to recall a specific brand from memory.

The exact composition of the consideration set $S_c$ depends on the consumer’s search conduct. The traditional search literature assumes that consumers randomly sample information (e.g. prices) from a set of ex-ante identical sellers at a fixed and constant cost for each of the firms sampled (Stigler, 1961). Weitzman (1979) was the first to consider search over differentiated products, allowing consumers to prioritize search systematically for those products with the highest anticipated utility. If awareness or familiarity make brands easier to search, then consumers prioritize their search across brands in an order determined by brand awareness, or “prominence” (Arbatskaya, 2007). As Armstrong and Zhou (2011) put it:

“In many circumstances, however, consumers consider options in a non-random manner and choose first to investigate those sellers or products which have high brand recognition, which are recommended by an intermediary, which are prominently displayed within a retail environment, which are known to have low price, or from which the consumer has purchased previously.” (Armstrong and Zhou, 2011, page F368)

Similarly, Erdem and Swait (2004) find that self-reported measures of brand credibility affect the likelihood that a brand enters a consumer’s consideration set. Because of the presence of search and evaluation costs in the first stage of the choice process, equation (6), information gathering often ends before exhausting all options in $S_a$ and $S_c \subset S_a$. 

19
During the second stage, or “choice stage,” the consumer picks the alternative in her consideration set, \( k \in S_c \), that yields the highest utility:

\[
k = \arg \max_{j \in S_c} \{ v_j \}.
\]  

The decision in equation (7) is a reformulation of the discrete choice problem in equation (1) from section 2.1, where all brands entered the choice problem. Models in the literature typically assume that search fully resolves the uncertainty about considered alternatives.  

A discussion of additional literature, in sections 3.2 and 3.3 below, suggests that brand reputation, loyalty, and pioneering advantage can cause a brand to be more likely to enter the awareness and consideration sets.

### 2.2.3 An econometric specification

To illustrate, we now modify the econometric modeling framework from section 2.1. An early literature abstracted entirely from the formation of consideration sets, modeling them in reduced-form instead (see Shocker, Ben-Akiva, Boccara, and Nedungadi, 1991). In this literature, the consideration set is treated as an additional random variable and the likelihood that a consumer chooses brand \( j \) is as follows:

\[
\Pr(j; \Theta) = \sum_{S_c \in \mathcal{P}(S)} \Pr(j|S_c; \Theta) \times \Pr(S_c; \Theta)
\]  

\[\text{(8)}\]

\[\text{In section 3.4 below, we will formally distinguish between search and experience characteristics (e.g., Nelson, 1970). We will then allow for ex post uncertainty about experience characteristics at the time of purchase. This uncertainty will be resolved slowly over time across repeated purchase and consumption experiences.}\]
where \( \Pr(j|S_c) \) is the brand choice probability conditional on the consideration set, \( \Pr(S_c) \) is the probability of observing consideration set \( S_c \in \mathbb{P}(S) \) and \( \mathbb{P}(S) \) is the power set of \( S \). If \( S_c = S \) and all brands are considered, then \( \Pr(S) = 1 \) and \( \Pr(j|S) \) is the choice probability from section 2.

In practice, consideration sets are unobserved and the likelihoods \( \Pr(j|S_c) \) and \( \Pr(S_c) \) are not separately identified without strong, and often ad hoc, functional form assumptions.\(^{15}\) While a literature has estimated models of consideration and choice without observing the consideration set, this approach is clearly prone to severe model misspecification concerns. Even when consideration sets are observed, the standard discrete choice models like logit and probit are probably not the correct reduced form for the conditional brand choices.\(^{16}\) In particular, the choice problem

\[
\Pr(j|S_c) = \Pr(v_j \geq v_k, \text{ for } k, j \in S_c)
\]  

selects on realizations of random utility shocks in \( v_j \) for products \( j \) that were systematically considered. So the choice problem (9) cannot simply “integrate out” the random utility shocks under the usual i.i.d. assumptions to obtain the discrete choice probabilities because \( S_c \) is also a function of the realizations of \( \varepsilon \) for searched brands. Mehta, Rajiv, and Srinivasan (2003) and Joo (2018) use exclusion restrictions based on the assumption that in-store promotional variables, like display and feature, affect the consumer’s information about products and not consumption utility and preferences: \( \Pr(S_c;Z) \) where \( Z \) contains the variables excluded from the brand choices conditional on consideration.\(^{17}\) In addition, the model in equation (8) suffers from a curse of dimensionality

\(^{15}\)Recently, Abaluck and Adams (2017) show that symmetry in the cross-derivatives of choice probabilities only holds when the consumer considers all possible options. They propose to identify consideration probabilities from the violations of symmetry in the cross-derivatives.

\(^{16}\)For instance, Srinivasan, Park, and Chang (2005) require the strong assumption that brand awareness is independent of brand preferences in order to retain the conventional logit functional form.

\(^{17}\)At least since Guadagni and Little, 1983, the brand choice literature has found that these promotional variables affect choices. The exclusion restriction assumes, logically, that the effect reflects the ease of search and product
since the dimension of $P(S)$ becomes unmanageable as the set of available products grows.

To provide an illustrative model, at the consideration stage, we assume that the consumer’s indirect utility from purchase is additively separable in the factors that are known deterministically to the consumer and the factors about which she is still uncertain. Formally, as in the discrete choice problem of equation (1) above, assume the consumer’s choice-specific indirect utilities $v_j$ have a component $x_j\beta + \gamma_j + \epsilon_j$ that is known deterministically to the consumer (the econometrician only observes the distribution of $\epsilon$ and $E(\epsilon) = 0$). In addition, the indirect utilities contain the vector of unknown match values, $\xi \sim F(\xi)$ about which the consumer is uncertain. $F(\xi)$ represents the consumers’ beliefs about the unknown match values. Thus,

$$v_j = x_j\beta + \gamma_j + \epsilon_j + \xi_j, \ j \in S_a.$$ (10)

During the search stage, the consumer endogenously resolves her uncertainty $\xi_j$ for a set of considered products, $S_c \subseteq S_a$. Products in the consideration set are selected based on (1) their respective option values (e.g., the variance of $\xi_j$), (2) known indirect utilities ($x_j\beta + \gamma_j + \epsilon_j$), and (3) search costs ($\Delta C_j$).

In contrast to the Lancasterian approach above, the deterministic component of preferences only partially determines the chosen product. Search costs and the option value from additional search and evaluation also influence the overall considered set and, therefore, the chosen product alternative. In our illustrative model, we assume that the total search and evaluation cost associated with a consideration set $S_c$ is $C(S_c) = \sum_{j \in S_c} c_j$ where $\Delta C_j \equiv c_j$ is independent of $c_k (\forall k \neq j)$, and

---

evaluation as opposed to preference. Indeed, it seems unlikely that consumers derive consumption utility from an in-store display. However, to the best of our knowledge, this point has not been tested empirically.
\( c_j \geq 0.18 \) In turn, the search costs, \( c_j \), reflect a consumer’s past experiences with the available products in the category or a firms’ advertising strategy.\(^{19}\) For instance, we could assume:

\[
    c_j = c_0 + G_{1j}(H,A;\Lambda_1),
\]

(11)

where the function \( G_{1j} \) captures the effect of state vectors summarizing a consumer’s brand purchase history, \( H \), past exposure to advertising, \( A \), and \( \Lambda_1 \) is a vector of parameters. Thus, shopping history, purchase experience, and advertising exposure influence her costs, \( c_j \), of gathering and evaluating information about the choice alternative. For instance, in a case study of residential plumbers, McDevitt (2014) finds that low-quality firms systematically use easier-to-find brand names that start with A or a number and tend to be located at the top of directories. Furthermore, in a study of consumers’ retail bank choices, Honka, Hortaçsu, and Vitorino (2017) find that advertising has a much larger effect on awareness and consideration than on consumers’ choices from their respective consideration sets.

The framework herein points towards a major limitation of the extant empirical literature and an opportunity for future research. While some progress has been made in the collection of consideration set data, a consumer’s awareness set, \( S_a \), is not observed in typical choice datasets. Consequently, the consideration papers derived from search theory generally assume that consumers are aware of all the product alternatives and search over an i.i.d. match value. This assumption is at odds with laboratory studies conducted by consumer psychologists that question the plausibility of

\(^{18}\)This specification treats search and evaluation costs as a fixed parameter. Alternatively, the rational inattention literature on consumer choice (e.g., Matejka and McKay, 2015) uses Shannon entropy to model the costs associated with the precision of the signals a consumer endogenously collects to learn about product values.

\(^{19}\)To the extent that advertising influences brand knowledge and psychological associations, it could also facilitate recall.
“full awareness” and document evidence suggesting very limited consumer brand awareness, even at the point of purchase. Econometrically, the distinction between consideration and awareness offers a potential direction for future research, along with a push to integrate research on memory into our demand models. If consumers only search and, thus, only consider brands they can recall from memory (i.e., brands in their awareness set), then awareness and marketing investments that stimulate awareness may create barriers to entry for new products.

2.2.4 Consideration and brand valuation

The possibility of a brand effect in the pre-purchase search- and product evaluation process raises concerns about the measurement of brand equity measures like those in section 2.1 above. Standard discrete choice models that ignore the search and consideration aspects of demand will load the entire brand effect, including that of recall and search, into preferences for the brand $\gamma_j$, defining brand equity as $BE_j = \frac{\gamma_j}{\theta}$. The omission of the consideration stage could bias estimates of brand value. Consider the illustrative case where the consumer has homogeneous ex-ante beliefs about her indirect utility for each of the products in her awareness set, i.e., $E(v_j) = v, \forall j \in S_a$, but where the search costs, $c_j$, to resolve the match value $\xi_j$ are lower for branded than unbranded alternatives among the products in $S_a$. In this case, the consumer’s consideration set and observed brand choices would systematically contain branded products over unbranded ones; not because of higher utility, $v_j$, but because of lower search costs $c_j$. The omission of the consideration stage could confound search costs and brand preferences. The extent of this confound would be exacerbated by the number of alternatives available and/or the magnitude of search costs. To the extent that search stops before any unbranded alternatives are discovered and considered, a traditional Lancasterian model would generate strong estimated brand preferences and potentially
high estimates of brand equity. Substantively, this scenario could lead to the conclusion that brand value stems from preferences, as opposed to from the ease of search.

3 Consumer brand loyalty

3.1 A general model of brand loyalty

In the previous section, we used a static perspective on brand choice that treated brand equity as a persistent residual (or “fixed effect”) in a characteristics specification of the economic model of consumer demand. However, the model abstracted from the manner in which the brand preference was formed. If brand preference is merely a nuisance or control, this may be sufficient for predicting demand. However, as we show below, the dynamics associated with the formation of brand preferences may be important for understanding product differentiation and the foundations of market structure and concentration. In this section, we discuss various dynamic theories of the formation of brand preferences.

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 her 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.

This chapter will not discuss the deeper psychological mechanisms through which preferences are formed. Instead, we focus on how different sources of brand preference formation create dependence on past choices or state dependence in consumer demand. Empirically, state dependence can lead to inertia in a consumer’s observed sequence of brand choices: consumers have a higher probability of choosing products that they previously purchased. Brand choice inertia is one of the oldest and most widely-studied empirical phenomena in the marketing literature (e.g., Brown, 1952, 1953) as it has typically been interpreted as “brand loyalty.” Below we survey the empirical evidence for inertia in consumer brand choices and discuss the econometric challenges associated with disentangling spurious sources of inertia from genuine structural state dependence, such as loyalty. We then discuss several consumer theoretic mechanisms that can generate brand choice inertia as a form of structural state dependence.

To formalize our discussion of the empirical literature, we consider the choices of a household, $h$, over brands, $j$, at time, $t$. We use $X^h_t$ to denote the contemporaneous factors such as product characteristics and marketing considerations like prices, promotions and shelf space. We use the state vector $H^h_t$ to denote a consumer’s historic brand experiences. We include these state variables, $X^h_t$ and $H^h_t$, into a consumer $h$’s indirect utility for brand $j$ on date $t$:

$$
\gamma^h_{jt} = \mu_j \left( X^h_t; \Theta^h \right) + F_j \left( H^h_t; \Lambda^h \right), \quad j = 1, \ldots, J
$$

(12)

where we now decompose the consumer’s brand equity into the static components discussed in the last section, $\mu_j \left( X^h_t; \Theta^h \right)$, and the consumer’s past experiences, $F_j \left( H^h_t; \Lambda^h \right)$, comprising a stock of historically formed brand “capital.” The vectors $\Theta^h$ and $\Lambda^h$ are parameters to be estimated.
Theorists have analyzed various mechanisms through which current willingness to pay for brands reflects past brand experiences. In the subsections below, we explore several formulations of the brand capital stock, $F_j \left( H_t^h; \Lambda^h \right)$, such as switching costs (e.g., Farrell and Klemperer, 2007), advertising and branding goodwill (e.g., Doraszelski and Markovich, 2007; Schmalensee, 1983), evolving quality beliefs through learning (Schmalensee, 1982), habit formation (e.g. Becker and Murphy 1988; Pollak 1970), and peer influence (e.g. Ellison and Fudenberg, 1995).

### 3.2 Evidence of brand choice inertia

The empirical analysis of brand loyalty, or inertia in brand choice, has been one of the central themes of the quantitative marketing research on brand choices. Most of the literature has focused on short-term forms of persistence in brand choices over time horizons of no more than one or two years. Early work by Brown (1952, 1953) exploited household-level diary purchase panel data to document the high incidence of spells during which a household repeat-purchased the same brand over time. Such persistent, repeat-purchase of the same brand has been detected subsequently across a wide array of industries, including those dominated by sellers with products differentiated mainly by brand names rather than objective features. Empirical generalizations across a broad array of CPG categories have found low rates of household brand switching (Dekimpe, Steenkamp, Mellens, and Vanden Abeele, 1997) and high rates of expenditure concentration with typically over 50% of spending allocated to the most preferred brand in a category (Hansen and Singh, 2015). Similar patterns of inertia in choices have been documented in other industries such as insurance (Handel, 2013; Honka, 2014), broadband services (Liu, Chintagunta, and Zhu, 2010), cellular services (Grubb and Osborne, 2015), and financial services (Allen, Clark, and Houde, 2016).
While early work interpreted short-term brand re-purchase spells as evidence of loyalty, the mere incidence of repeat-buying need not imply inertia per se. A consumer with a strong preference for Coca-Cola has a high probability of repeat-purchasing Coca-Cola over time, even if her shopping behavior is memoryless and static. A test for inertia consists of testing for non-zero-order behavior in a consumer’s choice sequence. Early work tested for higher-order behavior using the within-household variation in choices, often with a non-parametric analysis of the observed runs20 within a given consumer’s choice sequence (e.g., Frank, 1962; Massy, 1966; Bass, Givon, Kalwani, Reibstein, and Wright, 1984). Unfortunately, short sample sizes typically limited the power of these within-household tests and the findings were typically mixed or inconclusive; although early work often interpreted a failure to reject the null hypothesis of a zero-order choice process as evidence against loyalty. Alternative testing approaches that pooled choice sequences across consumers ran into the well-known identification problem of distinguishing between choice inertia and heterogeneity in consumer tastes (e.g., Massy, 1966; Heckman, 1981). More recent structural approaches have applied non-linear panel methods to test for choice inertia while controlling for heterogeneity between consumers using detailed consumer shopping panels (Roy, Chintagunta, and Haldar, 1996; Keane, 1997; Seetharaman, Ainslie, and Chintagunta, 1999; Shum, 2004; Dubé,Hitsch, and Rossi, 2010; Osborne, 2011).21 This literature has documented surprisingly high levels of inertia in brand choices. For instance, Dubé, Hitsch, and Rossi (2010) find a substantial decline in the predictive fit of a choice model when, all else equal, the exact sequence of a consumer’s purchases is randomized. This evidence confirms that the observed sequence of choices contains information for predicting demand. We discuss these methods further below in

20 A run is broadly defined as a sequence of repeat-purchases of the same brand. Typically, researchers look at pairs of adjacent shopping trips during which the same brand was purchased.

21 These methods also control for other causal factors, such as prices and point-of-purchase marketing that could potentially confound evidence of inertia.
Patterns of brand choice persistence have also been measured over much longer time horizons, spanning decades or even an individual’s lifetime. For instance, Guest (1955) surveyed 813 school children on their preferred brands in early 1941. Twelve years later, in the spring of 1953, he repeated the same brand preference survey among the 20% of original respondents that he was able to contact. Across 16 product categories, a respondent indicated the same preferred brand in both waves in 39% of the cases. In 1961, a third wave of the same survey continued to find the same preferred brand in 26% of cases. These survey results suggest that brand preferences developed during childhood partially persist into adulthood. However, “obviously, one cannot simply assume that what is learned during childhood somehow 'transfers intact' to adult life” (Ward, 1974). Returning to our model in equation (12), under this extreme view, a consumer’s preferences throughout her lifetime are entirely driven by $F_j(H_0^h; \Lambda^h)$ where $H_0$ represents her initial experiences in life, and $\mu_j(X_t^h; \Theta^h) = 0$. A proponent of this view, Berkman, Lindquist, and Sirgy (1997, pp. 422-3) suggests that preferences may be inter-generational: “[i]f Tide laundry detergent is the family favorite, this preference is easily passed on to the next generation. The same can be said for brands of toothpaste, running shoes, golf clubs, preferred restaurants, and favorite stores.”

The literature on consumer socialization research has studied mechanisms through which adult brand preferences are formed early in life during childhood (Moschis and Moore, 1979), especially through inter-generational transfer and parental influence (Ward, 1974; Moschis, 1985; Carlson, Grossbart, and Walsh, 1990; Childers and Rao, 1992; Moore, Wilkie, and Lutz, 2002) and peer influence (Reisman and Roseborough, 1955; Peter and Olson, 1996). Anderson, Kellogg, Langer, and Sallee (2015) document a strong correlation in the automobile brand preferences of parents
and their adult children. Sudhir and Tewari (2015) use a twenty-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 non-traditional “aspirational” goods and brands during adulthood.\textsuperscript{22} Similarly, having a birth year in 1962 or 1978 is a very strong predictor of whether a male Facebook user “likes” the New York Mets in the mid 2000s, implying the user was seven to eight years old when the Mets won a world series (in 1969 and 1986) - an age at which team preferences are typically formed (Stephens-Davidowitz, 2017).

Bronnenberg, Dubé, and Gentzkow (2012) match current and historic brand market share data across US cities.\textsuperscript{23} These data confirm that consumers brands (1) had very different shares across regions in the 1950s and 1960s, and (2) that the local market leaders of the 1950s and 1960s remain dominant in their respective markets today. In practice, decades-long panels are difficult to maintain and rarely available for research purposes. Therefore, the within-household shopping purchase information is too short to learn about the formation of preferences. Instead, Bronnenberg, Dubé, and Gentzkow (2012) surveyed over 40,000 US households to learn the migration histories of each household’s primary shopper, including her birth market, year of move and her age. They exploit the historic migration behavior of households and the long-term regional differences in brand preference to study the long-term formation of brand preference and loyalty.

Studying the two top brands across 238 product categories, Bronnenberg, Dubé, and Gentzkow (2012) document two striking regularities. First, immediately after a migrant family moves, 60% of the difference in brand shares between the state of origin and current state of residence is elimi-

\textsuperscript{22}These aspirational goods consist primarily of western brands consumed socially. 
\textsuperscript{23}The current brand shares were collected through AC Nielsen’s scanner data. The historic brand shares were obtained from the Consolidated Consumer Analysis (CCA) database, collected by a group of participating newspapers from 1948 until 1968 in their respective markets. The CCA volumes report the fraction of households who state that they buy a given brand in a given year.
nated. This finding holds both within and between households, suggesting that a significant portion of brand preferences is determined by the local choice environment. Second, the remaining 40% of the preference gap is very persistent, with migrants exhibiting statistically significant differences in brand preference than non-migrants even 50 years after moving.

Collapsing the data by age cohorts, Bronnenberg, Dubé, and Gentzkow (2012) find that “migrants who moved during childhood have relative shares close to those of non-migrants in their current states, while those who move later look closer to non-migrants in their birth states.” This finding is consistent with the brand capital stock theory whereby older migrants, having accumulated more brand capital in their birth state, should exhibit more inertia in brand choice.\footnote{An alternative hypothesis is that the aging process makes working memory decline more than long-term memory (Carpenter and Yoon, 2011), as does processing of information. Both aging effects favor relying on fewer considered options (John and Cole, 1986) and engaging in fewer product comparisons Lambert-Pandraud, Laurent, and Laperonne (2005). These factors are thought to contribute to persistence, or at least less flexibility, of purchasing patterns among aging consumers (see also Drolet, Schwarz, and Yoon, 2010).} Even migrants that moved before the age of 6 exhibit some persistence in the local preference from the birth location, suggesting a role for some inter-generational transfer of brand preferences. 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.” In terms of our model in equation (12), approximately 40% of consumers’ expected conditional indirect utility derives from $F_j (H^h_t; \Lambda^h)$ and 60% from $\mu_j (X^h_t; \Theta^h)$. Consistent with these findings, Atkin (2013) reports a similar long-term habit formation for nutrient sources from different foods.

Bronnenberg, Dubé, and Gentzkow (2012) formulate a simple model of habit formation (e.g., Pollak, 1970; Becker and Murphy, 1988) in which individual households’ brand choices depend
on current marketing and prices, as well as a stock of past consumption experiences. Assuming (1) that consumers did not move across state lines motivated by their preferences for CPG brands and (2) that a brand’s past local market share is on average equal to the share today, they determine that the effects of past consumption are highly persistent and depreciate at a rate of only 2.5% per year. Thus, they find that the half-life of brand capital is 26.5 years.

In sum, a large body of empirical work has documented patterns of persistence in brand preferences and choices. This persistence has been documented both at a high-frequency from “shopping trip to shopping trip” as well as at a much lower frequency spanning decades and even entire life-times. If consumers do indeed form strong attachments to brands, then understanding the mechanisms through which these attachments are formed will likely point to some of the important drivers of the industrial organization of consumer goods’ markets.

### 3.3 Brand choice inertia, switching costs and loyalty

Switching costs constitute one of the simplest theories of brand loyalty:

“A product exhibits classic switching costs if a buyer will purchase it repeatedly and find it costly to switch from one seller to another.” Klemperer (2005, p. 125)

Switching costs can be financial, such as the early termination fee for a mobile phone service contract, temporal, such as the time required to learn how to use a new product, or psychological, such as the cognitive hassle of changing one’s habit. Switching costs introduce frictions that can deter a consumer from switching to different brands and, hence, can lead to repeat-purchase

---

25 At least since Mittelstaedt (1969), consumer psychologists have studied the role of psychological switching costs in explaining repeat-purchase and inertia in brand choice. For an extensive review of this literature see (Muthukrishnan, 2015).
behavior. In the extreme case where switching costs are infinite, a consumer’s initial choice would determine her entire future brand choice sequence and the impact of $\mu \left( X_t^h; \Theta^h \right)$ would be zero. Consequently, switching costs can create brand loyalty even in the absence of any brand value other than the identifying feature of the brand name to a specific supplier. This behavior points to a simple theory of branding whereby the identifying features of a supplier’s product (i.e., the “mark”) can be sufficient to create loyalty in consumer shopping behavior as long as consumers form shopping habits. This type of loyalty is also detectable in shorter panels spanning one or two years.

The empirical brand choice literature typically allows for a brand switching cost to influence purchase decisions in consumer goods categories (e.g., Jeuland, 1979; Guadagni and Little, 1983; Jones and Landwehr, 1988; Roy, Chintagunta, and Haldar, 1996; Keane, 1997; Seetharaman, Ainslie, and Chintagunta, 1999; Shum, 2004; Osborne, 2008; Dubé, Hitsch, and Rossi, 2010). Suppose we define a household $h$’s indirect utility net of switching cost to be $\mu_j^h (X_t) = \gamma_j^h + \theta^h x_{jt} + \varepsilon^h_{jt}$. Following the convention in the brand-choice literature, we assume that consumers obtain a utility premium from repeat-buying the brand chosen previously: $F_j (H_t^h; \Lambda^h) = \lambda^h \mathbb{1}_{\{H_t^h = j\}}$ where $H_t^h \in \{1, \ldots, J\}$ is household $h$’s loyalty state and $\mathbb{1}_{\{H_t^h = j\}}$ indicates whether the previous brand purchased was brand $j$. The conditional indirect utility on trip $t$ is then

$$ v_{jt}^h = \mu_j^h (X_t^h; \Theta^h) + F_j^h (H_t^h; \Gamma^h) , \ j = 1, \ldots, J $$

$$ = \gamma_j^h + \theta^h x_{jt} + \lambda^h \mathbb{1}_{\{H_t^h = j\}} + \varepsilon_{jt}^h \hspace{1cm} (13) $$

This formulation nests the basic static model from section 2.1 with the baseline brand utility, $\gamma_j^h$. 

33
The additional parameter $\lambda^h_j$ allows for inertia in brand choices across time. As discussed above, the structural interpretation of $\lambda^h$ is typically analogous to a psychological switching cost. The following null hypothesis constitutes a test for choice inertia:

$$H_0: \mathbb{E}(\lambda^h) = 0,$$

where $\mathbb{E}(\lambda^h) > 0$ implies a positive inertia in brand choice (such as loyalty) and $\mathbb{E}(\lambda^h) < 0$ implies a negative inertia (such as variety-seeking). In practice, the researcher can specify a more general specification that relaxes both the linearity assumption and allows for higher-order choice behavior with a loyalty state that reflects the entire choice history. For instance, Keane (1997) and Guadagni and Little (1983) use a stock variable constructed as an exponential, smooth, weighted average of past choices. While most studies of brand loyalty assume consumers are myopic, a rational forward-looking consumer would plan her future loyalty, much like the rational addiction models fit to tobacco products (e.g., Becker and Murphy, 1988; Gordon and Sun, 2015).

Since most of the brand choice literature pools choice sequences across households, a concern is that state dependence captured by $\lambda^h$ may be spuriously identified by unobserved heterogeneity in tastes between households (Heckman, 1981). Even after rich controls for persistent unobserved heterogeneity and serial dependence in $e^h_{jt}$, Keane (1997) and Dubé, Hitsch, and Rossi (2010) find statistically and economically significant state dependence in choices. However, the magnitude of the state dependence, $\lambda$, is considerably smaller after the inclusion of controls for heterogeneity, falling on average by more than 50%. For instance, in a case study of refrigerated

---

When the researcher does not observe consumers’ initial choices, an “initial conditions” bias can also arise from the endogeneity in consumers’ initial observed (to 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.
orange juice purchases, Dubé, Hitsch, and Rossi (2010) estimate switching costs that, on average, are 21% of the magnitude of prices. Without controls for heterogeneity, these costs are inferred to be more than double.\footnote{Dube, Hitsch, Rossi, and Simonov (2018) find even larger magnitudes of switching costs when they control more formally for endogenous initial conditions (i.e., the initial loyalty state for each household).}

In addition to controlling for heterogeneity, Dubé, Hitsch, and Rossi (2010) also test between several alternative sources of structural state dependence, such as price search and learning. Intuitively, state dependence through consumer learning should dissipate over time as consumers learn through their purchase and consumption decisions. In contrast, switching costs create a persistent form of inertia in choices. We discuss the mechanism through which product quality uncertainty and consumer learning can generate inertia in consumer brand choices below in section 3.4. Dubé, Hitsch, and Rossi (2010) conclude that the inferred brand switching costs are robust to these alternative specifications and that the estimated values of $\lambda$ reflect true brand loyalty.\footnote{Using a structural model of consumer learning, Osborne (2011) also finds empirical evidence for both learning and switching costs.}

Similarly, imperfect information about prices or availability could also create state dependence in the purchase of a known brand. In-store merchandizing, such as displays, should offset the costs of determining a brand’s price in which case inertia for a given brand would be offset by a display for a competing brand. Dubé, Hitsch, and Rossi (2010) again find that their estimates of switching costs are robust to controls for search costs.\footnote{Using a structural model of search and switching costs, Honka (2014) also finds empirical evidence for both search and switching costs. However, in her case study of auto insurance, search costs are found to have a larger effect on choices than switching costs.}

Interestingly, Keane (1997) and Dubé, Hitsch, and Rossi (2010) estimate economically large and heterogeneous brand intercepts, $\gamma_j$. On average, the persistent differences in households’ brand tastes appear to be much more predictive of choices than the loyalty arising through $\lambda$. In a case
study of 16 oz tubs of margarine purchases, Dubé, Hitsch, and Rossi (2010) find the importance weights for loyalty (\( \lambda \)), price (\( \theta \)) and brand (\( \gamma \)) are 6.4%, 53.6% and 40% respectively.\(^{30}\) Therefore, switching costs alone do not seem to explain the persistent consumer brand preferences typically inferred through CPG shopping panels. In sum, while there is a component of consumer switching that reflects dynamics related to loyalty, a large portion of consumers’ brand choices seem to reflect a far more persistent brand taste that is invariant over the time horizons of 1-2 years typically used in the literature.

A limitation of this literature is the lack of a deeper test of the underlying mechanism creating this persistence in choices. As early as Brown (1952, p. 256), scholars have questioned whether inertia in brand choice “is a 'brand' loyalty rather than a store, price or convenience loyalty.” The subsequent access to point-of-sale data allows researchers to control for prices and other causal factors at the point of purchase. But, within a store a buying habit could merely reflect loyalty to a position on the shelf or other incidental factor that happens to be associated with a specific brand. In addition to unobserved sources of loyalty at the point of sale, the persistent brand tastes may also contain additional information about longer-term forms of brand loyalty, such as evolving brand capital stock (e.g., Bronnenberg, Dubé, and Gentzkow, 2012), that are not detectable over one or two-year horizons. While these distinctions may not matter for predicting choices over a short-run horizon, they have important implications for a firm’s willingness to invest in branding or consumer-related marketing to cultivate the shopping inertia.

\(^{30}\)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 13 has three components to utility: brand, marketing variables and loyalty with respective part-worths (or marginal utilities) \( PW_{\text{brand}}(\text{brand} = j) = \lambda_j - \min(0, \{\lambda_k\}_{k=1}^J) \), \( PW_{\text{marketing}}(X_{jt} = x) = \alpha(x - \min(x)) \) and \( PW_{\text{loyalty}}(s_{jt} = j) = \gamma \). Dubé, Hitsch, and Rossi (2010) then assign an importance weight to each of these components, scaled to sum to one, as follows: 

\[
\begin{align*} 
IW_{\text{brand}} &= \frac{\max(PW_{\text{brand}})}{\max(PW_{\text{brand}}) + \max(PW_{\text{price}}) + \max(PW_{\text{loyalty}})}, \\
IW_{\text{price}} &= \frac{\max(PW_{\text{price}})}{\max(PW_{\text{brand}}) + \max(PW_{\text{price}}) + \max(PW_{\text{loyalty}})}, \\
IW_{\text{loyalty}} &= \frac{\max(PW_{\text{loyalty}})}{\max(PW_{\text{brand}}) + \max(PW_{\text{price}}) + \max(PW_{\text{loyalty}})}.
\end{align*}
\]
3.4 Learning from experience

A more complex form of state-dependence in brand choices arises when a consumer faces uncertainty about aspects of product quality that are associated with the brand and that are learned over time. Following Nelson (1970), we modify the characteristics approach to demand by distinguishing between “search characteristics,” which can be determined prior to purchase, and “experience characteristics,” which are determined after the purchase through trial and consumption. The classification of brand as a search versus experiential characteristic is complicated. One the one hand, the brand name as an identifying mark acts like a search characteristic since it is likely verifiable prior to purchase through its presence on the packaging. On the other hand, intangible aspects of the product that are associated with the brand constitute experience characteristics that are learned over time through consumption (Meyer and Sathi, 1985) and informative advertising (Keller, 2012, Chapter 2). This view is consistent with the product-based associations that constitute part of a consumer’s brand knowledge (Keller, 2012, Chapter 2). We focus herein on rational models of consumers using Bayes’ rule to update their beliefs about products over time and to learn.\(^{31}\)

Suppose the consumer is uncertain about the intrinsic brand quality in any period \(t\). At the start of each period, the consumer has a prior belief about brand quality, \(f_{jt}(\gamma)\). At the end of each period, she potentially receives a costless, exogenous, unbiased and noisy signal about brand \(j\), \(s_{jt} \sim g_{j}(\cdot|\gamma)\). For example, the signal might reflect a free sample, word-of-mouth (Roberts and Urban, 1988), observational learning from peers’ choices (Zhang, 2010) or an advertising message (Erdem and Keane, 1996). The consumer then uses the signal to update her beliefs about the

\(^{31}\)The Bayesian learning model predicts that consumers eventually become fully-informed about a brand. However, lab evidence suggests that “blocking” may prevent a consumer 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).
brand’s quality using Bayes’ Rule: $f_{j(t+1)}(\gamma) \equiv f_{jt}(\gamma|s_{jt}) \propto g_j(\cdot|\gamma)f_{jt}(\gamma)$. In this case, the state variable tracking consumer brand experiences, $H_t$, consists of her prior beliefs about each of the brand qualities, $\gamma_j$: $H_t = (f_{1t}(\gamma), ..., f_{Jt}(\gamma))$.\footnote{In principle, one could allow for correlated learning as in (Dickstein, 2018).} We use a discrete-choice formulation of demand, as in section 2.1. If the consumer’s brand choice is made prior to receiving the signal, her expected indirect utility from choosing brand $j$ at time $t$ is

$$E(u_{jt}|H_t; \theta) = \int \left( \lambda (\gamma - \rho \gamma^2) + x_j \beta - \alpha p_{jt} + \epsilon_{jt} \right) g_j(s|\gamma)f_{jt}(\gamma) \, d(s, \gamma)$$

(15)

where $\rho > 0$ captures risk aversion. As discussed in Erdem and Keane (1996) and Crawford and Shum (2005), risk aversion is essential for predicting inertia in consumer’s choices for familiar brands since a consumer may be reluctant to purchase a new brand with uncertain quality. The vector $\theta$ contains all the model parameters, including those characterizing the consumer’s beliefs.

We can augment the model in (15) to allow the consumer to learn over time through her own endogenous brand choices (Erdem and Keane, 1996). Suppose each time the consumer purchases brand $j$, her corresponding consumption experience generates an unbiased, noisy signal about the quality of brand $j$: $s_{jt} \sim g_j(\cdot|\gamma)$. Let $D_{jt}$ indicate whether the consumer purchased brand $j$ at time $t$. To simplify, we follow the convention in most of the literature and assume the consumer’s initial period prior is $f_{j0}(\gamma) = N(\gamma_{j0}, \sigma_{j0}^2)$ and that her consumption signal in a given period is $s_{jt} \sim N(\gamma_{jt}, \sigma_{jt}^2)$. The advantage of this Normal Bayesian Learning model is that the consumer’s state now consists of the beginning-of-period prior means and variances for each of the $J$ brands, $H_t = (\gamma_{1t}, ..., \gamma_{jt}, \sigma_{1t}^2, ..., \sigma_{jt}^2)$, rather than each of the $J$ Normal prior distributions, $f_{jt}(\gamma) = N(\gamma_{jt}, \sigma_{jt}^2)$. In addition, under Normal Bayesian learning the consumer’s period $t$ prior mean and variance
for brand \( j \) can be written in closed form as
\[
\gamma_{jt} = \left( \frac{1}{\sigma^2_{j_0}} + \frac{\sum_{t<\tau} D_{jt}}{\sigma^2_{s}} \right)^{-1} \left( \frac{\gamma_{j_0}}{\sigma^2_{j_0}} + \frac{\sum_{t<\tau} s_{jt}}{\sigma^2_{s}} \right) \quad \text{and} \quad \sigma^2_{jt} = \left( \frac{1}{\sigma^2_{j_0}} + \frac{\sum_{t<\tau} D_{jt}}{\sigma^2_{s}} \right)^{-1}
\]
respectively (DeGroot, 1970). Intuitively, the precision of the signal \( \left( \frac{1}{\sigma^2_{s}} \right) \) determines the information content of the signal and, accordingly, the rate at which the consumer learns. Under this Normal Bayesian learning model with endogenous learning, we can now rewrite the consumer’s expected indirect utility from choosing brand \( j \) at time \( t \) from equation (15) as follows:

\[
\mathbb{E}(u_{jt} | \gamma_{jt}, \sigma^2_{jt}, \theta) = \lambda \gamma_{jt} - \lambda \rho \gamma^2_{jt} - \lambda \rho (\sigma^2_{jt} + \sigma^2_{s}) + x_j \beta - \alpha p_{jt} + \epsilon_{jt}.
\]  

(16)

Erdem and Keane (1996) also allow for consumer learning through brand advertising. As above, we can allow for the consumer to obtain a noisy signal about the quality of brand \( j \) each time she sees a brand \( j \) advertisement: \( s_{jt, ad} \sim g_{jt}^{ad} (\cdot | \gamma) \). The learning process through advertising is then derived analogously to the consumption experiences discussed in the previous paragraph.

The state dependence in demand created through a consumer’s learning stock introduces a dynamic incentive into consumers’ purchase and brand choices. To simplify, assume learning only arises through purchase and consumption. Accordingly, \( f_{H_t+1} (H_{t+1} | H_t, j) \) denotes the Markov process associated with the evolution of the learning state variable based on past priors and the past purchase choice \( j \in \{0, 1, ..., J\} \), where \( j = 0 \) indicates non-purchase during the previous trip. Assuming the consumer discounts her future utility at a rate \( \delta \in (0, 1) \), the value function associated with her purchase decision problem during period \( t \) is

\[
v(s_t, \epsilon_t) = \max_j \left\{ v_{jt} (H_t; \theta) + \epsilon_{jt} \right\}
\]

(17)
where $\varepsilon_{jt} \sim \text{i.i.d. EV}(0, 1)$ is a stochastic term known to the household at time $t$ but not to the analyst. $v_j(H_t; \theta)$ is the choice-specific value function associated with choosing product $j$ in state $H$

$$v_j(H_t; \theta) = \mathbb{E}(u_j|H_t; \theta) + \delta \int v(H_{t+1}, \varepsilon)f_{H_1}(H_{t+1}|H_t, j)f_{\varepsilon}(\varepsilon)d(H_{t+1}, \varepsilon).$$

(18)

If $\delta > 0$, the consumer may optimally experiment with unfamiliar brands in the current period to invest in reducing her future choice uncertainty.

Like the switching costs model, unobserved heterogeneity in tastes could spuriously identify learning. In addition, the learning model requires the assumption that the consumers’ initial beliefs are known to the analyst, similar to the initial conditions assumption in a model with loyalty. Typically, researchers use the restriction that consumers all have a common prior beliefs distribution. This restriction is at odds with past work using self-reported survey measures of uncertainty about brand quality and reliability that are found to be correlated with actual brand purchase choices and with brand purchase intentions (e.g., Erdem and Swait, 1998; Erdem, Swait, and Louviere, 2002). In addition to prior restrictions, these models require considerable computational complexity when a consumer who learns through her endogenous consumption decisions is forward-looking, in which case her current purchase decision is the solution to a dynamic decision problem as in (17) and (18) above (e.g., Erdem and Keane, 1996; Ackerberg, 2003; Crawford and Shum, 2005). A concern is that mis-specification in initial beliefs could introduce bias into the statistical estimates, similar to the incidental parameters problem in models with loyalty (e.g., Heckman, 1981; Shin, Misra, and Horsky, 2012).

The empirical evidence of consumer learning about brand qualities has been mixed. Using diary shopping panels, Demsetz (1962) documents suggestive evidence that consumers do eventu-
ally learn about products and that the advantages to branded goods might erode relatively quickly over the course of a few years. In our structural model, the rate of learning is determined by the information content of signals, which is measured by the signal precision. Estimates from structural econometric models of Normal Bayesian learning fit to CPG brand choice purchase panel data typically imply a relatively fast rate of learning that may only require a few purchase/consumption occasions (e.g., Erdem and Keane, 1996; Ackerberg, 2003; Crawford and Shum, 2005). In contrast, Shin, Misra, and Horsky (2012) find considerably slower rates of learning when they calibrate heterogeneous initial-period beliefs using auxiliary survey data. In the context of durable consumer goods, Erdem, Keane, Öncü, and Strebel (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.

Erdem and Keane (1996) also find that an advertising campaign with low-variance information signals can generate brand loyalty. In a study of advertising for a new brand launch, Ackerberg (2001) and Ackerberg (2003) finds a large advertising effect on brand choice for initial trials, and a small and insignificant effect for repeat purchases, suggesting a high information content from advertising and a very rapid rate of learning.

An interesting aspect of the learning models discussed above is that forward-looking consumers have an incentive to experiment with unfamiliar brands as an investment in more informed future choices. As discussed in Ackerberg (2003), the degree to which a consumer should experiment with brands depends not only on her preferences, but also on her discount rate, prior variances, and the intensity with which she can derive signals from other sources, such as advertising.
studies with forward-looking consumers, the future is typically assumed to be discounted by consumers according to a pre-determined, common discount rate. An interesting direction for future research would consist of testing whether consumers are indeed forward-looking and whether they endogenously “plan” their degree of brand uncertainty.

Another related area for future research would consist of studying whether consumers endogenously seek advertising and other informative marketing as another channel through which to resolve their brand uncertainty. To the extent that the consumption of informative marketing is costly (e.g., time consuming or distracting), this would create a dynamic analog of the complementary goods theory of brand advertising (Becker and Murphy, 1993). In the digital era, it should be easier to monitor consumers’ advertising consumption habits.\(^ {33} \)

A final direction for future research would be to adapt the consumer brand equity measures from section 2.1 to account for the dynamics discussed above.

### 3.5 Brand advertising goodwill

Keller (2012, Chapter 2) discusses the ways in which brand advertising can create consumer associations over time that contribute to brand knowledge, and hence perceived utility. When advertising conveys information about the objective characteristics of the product, we treat it as a product association. Even when advertising does not convey objective information, it can still have advertising-based associations that contribute to brand knowledge. For example, BMW uses persuasive advertising to evoke a consumer association with “prestigious” or “for the young at heart.” Carpenter, Glazer, and Nakamoto (1994) discuss how even a meaningless attribute that is

---

\(^{33}\)Bronnenberg, Dube, and Mela (2010) match household purchase data with households’ advertising skipping behavior using digital video recorder (DVR) technology.
unrelated to the quality of the product can be persuasive, especially when a consumer lacks the expertise to verify claims. For instance, Natural Silk shampoo used the slogan “We put silk in a bottle,” whereas this ingredient does not affect the functionality of shampoo in any way. The attribute “silk” is somehow seen as convincingly positive, perhaps because the consumer is confused by “silky” being a desirable trait for hair, not shampoo. This so-called meaningless differentiation may be effective under certain conditions where consumers lack information. As long as the irrelevant attribute can induce a first-time trial, even if only to test the hypothesis that the attribute is in fact irrelevant, that initial usage could have a persistent effect on future purchases.

To the extent that a consumer learns about these subjective attributes through advertising, we can still use the Normal Bayesian learning model in equation (15) since we did not take a stand on the exact interpretation of $E\left(\gamma_j|s^h_j\right)$ and hence which aspects of $\gamma$, objective or subjective, were being learned. Nevertheless, an extant empirical literature on persuasive advertising has often modeled the persistent effects of past advertising (or “carry-over” effects) as a goodwill stock (see Bagwell, 2007b, for a survey). Honka (2014) allows consumer brand utility to be shifted by the consumer’s self-reported advertising recall for that brand. Alternatively, if goodwill is not observed, it can be modeled through the formation of an advertising goodwill stock, $g_{jt}^h$. Following Nerlove and Arrow (1962), we model the goodwill stock using a distributed lag:

$$g_{jt}^h = \phi_j(A_t) + \rho g_{jt-1}^h, \ j = 1, \ldots, J$$

where $A_t$ is a vector of brand advertising levels in period $t$, $\phi_j(A_t)$ is brand $j$’s added goodwill from current advertising, and $\rho \in [0, 1]$ is the rate at which goodwill depreciates over time. In this case, the consumer’s brand experience state consists of her goodwill, $H_t^h = g_t^h$. State dependence
in demand arises through the cumulative effect of past advertising.

Exploiting a clever discontinuity around advertising differences between the borders of adjacent designated market areas, Shapiro (2017) conducts a case study of anti-depressant drugs and finds that brand advertising primarily expands the category as opposed to shifting share away from competitors. This finding suggests that advertising goodwill could exhibit spillovers that benefit rival brands.

An extensive empirical literature has attempted to estimate the long-run effects of advertising on demand using aggregate data (e.g., Leone, 1995). A detailed discussion about the econometric challenges with identifying consistent advertising effects is beyond the scope of this survey. A recent literature has found that the ability to generate advertising goodwill may depend on the exact scheduling of the ads and the temporal spacing between exposures (e.g., Dubé, Hitsch, and Manchanda, 2005; Sahni, 2012).

The structural interpretation of the effect of advertising in these models is unclear since the goodwill stock, $g_t$, may simply be a reduced-form of the learning processes described in the past section. For instance, in a study of advertising for a new brand launch, Ackerberg (2001) finds almost no advertising effect after a household's first trial of the new product, suggesting that advertising may not have a persuasive effect in this case. Alternatively, $g_t$ may be proxying for the cumulative effect of past advertising on brand awareness and consideration, creating a dynamic version of the model in section 2.2.
4 Brand value to firms

4.1 Brands and market structure

CPG industries provide an excellent context in which to examine brand value to firms. As discussed earlier, the leading CPG brands are often differentiated by little other than the brand itself.

In spite of the physical homogeneity between branded CPG products in a given category, cross-country comparisons for a given consumer goods industry often reveal striking differences in market structure and brand shares for the same set of leading national brands (Adams, 2006; Sutton, 1991). Similar differences have been documented across metropolitan areas and regions within a large country like the US. Bronnenberg, Dhar, and Dubé (2007, 2011) and Bronnenberg and Dubé (2017) provide a detailed description of the geographic patterns in CPG brand shares across major US cities using a large, longitudinal database. The findings consistently show high levels of market concentration. The average four-firm concentration ratio (C4) is 79%, the 25th percentile one-firm concentration ratio (C1) is almost 30%, and the 25th percentile C4 is 66%. Although not reported in their tables, there is little systematic difference in the concentration patterns across their categories.

In addition to high concentration ratios, the identity of the firms represented in C1 and C4 differs remarkably across markets. Using data from the top 2 brands (by revenue) across markets and months, the interaction between brand and market explains almost 70% of the share variation in a typical CPG category. The nationally leading CPG brand has local market shares that vary by 41%

---

34The data span about 35,000 stores with formats including supermarket, drug, mass and convenience located across 76 Scantrack markets. The data track weekly retail sales and prices at the level of the individual Universal Product Code (UPC) 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 & Beauty, Non-Food Grocery, and Packaged Meat.
share points, e.g., whereas in some markets shares could be as low as 9%, in others they may be 50%. These findings indicate large and persistent differences in the market structure of a product category across U.S cities. Bronnenberg, Dubé, and Gentzkow (2012) extend this persistence by comparing brand market shares in the mid twentieth century to shares in the early 2000s. A pooled regression of the historic share level on the current share level fails to reject (with reasonable precision) an intercept of zero and slope of one, leading the authors to conclude “that the best predictor of a past purchase share given the data we observe is the present purchase share.” Collectively, these findings indicate that the presence of brands and the role of brand history are central to the formation of market structure in consumer goods industries.


The high concentration in consumer goods industries has also recently been flagged as a cause for public policy concern, and part of a broader upward trend in concentration across US industries. Interestingly, concentration in CPG industries seems to be a historic phenomenon that has spanned decades, if not a century. Whether CPG industries are exceptional or not requires further investigation. But, it is clear that CPG concentration is not a recent development.

In fact, the dominance of national CPG brands has begun to erode slowly in recent years due to the growth of private labels. In the US, CPG private label expenditure shares remain modest, below

---

35 They match CPG brand shares across geographic markets from 2006-2008, based on Nielsen Homescan data, and from 1948-1968, based on the Consolidated Consumer Analysis (CCA) database. The CCA was compiled by aggregating consumer surveys conducted by newspapers nationwide from 1948 to 1968.

36 See for instance “America’s Concentration Crisis” by the Open Markets Institute at https://concentrationcrisis.openmarketsinstitute.org (accessed on 2-7-2019).
20%. But, expenditure growth followed a persistent trend of 0.45 share points per year between 2004 and 2012 (Dubé, Hitsch, and Rossi, 2018). In Europe, private label shares are much higher, exceeding 40% of spending in Switzerland, Spain and the UK. The Nielsen Company (2014). Bronnenberg, Dubé, Gentzkow, and Shapiro (2015) find that private label shares are especially high among consumers with high domain-specific expertise: pharmacists buy over 90% private label head-ache medicines and chefs by considerably more private label pantry food staples, suggesting that information barriers may play a role in the sustained advantages of large, national brands.

4.2 Measuring brand value

Echoing Bain (1956)’s view of brand preference as a critical barrier to entry, Goldfarb, Lu, and Moorthy (2008) observe that “brand equity is perhaps the single most important asset that marketing contributes for a firm.” Perhaps reflecting this importance, an industry has emerged to estimate the value of brands, with leading providers including Interbrand, BrandFinance, and Kantar Millward Brown. According to Interbrand, the top-100 most valuable global brands represented a joint value in excess of $1.8 trillion in 2017. Historically, these brand valuation estimates have been too unreliable to be recognized as an intangible asset according to the U.S. Generally Accepted Accounting Principles (GAAP) (see, e.g., Barth, Clement, Foster, and Kasznik, 1998). Simon and Sullivan (1993)’s approach comes closest in spirit to the methods typically used by practitioners. Starting with a firm’s market value, Simon and Sullivan (1993) first remove the component due to tangible assets (plant, machinery, real-estate, etc.) by using 10-K financial data on the replacement-value of the firm’s tangible assets. The residual from this calculation yields the total value of intangible assets. Since many intangible assets can contribute to this total—brand equity,

R&D, protection from competition, etc.—the remaining task consists of isolating the brand equity through a regression of total intangible value on brand-related variables (e.g., advertising expenditure, age of the company, order of entry, and advertising share), and non-brand-related variables (e.g., patent share, four-firm concentration ratio, presence or absence of regulation, etc.). The estimated regression coefficients for the brand-related variables are next used to construct a measure of brand value.

This top-down approach to brand value estimation has obvious limitations, notably several strong assumptions. For example, Simon and Sullivan (1993) assume that “market share due to brand equity” is a function of order of entry and advertising share. This assumption excludes many other brand-related variables, such as product quality, which may also affect market share. This method would yield a low brand value for a company like Google because it wasn’t the first to enter the search-engine market, and it is not known to have a large advertising share. The regression component also requires strong and potentially implausible exogeneity assumptions when firm-determined variables like advertising share are included on the right-hand side of the market-share equation. Finally, being a top-down approach, it can only work for single-brand firms. The approach could not be applied to some of the leading consumers goods companies, like Procter & Gamble or Kraft, which own many brands in their portfolios.

Theoretically, a brand’s value to a firm is “the difference between the profits that it would earn operating with the trademark and the profits it would earn operating without the trademark” (Rubin, 1998, Chapter 4). The measurement calls for both specifying a counterfactual—consumers’ perceptions of product quality in the absence of the brand—and measuring product performance in the counterfactual. Since brand equity reflects brand’s role in fostering expectations about product performance, as well as brand’s role in creating consumption value directly, specifying the coun-
terfactual requires making assumptions about which associations remain active and which become inactive when the product loses its brand. In practice, this counterfactual is seldom observed in the data. Rubin (1998, ch. 4) discusses several demand-side approaches, all of which require ceteris paribus assumptions. In reality, a firm would likely change its pricing and marketing in the absence of brand, and so will its competitors, which calls for a re-computation of the entire market equilibrium. Some empirical approaches to measuring brand value take a reduced-form approach, completely ignoring the supply-side changes in the counterfactual, while others, because they are structurally grounded, recognize those effects (at the cost of complexity). Structural approaches also require the researcher to specify the underlying assumptions more transparently and to be more precise about the mechanisms through which brands influence demand and create incremental profits. We will examine the reduced-form approaches first before considering the structural models.

4.2.1 Reduced-form approaches using price and revenue premia

The simplest implementation of the price- and revenue-premium approach consists of using an objectively identical, unbranded “generic” product as the counterfactual comparison for the branded product. The price or revenue differential between branded and generic can then be interpreted as the “brand premium.” For instance, using over-the-counter drugs as a category in which biologically equivalent private-label products exist, Bronnenberg, Dubé, Gentzkow, and Shapiro (2015) document large brand price premia for branded drugs. Their data suggest that consumers would spend $44 billion less on CPG categories if they all switched to a store-branded substitute. Similarly, price premia for products differentiated by little other than brand name have been documented in S&P index funds, (Hortacsu and Syverson, 2004), Mexican privatized social security
fund (Hastings, Hortacsu, and Syverson, 2013), and online books (Brynjolfsson and Smith, 2000). Ailawadi, Lehmann, and Neslin (2003) take the calculation one step further by taking into account not only price differences, but also quantity differences. Using the prices and quantities sold of national brands and their private-label equivalents, they compute the revenue premium enjoyed by each brand, and treat that as a measure of brand value. In practice, the incremental price charged or revenue earned by the national brand would need to be measured relative to a counterfactual in which the national brand becomes a generic or private-label brand, reducing the number of national brands by one and increasing the number of private-label brands by one. Depending on the size and market power of the national brand in question, the counterfactual equilibrium would likely change the price and revenue premia associated with brand.

When an appropriate comparable unbranded product is not available, researchers may instead use a hedonic approach that projects prices onto objective attributes and the brand name (Rosen, 1974). The estimated hedonic price function can then be used to predict the price of a product but-for the brand. A decision must still be made about the set of relevant product attributes to include in the regression. Wiggins and Raboy (1996) conduct a hedonic analysis of branded and unbranded bananas sold to retailers. Due to the lack of retail price variation, they focus on the brand premium in wholesale prices. After controlling for shipping method (containerized or palletized), region being shipped to, and year of shipment, they find that two companies, Noboa and Turbana, sell their bananas at a discount relative to Del Monte. However, the lack of variation in search attributes leads to an attribution problem. The authors are unable to determine whether this difference reflects low brand equity or the fact that the two companies also happen to source their bananas from a single country. Holbrook (1992) faces a similar issue in his hedonic analysis of the prices for home electronics which requires pooling brands across categories—e.g., receivers, CD players,
and car amplifiers—requiring the strong assumption that brand equity is common across these various categories. In general, the hedonic approach will not be able to disentangle the role of a given brand and the role of any search attributes that are only supplied by that brand.

### 4.2.2 Structural models

In contrast to the reduced-form approaches above, more recent research has used a structural approach that compares the market equilibrium both in the factual and in the counterfactual environments. Our discussion herein closely follows the approaches used in Goldfarb, Lu, and Moorthy (2008) and Borkovsky, Goldfarb, Haviv, and Moorthy (2016).

Let \( j = 1, \ldots, J \) index the brands in a product market. We assume that each brand is owned and marketed independently by a separate, single-product firm. Following the discrete-choice, characteristics approach in section 2.1, consumer demand for each product \( j \) is given by \( Q_j(p, x, \Theta) = N \cdot \Pr(j; p, x, \Theta) \), where \( \Pr(j; p, x, \Theta) \) is brand \( j \)'s market share, \( N \) is total mass of the market (i.e. the exogenous potential market size) and \( 1 - \sum_{j=1}^{J} \Pr(j; p, x, \Theta) \) is the unserved share of the market. As before, \( p \) denotes the vector of prices, \( x \) is a matrix of product characteristics, and the first \( J \) rows of \( \Theta \) contain the brand equities, \((\gamma_1, \ldots, \gamma_J)\).

Goldfarb, Lu, and Moorthy (2008) fit their model to breakfast cereal data. Since consumers of cereals can inspect the product’s search attributes—such as the cereal type (e.g., wheat, rice, raisin bran, etc.) and nutritional profile—before purchase, they assume that brand equity in the cereal category derives only from (i) creating expectations about the product’s experience attributes (e.g., taste, “crunch,” etc.), and (ii) providing brand-based consumption utility. In the expression above, therefore, they interpret \( x \) as the product’s search attributes and subsume the product’s experience...
attributes into the brand equity, $\gamma_j$.\textsuperscript{38}

In this static framework, the counterfactual brand value for brand $k$ is measured by setting $\gamma^h_k = 0$ for each consumer. Since firms in the industry would likely re-adjust their prices in response to the elimination of one firm’s brand equity, the counterfactual also requires computing the counterfactual equilibrium outcomes. Suppose that firms have constant marginal costs $c_j$ and compete on prices each period to maximize profitability, leading to the standard Bertrand-Nash equilibrium. Further, assume the equilibrium exists and is unique. In the factual environment where brand $k$ owns the commercial rights to the brand $k$, the equilibrium prices $p^0$ satisfy the following first-order necessary conditions:

$$Q_j(p^0, x, \Theta) + (p^0_j - c_j) \frac{\partial Q_j(p^0, x, \Theta)}{\partial p_j} = 0. \tag{19}$$

In the counterfactual, the parameter vector $\Theta_{-k}$ denotes setting the marginal population distribution of $\gamma_k$ to a point mass at zero. In this counterfactual, the equilibrium prices $p^1$ satisfy the following first-order necessary conditions:

$$Q_j(p^1, x, \Theta_{-k}) + (p^1_j - c_j) \frac{\partial Q_j(p^1, x, \Theta_{-k})}{\partial p_j} = 0. \tag{20}$$

Brand value to firm $k$ can then be computed by comparing the factual profits, $\pi_k(p^0) = (p^0_k - c_k) Q_k(p^0, x, \Theta)$, and its counterfactual profits, $\pi_k(p^1)$, such that:

$$BV_k = \pi_k(p^0) - \pi_k(p^1). \tag{21}$$

\textsuperscript{38}This approach also relies on the additive structure of consumer utility and rules out the possibility that brand knowledge includes any associations with specific objective characteristics. For instance, for a given product characteristic $i$, $\frac{\partial \gamma^h_i}{\partial x^h}$ cannot arise through consumer $h$'s brand association with that characteristic.
This measure requires the additional assumption that both $c_k$ and fixed costs, $F_k$, are invariant to branding. Of these, the assumption of constant fixed costs is more problematic as it abstracts from any investments a firm might make to maintain the brand, such as advertising.

In contrast to this static framework, Borkovsky, Goldfarb, Haviv, and Moorthy (2016) take a price-theoretic approach that treats consumer brand equity as an intangible goodwill stock that evolves in response to firms’ endogenous brand advertising decisions. As in section 3.5, advertising goodwill depreciates slowly over time so that a firm with brand equity could potentially “coast,” easing off on its advertising investments until brand equity has depreciated sufficiently to be viewed as “too low.” This formulation allows the firm to control its brand equity over time through fixed and sunk advertising investments. The carry-over effects of advertising spending on future consumer brand equity create dynamic incentives for the firm which must consider the future implications of current spending.

Borkovsky, Goldfarb, Haviv, and Moorthy conduct a case study of the stacked potato chip industry in the period 2001–2006, when the industry transitioned from a monopoly, with Procter & Gamble’s Pringles brand, to a duopoly after the entry of STAX brand extension of the Lay’s brand. Using demand estimates, they derive quarterly measures of Pringles and STAX’s brand equities as in Goldfarb, Lu, and Moorthy (2008). Firms’ advertising decisions control the evolution of brand equity through the following “quality ladder” formulation, as in Pakes and McGuire (1994):

$$\gamma'_j = \gamma_j + \tau_j(A_j) - t_j + \eta$$

Here $\gamma'_j$ denotes firm $j$’s brand equity at the end of a period and $\gamma_j$ its brand equity at the start of a period; $\tau_j$ is a random variable governed by firm $j$’s advertising $A_j \geq 0$, $t_j \in \{0, 1\}$ is a firm-
specific depreciation shock, and $\eta \in \{-1,0,1\}$ is an industry-wide shock that can cause all brand equities in the industry to increase, decrease or remain unchanged independent of advertising.

These brand equities then become the state variables in a dynamic model of brand management. Firms are assumed to devise state-dependent, quarterly advertising strategies to maximize the discounted sum of their profits where the state consists of their own and competitors’ respective brand equity levels. Let the vector $\gamma = (\gamma_j, \gamma_{-j})$ represent the industry state from firm $j$’s perspective. Then firm $j$’s dynamic brand value in a particular period is

$$v_j(\gamma) \equiv V_j(\gamma) - V_j(0, \gamma_{-j})$$

where $V_j(\gamma)$ and $V_j(0, \gamma_{-j})$ are the discounted expected present values of firm $j$ in the actual state $\gamma$ and in the counterfactual state $(0, \gamma_{-j})$, respectively. Note that this value already takes into account the brand building and rebuilding decisions that a firm and its competitors might expect to undertake in the current period and in the future. From these estimates of brand value, Borkovsky, Goldfarb, Haviv, and Moorthy (2016) assert that the maximum discounted profit potential of a stacked chips brand is $3.36$ billion, which arises when it has the highest possible brand equity and its rival possesses the lowest possible brand equity. To illustrate the difference between the static and dynamic approaches to measuring brand value, Borkovsky, Goldfarb, Haviv, and Moorthy (2016) produce Figure 1. The static brand values of Pringles and STAX are higher than their dynamic values. Moreover, while the static estimates fluctuate a lot, the dynamic estimates are relatively stable. (Both decrease over time because of improvements in the outside good.) These differences arise because the static measure ignores (i) depreciation of brand equity, (ii) advertising’s ability to boost brand equity, and (iii) firms’ forward-looking behavior.
By embedding brand value measurement in a dynamic model of brand management, Borkovsky, Goldfarb, Haviv, and Moorthy’s approach allows us to evaluate how changes in the ability of firms to build and sustain brand equity affect brand value. For instance, they find that at their estimated parametrization, if brand equity were to depreciate more slowly, or if advertising were to become more effective at building brand equity, then brand value would increase. More interesting is the way advertising effectiveness and brand equity depreciation rate interact in determining brand value. In particular, when the effectiveness of advertising is high, there is an inverted-U-shaped relationship between the brand equity depreciation rate and brand value. Brand value is low when the brand equity depreciation rate is low because any brand equity that is lost can be rebuilt relatively easily in a high-advertising-effectiveness environment. As well, brand value is low when the depreciation rate is high, because it is harder to sustain brand equity despite the high advertising effectiveness. In a high advertising effectiveness environment, therefore, a brand is most valuable at intermediate depreciation rates, when brand equity can both be sustained—and thus continue to generate high profits for the firm—and harder to rebuild should it be lost. In short, when industry conditions make it easier to brand build, an increase in the brand equity depreciation rate can increase brand value, even as it reduces firm value.

– insert Figure 1 here –

An alternative structural approach uses stock market data to recover the market value of a firm’s brand. Building on Belo, Lin, and Vitorino (2014), Vitorino (2014) proposes a dynamic structural model of the firm in which firms invest in both physical capital and intangible brand capital. She develops an empirical measure of the stock of brand capital for all publicly traded firm in the U.S. economy that report advertising expenses, and estimates the model using asset price
and accounting data from a large cross section of firms. The estimation results show that brand equity accounts for a substantial fraction of a firm’s market value (about 23%), and that this value varies substantially across industries.

The high irreversible marketing costs associated with the creation and maintenance of a brand make it puzzling why some industries routinely launch new brands that, for the most part, ultimately fail shortly after launch. In a case study of RTE Cereals, Hitsch (2006) rationalizes the high entry and exit rate of new brands as the outcome of optimal experimentation by firms. He shows that, under demand uncertainty, a forward-looking firm may optimally continue to invest in marketing a seemingly unprofitable product to gather more information about its true profitability. The option value associated with being able to scrap the product at a later date enables the firm to invest in experimenting with new products even when the odds of success are low.

5 Branding and firm strategy

In sections 2 and 3, we explored theories of consumer brand equity whereby the mere presence of a brand name or logo shifts demand for the branded good. In this section, we explore economic theories of how firms create brand value on both the supply and demand sides, as equilibrium outcomes. We categorize the branding literature into two broad groups based on the mechanism through which a brand affects consumer demand. The first line of research studies branding as investments in intangible consumption benefits bundled with the product. For instance, the firm might make a marketing investment, such as advertising, to create a favorable brand image in the eyes of consumers. In this context we discuss Sutton’s (1991) work on endogenous sunk costs (ESC), showing how the convexity of the marketing cost function can ultimately shape the
equilibrium industrial market structure of an industry: the variety and quality of brands available to consumers, market concentration, and the incremental profits from branding for firms.

The second line of research studies branding as a strategy to communicate information about product quality in contexts where quality is difficult for consumers to observe before purchase. We organize this discussion into three groups. We begin with brand reputation as a solution to a moral hazard problem. Suppose a firm decides whether to sell a high versus low quality product each period to a customer with imperfect product information prior to purchase. With repeated interaction, a “brand reputation” equilibrium can emerge whereby the consumer “punishes” the firm for supplying low quality by refusing to buy in the future. We next discuss how investments in branding can resolve an adverse selection problem. Suppose that firms pre-commit to selling specific vertically differentiated products and that consumers have imperfect information about product quality prior to their initial purchase. Even when consumers derive no intrinsic value from brand advertising, the advertising expenditure enables a high-quality firm to signal its type to first-time buyers as long as low-quality firms derive insufficient benefit from repeat purchases to justify advertising. Finally, we discuss how umbrella branding can manipulate consumers’ expectations about product quality by linking transactions across products sharing a common brand name.

The distinction between branding as a source of consumption utility versus branding as a device to communicate information about product quality is related to our understanding of market structure. In a commodity market, for instance, there would be no role for informative advertising. The fact that advertising persists in such markets is suggestive of a consumption role of advertising that creates direct utility (e.g., advertising as a characteristics of the product itself). In addition, the distinction between branding theories raises some interesting, albeit unexplored, questions regarding the consumer welfare benefits associated with brand names. Under the consumption utility
view, the brand intercepts create intrinsic value to the consumer. However, under the informative views, brand shifts demand but does not directly create value. In the demand models in section 2.1, the brand intercept is a reduced-form representation of the role of the brand name; these models do not distinguish between the consumption and informational roles. In section 2.2 we come closer to separating these roles. However, to the best of our knowledge, the literature has not yet combined the demand and supply sides in the context of these more elaborate consumer models. When the informational role of brands emerges as an equilibrium signal, researchers need to be careful that this signaling role would be preserved in any counterfactual equilibrium analysis.

5.1 Brand as a product characteristic

A recent literature has analyzed the important relationship between branding investments and the long-term equilibrium market structure in an industry. In a series of empirical case studies, Sutton (1991) observed that many branded consumer goods industries remain concentrated even as the market size grows, an observation seemingly at odds with standard economic conventional wisdom. As we saw in section 2.1, many applications treat a product’s brand as an intangible characteristic that creates utility for the consumer. In section 3.5, we discussed how the marginal utility to the consumer associated with a product’s brand can derive from the goodwill associated with the product’s advertising investments over time.

Formally, suppose we define a consumer $h$’s utility from choosing brand $j$ from a choice set of $J$ branded goods as follows:

$$u(\psi_j, d_j^h, y^h - p_j)$$

where $\psi_j$ captures the perceived quality of brand $j$ (i.e., $u_{\psi} \geq 0$), $d_j^h = |\delta_j - \alpha^h|$ represents the...
horizontal differentiation associated with brand $j$ (i.e., the distance between the product’s position, $\delta_j$, and the consumer’s ideal point, $\alpha_h$), $p_j$ is the price of brand $j$ and $y$ denotes the consumer’s income. Assume that $|u_d|$ and $u_y$ are bounded above. This model is sufficiently general to embody many of the most popular discrete choices models of demand used in the empirical literature on brand choice.

On the supply side, suppose firms play a three-stage game. In the first stage, they decide whether or not to enter a market. In the second stage, they invest in marketing to select their perceived product attribute levels $(\psi_j, \delta_j)$ at a fixed cost $F(\psi)$ where $F(\psi) > 0$, $F_{\psi}(\psi) > 0$ and $\frac{F_{\psi}(\psi)}{F(\psi)}$ is bounded above. The latter assumption ensures that as brand quality levels increase, the incremental costs to raise quality do not become arbitrarily large. In the third stage, firms play a Bertrand pricing game conditional on the perceived product attributes and marginal costs $c(Q; \Psi)$, where $Q$ is the total quantity sold by the firm. If we further assume that $c(Q; \Psi) < \bar{y} < \max (\bar{y}^h)$, where $\bar{y}$ is an upper bound on costs, then “the increase in unit variable cost is strictly less than the marginal valuation of the richest consumer” Shaked and Sutton (1987, page 136). Accordingly, the model bounds above how quickly unit variable costs can increase in the level of quality being supplied.\footnote{Note that we are abstracting away from the richer setting where a firm can invest in marketing over time to build and maintain a depreciating goodwill stock, as in section 4.2.2 above.} At the same time, on the demand side there will always be some consumers willing to pay for arbitrarily large brand quality levels. In other words, costs increase more slowly than the marginal valuation of the “highest-income” consumer.

Seminal work by Shaked and Sutton (1987) and Sutton (1991) derived the theoretical mechanisms through which the manner in which brands differentiate goods, $u_\psi$ and $u_d$, and the convexity of the advertising cost function, $F(\psi)$, ultimately determine the industrial market structure. The
following propositions are proved in Shaked and Sutton (1987):

**Proposition 1.** If $u_\psi = 0$ (i.e. no vertical differentiation), then for any $\varepsilon > 0$, there exists a market size $S^*$ such that for any $S > S^*$, every firm has an equilibrium market share less than $\varepsilon$.

Essentially, in a purely horizontally-differentiated market, the limiting concentration is zero as market size increases. The intuition for this result is that as the market size increases, we observe a proliferation of products along the horizontal dimension until, in the limit, the entire continuum is served and all firms earn arbitrarily small shares.

**Proposition 2.** When $u_\psi > 0$, there exists an $\varepsilon > 0$ such that at equilibrium, at least one firm has a market share larger than $\varepsilon$, irrespective of the market size.

As market size increases for industries in which firms can make fixed and sunk investments in brand quality (i.e., vertical attributes), we do not see an escalation in entry. Instead, we see a competitive escalation in advertising spending to increase the perceived quality of products. The intuition is that a firm perceived to be higher-quality can undercut its “lower-quality” rivals. Hence, the firm perceived to be the highest-quality will always be able to garner market share and earn positive economic profits. At the same time, only a finite number of firms will be able to sustain such high levels of advertising profitably, which dampens entry even in the limit. These two results indicate that product differentiation per se is insufficient to explain concentration. Concentration arises from competitive investments in vertical product differentiation. When firms cannot build vertically-differentiated brands (by advertising) we expect markets to fragment as market size grows. In contrast, when firms can invest to build vertically-differentiated brands, we do not expect to see market fragmentation, but rather an escalation in the amount of advertising and the perseverance of a concentrated market structure.
The crucial assumption is that the burden of advertising falls more on fixed than variable costs. This assumption ensures that costs do not become arbitrarily large (i.e. prohibitively large) as quality increases. Consequently, it is always possible to outspend rivals on advertising and still impact demand. This seems like a reasonable assumption for the CPG markets in which advertising decisions are made in advance of realized sales. It is unlikely that advertising spending would have a large influence on marginal (production) costs of a branded good.

Extending the ESC theory to a setting in which firms make their entry and investment decisions sequentially, strengthens the barriers to entry by endowing the early entrant with a first-mover advantage. With vertical differentiation and endogenous sunk advertising costs, an early entrant can build an even larger brand that pre-empts future investment by later entrants (e.g., Lane, 1980; Moorthy, 1988; Sutton, 1991).

Using cross-category variation, Sutton (1991) provides detailed and extensive, cross-country case studies that empirically confirm the central prediction of ESC theory in food industries, finding a lower bound on concentration in advertising-intense industries but not in industries where advertising is minimal or absent. Bronnenberg, Dhar, and Dubé (2011) find similar evidence for US CPG industries, looking across US cities of differing size, with a lower bound on concentration in advertising-intense CPG industries and a fragmentation of non-advertising-intense CPG industries in the larger US cities. Order of entry is also found to play an important role on market structure, with early entrants sustaining higher market shares even a century after they launched their brands (e.g., Bronnenberg, Dhar, and Dubé, 2009). To understand the important role of the convexity of the marketing cost function, Berry and Waldfogel (2010) show that market concentration fragments and the range of qualities offered rises in larger markets in the restaurant industry, where quality investments comprise mainly variable costs. In contrast, for the newspaper industry,
where quality investments comprise mainly fixed costs, they observe average quality rising with market size without fragmentation.

Sutton (1991) even provides anecdotal evidence linking historical differences in advertising across countries to national market structure and demand for branded goods. For instance, relative to the United States, TV advertising used to be restricted in the United Kingdom. Consistent with advertising as an intangible attribute, Sutton (1991) notes that the market share of private labels is much higher in the UK than in the US.

Interestingly, the theory does not imply that the market be served exclusively by branded goods. When the consumer population consists of those who value the vertical brand attribute and those who do not, it is possible to sustain entry by advertising and non-advertising firms. The latter will serve the consumers with no willingness-to-pay a brand premium. However, as the market size grows, the unbranded sub-segment of the market fragments. In their cross-industry analysis, Bronnenberg, Dhar, and Dubé (2011) observe an escalation in the number of non-advertising CPG products in a given category in larger US cities, with concentration converging towards 0.

5.2 Brands and reputation

Traditionally, the term brand was associated with the identity of the firm manufacturing a specific commodity. As a brand developed a reputation for reliability or high quality, consumers would potentially pay a price premium for the branded good.

A central idea in markets for experience goods and services is that a firm’s reputation matters when consumers have incomplete information about product quality and fit prior to consuming the product (Nelson, 1970, 1974). In such markets, inexperienced consumers can be taken advantage
of by firms selling low quality at high prices. The longer it takes for consumers to discover true product quality, the more beneficial it is for firms to sell low quality at a high price. Perhaps the most straightforward role of brands in such a setting is that they allow consumers to connect one purchase to the next. This connection provides the basis for holding firms accountable for their actions even without a third-party (e.g., government) enforcing contracts. In turn, it provides the basis for “reputations”—how they arise and why they are relevant to consumers.\footnote{We focus herein on the role of brand reputations, referring readers looking for a broader discussion of the literature on reputation to the survey in Bar-Isaac and Tadelis (2008).}

In common parlance, reputation is a firm’s track-record of delivering high quality; in theoretical models, it is consumers’ beliefs about product quality. There is a large literature that ties the provision of product quality to seller identification or lack thereof. On the negative side, Akerlof (1970) shows that non-identifiability of firms in experience goods markets leads to deterioration of quality because low quality firms cannot be punished, and high quality firms rewarded, for their actions.\footnote{“The purchaser’s problem, of course, is to identify quality. The presence of people in the market who are willing to offer inferior goods tends to drive the market out of existence - as in the case of our automobile “lemons.”” (Akerlof, 1970, p. 495)} On the positive side, Klein and Leffler (1981, p. 616) observe that branding enables reputations to form and to be sustained:

“...economists also have long considered "reputations" and brand names to be private devices which provide incentives that assure contract performance in the absence of any third-party enforcer (Hayek 1948, p. 97; Marshall 1949, vol. 4, p. xi). This private-contract enforcement mechanism relies upon the value to the firm of repeat sales to satisfied customers as a means of preventing nonperformance.”

In a competitive market, the identification role of a brand can benefit a firm because “a firm which has a good reputation owns a valuable asset” (Shapiro, 1983, p. 659). For instance, Png and
Reitman (1995) find that branded retail gasoline stations are more likely to carry products and services with important experiential characteristics that could be verifiable through a reputation for quality, such as premium gasoline and repair services. Similarly, Ingram and Baum (1997) report that chain-affiliated hotels in Manhattan had a lower failure rate than independent hotels. In case studies of jeans and juices, Erdem and Swait (1998) find that consumer demand responds to self-reported survey measures of brand credibility.

Klein and Leffler (1981) and Shapiro (1983) examine the incentives firms need to maintain reputations. A simple model illustrates their arguments. Suppose a monopolist firm can offer high (H) or low quality (L) every period, with H costlier to produce than L: \( c_h > c_l \). Suppose also that there are \( N \) consumers in the market, each looking to buy at most one unit of the product. Assume for simplicity that all consumers buy at the same time or, equivalently, that there is instantaneous word-of-mouth from one consumer to all. Consumers’ reservation prices are \( v_h \) and \( v_l \) for products known to be high-quality and low-quality, respectively, with \( v_h > v_l \). Assume \( v_h - c_h > v_l - c_l \), i.e., the firm would prefer to offer high quality if consumers were perfectly informed about quality. The product, however, is an experience good and consumers only observe price before purchase. They observe quality after purchase. We now analyze the firm’s behavior in a “fulfilled expectations” equilibrium, i.e., an equilibrium in which consumers’ expectations about firm behavior match the firm’s actual behavior.

In one such equilibrium, consumers expect low quality in every period and the firm delivers low quality every period. The more interesting equilibrium, however, is one in which the firm delivers a high quality product in every period when consumers expect it to do so. Such an equilibrium is sustained by consumer beliefs that punish the firm for reneging on its “reputation for high quality.” Specifically, consumers’ beliefs are that the firm offers \( H \) unless proven otherwise, in which case,
their future expectations are that the firm will deliver $L$.\footnote{Such a punishment may appear draconian—and we will have more to say about whether real-world firms get punished this way—but for now the theoretical point is that it is the possibility of punishment that provides firms the incentive to maintain reputations.} The existence of this equilibrium requires the firm to have no incentive to deviate from $H$ in any period and, hence, never to offer $L$ in any period. Given a discount factor $\rho$, the payoff along the equilibrium path is

$$\pi_h = \left( \frac{\rho}{1-\rho} \right) (p_h - c_h) N$$

whereas (assuming $v_l > c_l$) the payoff along the deviation path of making a low quality product, but selling it as a high quality product once, is

$$\pi_l = \rho (p_h - c_l) N + \left( \frac{\rho^2}{1-\rho} \right) \max \{0, (v_l - c_l)\} N.$$ 

The reputation equilibrium is sustained, therefore, if and only if

$$p_h \geq p_h^* = \left( \frac{1}{\rho} \right) (c_h - c_l) + \max \{c_l, v_l\}$$

Since we also need $v_h \geq p_h$, this equilibrium is feasible if and only if

$$v_h \geq p_h^* \quad (22)$$

Obviously, if $\rho$, the discount factor, is small enough, this condition cannot hold. On the other hand, if $\rho$ is large enough and $c_h - c_l$ small enough, it is possible to find a $p_h \in (p_h^*, v_h]$. In short, if the firm has a sufficient stake in the future, and consumers are willing to pay a sufficient premium for high quality, then the firm is willing to maintain its reputation for high quality by offering high
quality, foregoing the short-term incentives to “harvest” its reputation. Shapiro (1982) extends a model like this to continuous time, and assumes that both quality and price are continuously adjustable. In addition, he allows for general reputation functions—i.e., reputation functions that do not instantaneously adjust to the last quality provided (as in the model above). The main result is that, with reputation adjustment lags, the firm will only be able to sustain less than perfect-information quality in a fulfilled-expectations equilibrium. Essentially, the firm has to pay a price for consumers’ imperfect monitoring technology when it is coupled with consumer rationality.

Board and Meyer-ter Vehn (2013) extend this framework to long-lived investment decisions that affect quality and consider a variety of consumer learning processes. They find that when signals about quality are more likely to convey “good news” than “bad news,” a high-reputation firm has the incentive to run down its quality and reputation, while a low-reputation firm keeps investing to increase the possibility of good news. Conversely, when signals about quality are more likely to convey bad news, a low-reputation firm has weak incentives to invest, while a high-reputation firm keeps investing to protect its reputation.

In practice, the extent to which reputation incentives discipline firms to deliver persistently high quality is an interesting direction for future empirical research. The recent experience of brands like Samsung, Tylenol and Toyota, which rebounded quickly from product crises suggests that consumers might be forgiving of occasional lapses in quality, even major ones.\footnote{See, for example, “Samsung rebounds with strong Galaxy S8 pre-orders after Note 7 disaster,” \textit{New York Post}, April 13, 2017. \url{https://nypost.com/2017/04/13/samsung-rebounds-with-strong-galaxy-s8-pre-orders-after-note-7-disaster/}.}

A limitation of the reputation literature is the assumption that firms are consumer-facing and can be held accountable for their actions. For this reason, Tadelis (1999) notes that brands are natural candidates for reputations because they are observable, even when the firms that own them
are not. For example, a consumer can hold a restaurant accountable for its performance across unobserved (to the consumer) changes in the establishment’s ownership as long as the restaurant’s name remains the same. Luca and Zervas (2016) find that restaurants that are part of an established branded chain are considerably less likely to commit review fraud on Yelp. They also find that independent restaurants are more likely to post fake reviews on Yelp when they are young and have weak ratings. By the same token, a new brand from an existing firm starts with a clean slate; thus a firm can undo the bad reputation of its existing brand by creating a new brand. An illustration of this point appears in a case study of residential plumbers. McDevitt (2011) finds that firms with a high track record of customer complaints typically changed their names. However, changing names is not costless: as we have noted above, besides the direct costs of doing so—researching names, registering the new name, etc—the more significant expense is the cost of developing awareness of the new brand. Perhaps for this reason, companies whose corporate brands permeate their entire, large, product line—companies such as Ford, Sony and Samsung—inevitably create sub-brands (a) to establish independent identities in multiple product categories, and (b) to insulate the corporate brand at least partially from the individual transgressions of one product. Examples include Mustang for Ford, Bravia for Sony, and Galaxy for Samsung.

The idea that brands serve as repositories for reputation, and provide the right incentives for firms to maintain quality, is perhaps the most fundamental of all the ideas that the economics literature contributes to branding. Its power and empirical relevance is illustrated in a field experiment run on the Chinese retail watermelon market by Bai (2017). She randomly assigns markets either to a control condition, a traditional sticker on each watermelon identifying sellers, which is, however, frequently counterfeited, or to a laser-cut label, which is more expensive to implement and, hence, less likely to be counterfeited. Over time, Bai finds that sellers assigned to the laser-cut la-
bel start selling higher quality melons (based on sweetness) and earned a 30-40% profit increment due to higher prices and higher sales.\textsuperscript{44} These findings are consistent with the predictions of the reputational models above.

In the domain of consumer goods, retailers have created brand images for their stores and chain through the assortment of manufacturer brands they carry:

“Retailers use manufacturer brands to generate consumer interest, patronage, and loyalty in a store. Manufacturer brands operate almost as “ingredient brands” that wield significant consumer pull, often more than the retailer brand does.” (Ailawadi and Keller, 2004, p. 2)

In some cases, retailers also use exclusive store brands or private labels to enhance the reputation of their stores and chain, by differentiating themselves through these exclusive offerings (Ailawadi and Keller, 2004). Dhar and Hoch (1997) report that a chain’s reputation correlates positively with the breadth and extent of its private label program. To shift their reputation from merely providing value, more recently, retailers have expanded their private label offerings into a full line of quality tiers, including premium private labels that compete head-on with national brands (see, e.g., Geyskens, Gielens, and Gijsbrechts, 2010). Recent work suggests that private labels have closed the quality gap and have become vehicles for reputation themselves. For instance, Steenkamp, van Heerde, and Geyskens (2010) report that as private label programs mature and close the perceived quality gap, consumers’ willingness to pay premia for national brands decreases.

The reputation literature underscores the role of time in establishing reputations. It is over time that a reputation develops, as the firm provides repeated evidence of fulfilling consumers’ expec-

\textsuperscript{44}One year after the experiment, once the laser branding was removed, the market reverted back to its original baseline outcome that was indistinguishable from a market with no labels at all.
tations. A new brand coming into a market may therefore face a “start-up problem”—how to get going on the reputation journey when consumers are reluctant to try it even once. Possible solutions to this range from “introductory low prices,” to offering money-back guarantees, to “renting the reputation of an intermediary” (Moorthy and Srinivasan, 1995).

Brand name reputation has value if and only if quality is not directly observable (Bronnenberg, Dubé, Gentzkow, and Shapiro, 2015). With the advent of the Internet, independent websites providing direct information about quality have proliferated. As consumers increasingly rely on such websites for product information, the value of brand name reputation is bound to go down. Waldfogel and Chen (2006) noted this as early as 2006. They observed that consumers using information intermediaries such as BizRate.com substantially increased their shopping at “unbranded” retailers such as “Brands For Less,” at the expense of branded retailers such as Amazon.com. More recently, Hollenbeck (2017) and Hollenbeck (2018) have examined the revenue premium enjoyed by chain hotels over independent hotels and observed that it has shrunk over the period 2000-2015, just as online review sites such as TripAdvisor have increased in popularity.

5.3 Branding as a signal

Much of what the industry refers to as “branding” activities would appear to the economist as “uninformative advertising,” i.e., advertising that is devoid of credible product quality information. In a series of seminal papers, Nelson (1970) argued that seemingly uninformative advertising for experience goods may nevertheless convey information if there exists a correlation between quality and advertising spending. Assuming consumers can perceive this correlation, it would be rational for them to respond to such advertising. Then the “money-burning” aspect of advertising will have
signaling value and a brand value can be established through the mere act of spending money on advertising associated with the brand.

A small literature has emerged that attempts to formalize Nelson’s ideas. Among these efforts are Kihlstrom and Riordan (1984), Milgrom and Roberts (1986), Hertzendorf (1993), and Horstmann and MacDonald (2003). In all of these papers, a key necessary condition for Nelson’s argument to work is the existence of a positive correlation between quality and advertising spending in equilibrium. This condition requires that the returns to advertising be greater for a high quality manufacturer than for a low quality manufacturer even after accounting for the latter’s potential incentive to copy the former’s advertising strategy (and thus fool consumers into thinking that its quality is higher than it actually is). In general, this condition is difficult to establish, as illustrated by both Kihlstrom and Riordan (1984) and Milgrom and Roberts (1986). The former works in a free-entry framework, with firms behaving as price-takers, and living for two periods. Firms decide whether to advertise or not at the beginning of the first period. In doing so they trade off the advertising benefit of being perceived by consumers as a high quality firm, which fetches higher prices, and the financial cost of advertising spending. As Kihlstrom and Riordan’s analysis demonstrates, it is possible to sustain an advertising equilibrium of the kind Nelson envisaged only under unrealistic cost assumptions or unrealistic information-transmission assumptions. For instance, if consumers learn true quality in the long run—the second period, in Kihlstrom and Riordan’s framework—then marginal costs cannot be lower for the lower-quality product (for if they were lower, then the lower-quality firm may also be tempted to advertise).

On the other hand, if marginal costs are assumed to be lower for the lower-quality product, then it must be assumed that high-quality manufacturers will never be discovered to be high quality (if they do not advertise). This condition rules out, for example, consumers spreading the word
about “bargains”—high quality products sold at low prices in the first period because they were mistakenly identified as low quality products due to their lack of advertising.

Milgrom and Roberts’s (1986) monopoly model shows that additional issues arise when prices are chosen by the firm. If advertising signals quality, then it is likely that the higher quality firm would also want to choose a higher price. But if prices also vary with quality, consumers can infer quality from the price rather than the advertising. It is unclear why, in a static model, a firm would need to burn money on advertising if it can signal quality through its prices. However, in a dynamic model with complete information about quality in the second period—akin to the first set of information-transmission assumptions in Kihlstrom and Riordan’s framework above—advertising may be needed to signal quality, but only if marginal costs increase in quality (in contrast with Kihlstrom and Riordan’s conditions). But even if marginal costs do increase in quality, the necessity of advertising to signal quality is not guaranteed. In Milgrom and Roberts’s own words: “advertising may signal quality, but price signaling will also typically occur, and the extent to which each is used depends in a rather complicated way, inter alia, on the difference in costs across qualities.”

Given the theoretical difficulties in establishing a signaling role for uninformative advertising, perhaps it is not surprising that empirical attempts to find a correlation between advertising and quality have turned out to be inconclusive. Several empirical studies have relied on the correlation between advertising spending and consumers’ perception of product quality using laboratory studies (e.g., Kirmani and Wright, 1989) and transaction histories (e.g., Erdem, Keane, and Sun, 2008). However, the correlation one seeks is between “objective quality”—the quality actu-

45Ackerberg (2001) is able to identify an informative role of advertising, separately from its consumption role—what he calls “the prestige effects of advertising”—by contrasting the purchase behavior of “new consumers” and “experienced consumers.” However, as he notes, he can’t identify how advertising is informative: “There are a number of different types of information advertising can provide: explicit information on product existence or ob-
ally built into the product, the sort of quality that might impact production costs—and advertising spending, and not “perceived quality” and advertising spending. As Moorthy and Hawkins (2005) have noted, a correlation between consumers’ perceptions of quality and advertising spending can occur through a variety of mechanisms, not necessarily Nelson’s mechanism.

Turning now to the studies examining objective quality-advertising spending correlations, Rotfeld and Rotzoll (1976) find a positive correlation between advertising and quality (as reported in *Consumer Reports* and *Consumers Bulletin*) across all brands in their study of 12 product categories, but not within the subset of advertised brands. In a more comprehensive study, using a sample frame of 196 product categories evaluated by *Consumer Reports*, Caves and Greene (1996b) find generally low correlation between advertising spending and objective quality. They conclude: “These results suggest that quality-signalling is not the function of most advertising of consumers goods.”

More recently, in a case study of the residential plumbing industry, McDevitt (2014) documents a novel use of branding as a signal of product quality. He finds that plumbers with names beginning with an A or a number, placing them at the top of alphabetical directories, “receive more than five times as many complaints with the Better Business Bureau, on average, and more than three times as many complaints per employee” McDevitt (2014, p. 910). This result is shown to be consistent with a signaling theory with heterogeneous consumer types in addition to firms with heterogeneous qualities. In equilibrium, low-quality firms use easy-to-find names that cater to low-frequency customers with low potential for repeat business and who will not find it beneficial to engage in costly search to locate the best firms. High-quality firms are less interested in such customers, servable characteristics, or signaling information on experience characteristics. It would be optimal to write down and estimate a consumer model including all these possible informative effects. Unfortunately, such a model would likely be computationally intractable, and more importantly, these separate informative effects would be hard, if not impossible, to empirically distinguish given my data set.”
focusing instead on customers with extensive needs who will devote more effort to searching for a good firm with which they can continue to engage in the future. These results corroborate ?’s (?) observation that cheap talk in advertising can serve the function of matching sellers to buyers.

With the rise of online marketplaces with well-established customer feedback mechanisms, it may be interesting to study whether the informational role of brands on consumer choices begins to erode in online markets. For instance, Li, Tadelis, and Zhou (2016) discuss how Taobao’s “Rebate-for-Feedback” feature creates a similar equilibrium quality signal as in Nelson (1970)’s money-burning theory of advertising.

Of course, signaling is not the only framework in which to interpret uninformative advertising. For instance, in the marketing literature, it is widely believed that such advertising is useful to create brand associations that help differentiate the brand in the consumer’s mind (see Keller, 2012, Chapter 2 for a survey). More recently, the economics literature has also recognized such a role for advertising via Becker and Murphy’s (1993) notion of “advertising as a good.”

5.4 Umbrella branding

5.4.1 Empirical evidence

Many new products are brand extensions that leverage the reputation and/or goodwill associated with an established brand, a practice often termed “umbrella branding” or “brand stretching.” Examples abound including Arm & Hammer, originally a baking soda, which has been extended to toothpaste, detergent, and cat litter; and Sony, a brand name created for a transistor radio in 1955, which has been extended to televisions, computers, cameras, and many other categories. Accord-

---

46Sellers have the option to pay consumers to leave feedback about the seller, where the payment is based on a Taobao algorithm that determines whether feedback is objectively informative.
ing to Aaker (1990), forty percent of the new products 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. The entire institution of business-format franchising relies on umbrella branding for its consumer-side effects.

In spite of the high incidence of umbrella branding, the empirical evidence for spillovers in consumer quality beliefs is limited. Erdem and 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 and Sun (2002) extend the model to allow for learning effects from marketing decisions like advertising and promotion. The parameter estimates imply that advertising and promotion not only reduce uncertainty about product quality, these effects can spillover across the two categories.

Any failures or negative associations with the extension could harm the original brand’s “reputation” Aaker (1990). 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 prod-

---

47 Consumer psychologists have found mixed evidence for 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). Alternatively, the match in the specific associations evoked by a brand may also affect the success of an extension. Broniarczyk and Alba (1994) give the example of different brands of toothpaste, some of which carry the association of being superior in tooth cleaning, others of freshness. Even though a tooth cleaning brand may be liked better in the original category (tooth paste), the brand associated with freshness can be evaluated more positively when extended into the category mouth-wash.
ucts. 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 first to enter new sub-markets. However, in a comprehensive analysis of 95 brands across 11 CPG categories, Sullivan (1992) finds that new brands typically enter earlier into new product markets 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.

In the next two sub-sections, we analyze theoretically the reputational effects of umbrella branding on a firm’s incentives to supply high quality, and on the signaling benefits of umbrella branding relative to the creation of entirely new brands.

### 5.4.2 Umbrella branding and reputation

We first discuss how a brand reputation incentivizes a firm to supply high quality to all the products under the common brand umbrella. The idea of holding brand reputation as hostage in return for quality assurance acquires even more power when the same brand is applied to several products. Now, when brand reputation suffers, many products suffer,\(^48\) not just one (Cabral, 2009; Rasmusen, 2016; Hakenes and Peitz, 2008). In this case, the incentive for a firm to maintain quality is even higher than in the previous section.

To illustrate, consider the following variation on the model presented in section 5.2 above. Suppose the firm controls not just one product, but two products, denoted 1 and 2. The notation above carries over wholesale, with superscripts denoting product-specific quantities. Suppose inequal-

\(^{48}\)The empirical evidence for such spillovers comes from Sullivan (1990).
ity (22) is satisfied for product 1, but not for product 2, i.e., \( v^1_h \geq \left(1/\rho\right)\left(c^1_h - c^1_l\right) + \max\{c^1_l, v^1_l\} \), but \( v^2_h < \left(1/\rho\right)\left(c^2_h - c^2_l\right) + \max\{c^2_l, v^2_l\} \). In short, if these products were separately branded, the monopolist firm would offer only product 1 in high quality, pricing it at \( v^1_h \). Product 2 would be offered at low quality only because it is impossible to find a price for this product that simultaneously is low enough to appeal to consumers and high enough to provide incentives for the firm not to renege on a high-quality reputation.

Umbrella-branding changes the quality incentives for the firm. Brand reputation is now a two-dimensional object. Accordingly, punishments, too, can be two-dimensional. As Cabral (2009) notes, several punishment regimens are available, ranging from lenient to draconian, depending on whether consumers punish one or both products for the indiscretions of one or both. The most draconian punishment consists of assigning low quality expectations forever to both products, even after only one low-quality deviation by one of the products. Such a draconian punishment regimen also means that any deviation will be “all-or-nothing”: if both products will be punished for a single deviation, the optimal deviation is to deviate on both forever.

The payoff in a potential equilibrium involving high quality on both products is

\[
\left(\frac{\rho}{1-\rho}\right) (v^1_h - c^1_l)N + \left(\frac{\rho}{1-\rho}\right) (p^2_h - c^2_l)N
\]

whereas the payoff in a deviation is

\[
\rho (v^1_h - c^1_l)N + \left(\frac{\rho^2}{1-\rho}\right) \max\{0, (v^1_l - c^1_l)\} N + \rho (p^2_h - c^2_l)N + \left(\frac{\rho^2}{1-\rho}\right) \max\{0, (v^2_l - c^2_l)\} N.
\]
A reputation-for-high-quality-on-both-products is sustained, therefore, if and only if

\[
p_h^2 \geq \left( \frac{1}{\rho} \right) (c_h^2 - c_l^2) + \max \{c_h^2, v_l^2\} - \left[ v_h^1 - \left( \frac{1}{\rho} \right) (c_h^1 - c_l^1) - \max \{c_l^1, v_l^1\} \right]
\]

If the term in square brackets is strictly positive, then the price-threshold for product 2 is reduced to below \((1/\rho)(c_h^2 - c_l^2) + \max \{c_h^2, v_l^2\}\), making it possible for \(v_h^2\) to exceed the threshold. Note that this construction works only with the most stringent punishment regimen. Anything more lenient, such as punishing only when both products have low quality, or punishing only the product which has low quality, would make it impossible to sustain a reputation for high quality on both products.

Cabral (2009) and Hakenes and Peitz (2008) explore how imperfect monitoring affects umbrella branding incentives. The idea consists of distinguishing between the firm’s inputs and how the product performs. Only the former is controllable by the firm. Consumers observe product performance, but cannot observe the firm’s inputs—its “intent.” In this context, poor product performance could be the accidental bad outcome of “high quality intent” or the intentional bad outcome of “low quality intent.” Now it is possible for punishments to be “too strong.” Instead of encouraging high quality, punishments could have the opposite effect: the firm, recognizing that it is unable to prevent punishments even with the best intentions, sees no point in trying to sustain a high-quality reputation. The fault, however, is not with umbrella branding, but rather with the punishment regimen. It is still better to be umbrella-branded rather than not, because umbrella branding allows more flexible punishments.

Cabral (2009), Hakenes and Peitz (2008), and Rasmusen (2016) only consider cases where...
consumers hold the brand accountable for product failures. But, as Choi’s (1998b) work suggests, the brand may also be held accountable for bad extensions. In other words, product failure may hurt brand reputation not only because the brand cannot be trusted to deliver high quality, but also because the brand cannot be trusted to be extended to other high quality products. Under those circumstances, subsequent extensions may not be purchased.

While this literature is helpful in suggesting that umbrella branding strengthens the incentive to maintain reputations, it does not provide any guidance on what brand extensions a firm ought to pursue. A large literature in consumer psychology finds that fit between the parent brand and the extension category is important to the success of a brand extension (e.g., Aaker and Keller, 1990; Broniarczyk and Alba, 1994). This literature finds that perceived similarity of the category or usage situation, and the relevance of brand parent associations, makes a reputable brand (e.g., a toothpaste brand) more effective at extending its reputation into a closely related category (e.g., the mouthwash category versus a more distant category like shaving cream).

5.4.3 Umbrella branding and product quality signaling

In the previous section, we analyzed the reputational effects of umbrella branding on product quality provision. We now analyze the case where a firm with a product line can use umbrella branding to signal quality. In contrast with the previous reputational role of brands, where brands merely carried a reputation, now the firm decides whether or not to umbrella brand. Of interest is whether a firm with high quality products can use umbrella branding to signal that a new product is also “high quality.”

In a pioneering paper, Wernerfelt (1988) conceptualizes the problem as follows. Suppose a monopolist is endowed with two products, “old” and “new,” each of which, independently, can
be “high” (h) or “low” quality (ℓ). Product performance is assumed to provide only an imperfect indication of quality: high quality products always work and low quality products work (w) or fail (f), with probabilities θ and 1 − θ, respectively. Consequently, consumers do not observe product quality before or, potentially, after purchasing the product. When the new product is introduced in period 0, consumers observe the old product’s performance (w or f); but not its quality. The firm does not observe the old product’s performance and decides whether or not to umbrella brand (B) or to create a new brand (N). Umbrella branding costs β more than new branding. After the new product is introduced and purchased, its performance is observed (period 1), and consumers must then decide whether to buy the old product again. If consumers have beliefs that product quality is high with probability P, then their purchase decisions generate a revenue stream of x(P), where x(·) is an increasing function.

Suppose consumers have an initial prior that each of the two products is high quality with probability η ∈ (0, 1). Thus, the prior probability of (h, h) is η², the prior probability of (h, ℓ) is η(1 − η), and so on. Of interest is whether umbrella branding can generate quality beliefs of (h, h) with probability 1 in equilibrium. Wernerfelt (1988) shows that such beliefs can be sustained if the relative cost of umbrella branding, β, is not too large. If β is small enough, there exists a separating equilibrium in which only the firm with two high-quality products umbrella brands, while all other firm types choose to use a new brand. The following equations characterize the separating equilibrium (here the notation π(B; n, o) (resp. π(N; n, o)), n, o = h, ℓ, refers to the firm’s profits when it umbrella brands (introduce a new brand) when its new product is of quality n and its old product is of quality o):

\[ π(B; h, h) = x(1) + [x(1)] − β ≥ x(P^w|\{w, N\}) + x(P^w|\{w, N, w\}) = π(N; h, h) \] (23)
\[
\pi(B; h, \ell) = x(1) + [\theta x(1) + (1 - \theta) x(\varphi)] - \beta \leq 
\theta x(P_n|\{w, N\}) + \theta x(P_o|\{w, N, w\}) + \theta x(P_o|\{w, N, f\}) = \pi(N; h, \ell) \tag{24}
\]

\[
\pi(B; \ell, h) = \theta x(1) + [\theta x(1) + (1 - \theta) x(\varphi')] + [1 - \theta] x(P_o|\{w, N, w\}) - \beta \leq 
\theta x(P_n|\{w, N\}) + \theta x(P_n|\{w, N, f\}) + \theta x(P_o|\{w, N, f\}) = \pi(N; \ell, h) \tag{25}
\]

\[
\pi(B; \ell, \ell) = \theta x(1) + (1 - \theta) x(\psi) + \theta x(1) + (1 - \theta) x(\varphi'') - \beta \leq 
\theta x(P_n|\{w, N\}) + \theta x(P_n|\{w, N, f\}) + \theta x(P_o|\{w, N, f\}) = \pi(N; \ell, \ell) \tag{26}
\]

These equations ensure that umbrella branding will only signal that both new and old products are high quality if firms with \((h, h)\) prefer to umbrella brand (23), but other firm types do not (24)-(26). It is easy to see that if umbrella branding does not cost too much, (23) will be satisfied by virtue of “signaling expectations.” The challenge consists of verifying that similar expectations would not tempt other firm types to choose umbrella branding. The key for separation lies in the off-equilibrium beliefs: consumers’ beliefs when they observe one or both of the umbrella branded products failing, which should not arise under \((h, h)\). In the equations above, these are: (1) if the old product has failed in period 0, then consumers believe that the new product is good with probability \(\psi\) and the old product is good with probabilities \(\varphi'\) and \(\varphi''\), depending on whether the new product works or fails, respectively, and (2) if the old product works in period 0 but the new product fails in period 1, then consumers believe that the old product is good with probability \(\varphi\). In a perfect Bayesian equilibrium, off-equilibrium beliefs are a free parameter. Wernerfelt (1988) sets these beliefs to be very “negative”: \(\varphi = \varphi' = \varphi'' = 0\). The assumption \(\varphi = 0\) is particularly draconian: even if the old product has worked in period 0, consumers are asked to view the old product as low quality with probability 1 based on the new product’s failure under umbrella branding. This is analogous to the “collective punishment” regimen we saw in section 5.4.2 above in the reputations literature.

Several subsequent writers have objected to Wernerfelt (1988)’s assumption that umbrella
branding costs more than creating a new brand. After all, one of the leading arguments for umbrella branding is that one doesn’t have to undertake the expense of creating a new brand. In a variant of Wernerfelt (1988)’s model, with cost-neutrality instead of umbrella branding costing more, Cabral (2000) shows that if old and new products are of the same quality and quality is continuous, then brand extension can still have some signaling value. However, since the signal is binary, and quality lies on a continuum, it can only separate qualities into regions: a “higher quality band” chooses to umbrella brand while a “lower quality band” chooses not to. More recently, Miklós-Thal (2012), also assuming cost-neutrality, but with ex ante independence between the two products, finds that Wernerfelt (1988)’s equilibrium cannot exist in her model.

In an extensive critique of Wernerfelt (1988)’s model, Moorthy (2012) argues that signaling quality is a weak basis for justifying brand extensions. Besides the problematic assumption that brand extension costs more than creating a new brand, he shows that the signaling equilibrium relies on off-equilibrium beliefs that are poorly motivated in the model. For instance, Wernerfelt (1988) assumes that consumers must penalize the old product if the new product does not perform well, even if they have observed the old product perform well and they believe that the products’ qualities are uncorrelated (indeed, even negatively correlated). No empirical evidence supports such beliefs. These collateral-damage considerations are essential to the argument, but, in most real-world brand extension situations, the old product’s quality is rarely in doubt. In fact,
brand extensions are only undertaken after old products have solidified the reputation of the brand performing well (Sullivan, 1992).

This suggests that Choi (1998b)’s model provides a better basis for finding a signaling role for brand extensions. First, in his model brand extension works in conjunction with price to signal new product quality. The old product’s quality plays no role. Second, brand extension is not necessary to signal the new product’s quality. Price could have done so, too, but, by itself, it would have to bear a heavier signaling burden, with an attendant loss of profit. Brand extension reduces the signaling burden, similar to advertising (e.g., Milgrom and Roberts, 1986); but unlike advertising, it may not involve out-of-pocket costs. Third, brand extension’s signaling role is sustained by consumers’ implicit threat that if a firm “cheated”—extended a brand to a low quality product—then consumers would no longer take any signal from the brand. Future brand extensions would be treated like new brands. Price would then have to bear the entire signaling burden, reducing profits. Choi (1998b) thus provides a reputational basis for why consumers might want to assume that a new product bearing an existing brand name provides the same good quality that old products under the same brand have been known to provide.

5.5 Brand loyalty and equilibrium pricing

Section 3.1 discussed the empirical role of psychological switching costs as a source of brand loyalty. Brand loyalty from switching costs introduce two countervailing incentives for firms’ pricing decisions. The persistence in demand motivates firms (1) to prospect for new customers (the “investment” motive), and (2) to exploit the loyalty of existing customers (the “harvesting” motive). Therefore, the net effect of switching costs on the equilibrium prices in a market is not

---

51For example, only successful movies are sequeled, not movies that fail.
clear a priori.

Klemperer (1995) and Farrell and Klemperer (2007) conclude that there is a “strong presumption” that switching costs soften price competition, leading to higher equilibrium prices. Therefore, the conventional wisdom asserts that brand loyalty leads to higher prices. The conventional wisdom is based on an early theoretical literature that studied stylized analytical models of dynamic oligopoly in which the harvesting motive outweighs the investment motive. In a two-period model, Klemperer (1987) derives the well-known “bargain then rip-off” result whereby firms compete for market share in the first period and then raise their prices in the second terminal period. In an infinite-horizon model of homogeneous goods and overlapping consumer generations, switching costs allow firms to raise their prices above marginal costs and to derive supra-normal economic profits (Farrell and Shapiro, 1988; Padilla, 1995; Anderson, Kumar, and Rajiv, 2004). Similarly, in an infinite-horizon model with differentiated products and perfect lock-in (i.e. infinite switching costs), equilibrium prices are also higher than in the absence of switching costs (Beggs and Klemperer, 1992).

Most of the key assumptions in this literature are unlikely to hold in consumer markets where products are differentiated, price competition is not subject to a “terminal period” and consumer loyalty is imperfect in the sense that consumers switch brands over time even in the absence of any price adjustments. Using empirically estimated demand for branded consumer goods, Dubé, Hitsch, and Rossi (2009) compute the corresponding steady-state equilibrium prices under different magnitudes of switching costs. At the empirically-estimated magnitudes of switching costs, they find that equilibrium prices would be lower in the presence of switching costs. Moreover, they find that switching costs would need to be at least four times their empirically-estimated magnitudes before they would lead to higher equilibrium prices.
The extant literature has not analyzed the impact of brand loyalty on other marketing decisions. Early work by Brown (1952, p. 257) observed that

“One brand loyalty is high, the advertiser has a good case for "investment" expenditures where large amounts are expended over short periods of time to win new users in the knowledge that continued purchases after the advertising has been curtailed will "amortize" the advertising investment.”

Shum (2004) quantifies the extent of advertising expenditures required to overcome loyalty. But a lingering direction for future research might be to analyze how, in equilibrium, price and advertising competition is moderated by consumers who form brand loyalties or shopping habits.

5.6 Brand loyalty and early-mover advantage

In Section 3 we discussed the evidence for, and the motivations behind, persistence in consumers’ brand choices and the emergence of brand loyalty over medium-term horizons of several years. A separate literature has documented the persistence of market-share leadership by brands over much longer horizons spanning decades. 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.52 All 25 brands were still ranked among the top 5 in the category in 1983.53 Persistence in dominance has also been documented for the initial entrants into a new product category, the “pioneering advantage” (Robinson and Fornell 1985; Urban, Carter, Gaskin, and Mucha 1986; Lambkin 1988; Robinson 1988; Parry and Bass 1990; Kerin, Varadarajan, and

---

52 Advertising Age (1983), "Study: Majority of 25 Leaders in 1923 Still On Top" (September 19), 32.
53 Golder (2000) extended this analysis to 100 categories and using more reliable 1997 market share data. He finds that only 23% of the dominant firms in 1923 remain dominant in 1997; although nearly 50% remain in the top 5.
Peterson 1992; Brown and Lattin 1994). A similar persistence of dominance dating back to 1933 is reported for consumer brands in the UK (Keller 2012, p. 21). Kalyanaram, Robinson, and Urban (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. While the exact definition of a “pioneer” is under debate (Golder and Tellis, 1993), the common finding across these studies is the evidence of persistence in the market shares for early movers that “survive” long-term. Since this work typically relies on a single time-series data set for any given industry, the results are subject to the usual identification concerns regarding the role of state-dependence (early-mover status) versus heterogeneity in managerial skills.\footnote{One exception is (Brown and 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.}

More recently, Bronnenberg, Dhar, and Dubé (2009) use a panel approach to test for persistent, early-mover advantages in market shares versus heterogeneity with CPG data spanning multiple US cities. For each city, they obtain the exact date of entry for the leading brands in the category. Since most of the brands launched during the late 1800s, long before marketing and distributional technology existed to coordinate a national launch, the key identifying assumption of exogenous entry timing across cities may not be problematic. In six case studies, they find that the historic order-of-entry (often a century earlier) among survivor brands in a geographic market predicts the current rank-order of market shares in that market. These findings are visualized in Figure 2 which plots the geographic distribution of brand shares for the Ground Coffee category across US cities. The diameter of each circle is proportional to a brand’s market share in that city, and shading indicates the earlier entrant. Historic order-of-entry in a geographic market also predicts the current rank order of brands’ perceived quality levels as measured by Young & Rubicam’s 2004 Brand...
Asset Valuator survey. For 49 of the top two national brands in 34 CPG categories, Bronnenberg, Dhar, and Dubé (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 19th and early 20th centuries with entry in more distant markets occurring relatively later.\textsuperscript{55} Collectively, the persistent, early-mover effects for brands suggests an important role for branding in the shaping of the market structure of consumers goods industries.

– insert Figure 2 here –

A number of mechanisms may yield first-mover advantages, and not all of them are brand loyalty-related. For example, Robinson and Fornell (1985) and Robinson (1988) note that first movers can benefit from a lower-cost position, which they achieve by riding further along the learning curve, or simply through scale and scope economies. On the demand-side, first mover advantages can arise because the first mover, having first use of the product space, is able to choose the most desirable product position (Lane, 1980; Moorthy, 1988; Sutton, 1991), or several such positions, effectively pre-empting the product space through product proliferation Schmalensee (1978, 1982); Judd (1985). The first mover can also gain an advantage by using its additional time on the market to “perfect” its product—more generally, its entire marketing mix (Robinson and Fornell, 1985; Robinson, 1988)—or create more awareness and/or marketing-based goodwill (Doraszelski and Markovich, 2007), or simply because consumers, having experienced the first

\textsuperscript{55}See for instance (Bartels, 1976; Tedlow, 1990) for detailed discussions of how entrepreneurs in the late 19th century with new consumer brands gradually rolled them out across the US.
product in the market, are reluctant to try a new product whose experience attributes they are uncertain about (Schmalensee, 1982). In industrial goods, the cost-related mechanism may prove decisive (Robinson, 1988), whereas among consumer goods, the demand-related mechanisms may prove decisive.

6 Conclusions

Brands and branding are central to the understanding of the market structure of consumer goods industries. On the demand side, we have discussed three potential effects of branding. First, brands may enhance the consumption utility of branded goods relative to unbranded substitutes. Second, brands may reduce search costs and stimulate consumer consideration. Finally, brands may signal product quality and affect consumer demand through reputation. These different influences of branding on consumers generate several striking patterns in purchase behavior including brand loyalty and inertia, a form of switching costs, and longer-term persistence in brand choice as consumers learn and change their beliefs about quality of branded goods.

An interesting direction for future research consists of testing the exact mechanisms through which brands ease the consumer search and consideration process. The literature has yet to parse the extent to which estimated brand value to consumers reflects genuine preferences as opposed to the facilitation of search. A related direction for future research on the demand side consists of modeling how consideration sets are formed over time and distinguishing between how a consumer becomes aware of brands and how this awareness influences consideration during a given purchase occasion. Of particular interest is the extent to which long-term industrial market structure is shaped by firms’ marketing investments to build consumer awareness and consideration for their
brands. Future research may benefit from digging more into the psychological roots of consumer memory and the persistent effects memory creates for brand preferences. Moreover, memory may be an important moderator through which branding-related expenditures on the supply side become valuable brand assets in the long-term.

On the supply side, there still is no consensus regarding the role of brand assets to firms and the mechanism through which these assets are built. Attempts to test the traditional signaling, or “money-burning,” theory of advertising have mostly been inconclusive. Some progress has been made on testing reputation theories. It is possible that brands can emerge under different circumstances, supporting a co-existence of each of these theories. Cross-industry studies will likely provide a fruitful direction for future research to determine which institutional factors support a specific branding theory within a market.

The value of brands to firms manifests itself mostly through price-premia and brand-equity. Although there is a strong consensus among scholars and policy makers that brand names are valuable, intangible assets, there is no agreement on how to measure the value of brands. The General Accepted Accounting Principles (GAAP) offer no explicit guidance to firms about how to value brand capital that is internally created. Current business practice disregards internally created brand value as an asset. At the same time, and seemingly inconsistent, the activation of externally acquired brands on the balance sheet is commonly accepted. This practice likely misrepresents many marketing investments as expenses.

A recent literature has proposed a more rigorous structural approach to measure brand value that defines the incremental profits of brands through counterfactuals. Nevertheless, more research is needed to tie the value of brands back to these demand components. Another direction for future research consists of analyzing how firms invest in building brands and how such “branding”
will likely evolve. The reduced barriers to entry in the digital era have led to a rapid influx of new consumer products with a more customer-centric image and focus (Islam, 2018), potentially enhancing the role of a brand as a reputation signal and changing the manner in which firms build the brand’s image. Similarly, the digital era has expanded the set of channels through which consumers can buy brands and through which firms can actively invest in branding. While most of the extant branding management literature has focused on advertising investment, another interesting direction may consist of the role of distribution and availability on brand performance.  

Most important, the literature has mostly been silent on how branding and brand investments moderate product market competition. Another important direction for future research will be the analysis of whether branding softens or toughens price competition. In addition, more work is required to understand how competitive branding underlies the industrial organization of consumer goods markets and the extent to which it leads to persistent, concentrated market structures as suggested by recent empirical research.

\footnote{In a large-scale study of over 200 new product launches, Ataman, Mela, and Heerde (2008) find reduced-form evidence that distribution may be more highly associated with a new brand’s success than other marketing variables, such as promotions and advertising.}
References


BARTELS, R. (1976): The History of Marketing Thought. Grid Publishers, Columbus, OH.


Figure 1: Static and dynamic estimates of brand value in the stacked chips industry. Source: Borkovsky, Goldfarb, Haviv, and Moorthy (2016).
Figure 2: The Geographic Distribution of Ground Coffee Brand Market Shares in the US (source: Bronnenberg, Dhar, and Dubé 2009)