Does Peter Piper Pick a Package of Pepper Inattentively? The Consumer Response to Product Size Changes

By Ian Meeker*

September 14, 2021

In the consumer-packaged goods industry, firms can increase unit prices by decreasing package content, a practice known as product downsizing. Since consumers tend to underuse information on product size, they may fail to notice size changes. Downsizing in the black pepper industry provides an opportunity to test whether consumers are inattentive to changes in package content. I build a structural model of consumer preferences that incorporates inattention to size changes and apply it to grocery store scanner data. I find that consumers are insensitive to size decreases, despite their preference for larger package sizes. This differential sensitivity to size suggests that downsizing exploits consumer inattention. With full information, consumers would switch to larger packages that provide greater welfare.

Food manufacturers sometimes replace products with smaller versions, a practice known as product downsizing.[1] Some manufacturers shrink their packaging to reflect the reduced content, but many do not. Examples of downsizing abound. In 2020, Dawn reduced the amount of dishwashing liquid in a bottle by one ounce; Great Value reduced the amount of paper towels in a roll by 148 sheets; and Keebler reduced the amount of Club Crackers in a box by 1.2 ounces. As these examples show, product downsizing occurs across a wide range of products. In some industries, downsized products constitute a large fraction of the available products.

Manufacturers often use downsizing as a way to increase unit prices (i.e. price per ounce), keeping package prices constant while reducing package content. Most firms do not advertise such size changes. To identify dowsizing, consumers must correctly process the available sizes. Because many consumers use visual estimates in place of explicit size information, they may fail to notice the smaller size as firms downsize their products in a number of different, and often subtle, ways. If consumers are inattentive, downsizing represents a hidden price increase.

I test whether consumers are inattentive to product size changes in the black pepper industry.

* Meeker: Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215 (email: imeeker@bu.edu).
This paper relies on data from Nielsen Consumer LLC. This paper presents the researcher's own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[1] The literature has used various terms to describe this practice, including **package downsizing** (Cakir and Balagtas, 2014; Yonezawa and Richards, 2016), **downsizing price increase** (Gourville and Koehler, 2004), **content reduction** (Kachersky, 2011), and **shrinkflation** (Ochirova, 2017). Some papers (e.g. Gupta et al., 2007) add the additional requirement that the package price remains the same. As Imai and Watanabe (2014) show, large decreases in package content may lead to decreases in the package price even if the unit price increases.

To recover the degree of inattention, I develop a model of inattention and show how it can be recast as a standard random coefficient model. In the model, inattention results in consumers evaluating product utility according to the product's original size, causing the change in product size to enter utility as an additional product characteristic with a random coefficient. The distribution of this random coefficient characterizes the degree of inattention. Thus, estimating the extent of inattention amounts to estimating the distribution of the random coefficient.

To avoid imposing parametric functional form assumptions that might restrict the types of inattention present, I employ a semi-nonparametric method that estimates the model in two stages. In the first stage, I estimate the linear parameters that are unaffected by inattention using a standard logit model. In the second stage, I estimate the distribution of the inattention parameter nonparametrically using the fixed grid approach of Fox et al. (2011) and Heiss, Hetzenecker and Osterhaus (2021). The fixed grid approach approximates the true distribution using a fixed set of values. The particular fixed grid estimator that I use is a special case of elastic net. This procedure allows me to estimate the distribution of the inattention parameter without imposing strong functional form assumptions that might restrict the types of inattention.

I use my method to examine a downsizing event in the black pepper industry where McCormick, the industry's largest firm, shrank the content of five products, which constitute 25% of the market. This downsizing event provides an ideal opportunity to study inattention due to the wide range of product sizes offered by other firms in this industry. Consumer substitution between different sizes allow me to estimate consumers' size preferences. I then recover inattention by comparing how product shares change after downsizing to how product shares should change given consumers' preferences before downsizing. When consumers are inattentive, product shares will remain constant since consumers do not notice the change, whereas when consumers are attentive, the product shares decline according to consumers' size preferences. The difference between the observed trend and the expected trend after downsizing identifies inattention. Previous studies on downsizing (Cakir and Balagtas, 2014; Yonezawa and Richards, 2016) do not recover inattention and may not be able to since they consider industries with little size variation. Existing variation is necessary to construct the expected trend.

Applying my model on store-level data from Nielsen, I find that more than 97% of consumers fail to notice at least one size change. Moreover, the probability of being inattentive is similar across products. The probability of inattention does not change with the magnitude of the size change or the type of packaging.

By distorting product utilities, inattention causes some consumers to choose products that are not utility maximizing, thereby reducing consumer welfare. I compute the welfare loss from limited attention under the existing pricing structure and find that inattention lowers the average consumer's welfare by $0.10 or 3% of the average product's price. This suggests that policies that increase consumer attention to product size, such as larger labels for net weight or a registry of downsizing products, could improve welfare.

I. Previous Research

Previous studies of downsizing in the ice cream industry (Cakir and Balagtas, 2014) and the cereal industry (Yonezawa and Richards, 2016) find that consumers are less sensitive to size than price. Neither study explores why consumers appear to undervalue size. There are several possible reasons. One possibility is that consumers care about other product features more than size. For example, when buying ice cream, consumers may place more emphasis on quality and taste than on whether a container has 48 or 56 ounces. Another possibility is that consumers are attached to the downsizing brand. Brands with strong customer loyalty may be confident that downsizing will not affect consumer attachment.

I explore a third explanation, namely, that the lack of response to size changes is due to inattention. Consumers frequently ignore explicit size information and instead rely on visual cues to evaluate size (Lennard et al., 2001). Visual estimates can be inaccurate since they are subject to cognitive biases. For instance, consumers perceive tall, narrow objects to be larger than short, wide objects of the same volume (Krishna, 2006). Such perception biases grow when the size of the object changes across multiple dimensions (Chandon and Ordabayeva, 2009). Particular size changes can result in consumers failing to notice even a 24% decrease in package size (Ordabayeva and Chandon, 2013). Consumers' poor grasp of volumes translates to unit prices as well. Since many consumers do not compare unit prices across sizes of the same product and often pay a surcharge for larger quantities (Clerides and Courty, 2017; Joo, 2018), they will probably not compare unit prices across brands. This suggests that downsizing can be an effective strategy to hide an increase in the unit price. Determining the level of inattention is important since it dictates the degree to which firms can engage in downsizing.

Even if consumers are inattentive to size changes, exploiting inattention comes with risks. Consumers may feel deceived and react negatively toward the downsizing brand upon discovery of the size decrease. In lab experiments, consumers presented with downsized products expressed a lower willingness to buy the presented brand (Kachersky, 2011; Wilkins, Beckenuyte and Butt, 2016). The possibility of a backlash may explain why many firms do not advertise their downsizing decisions.

Consumers exhibit inattention and cognitive biases in a variety of settings. Many do not pay close attention to hidden attributes like shipping costs (Brown, Hossain and Morgan, 2010) or sales taxes (Chetty, Looney and Kroft, 2009). Consumers can also misperceive product attributes. Allcott (2013), for example, finds that consumers misjudge the value of fuel economy when choosing cars. In some cases, consumers give particular attributes too much weight. For instance, many consumers place too much emphasis on the left-most digit and pay higher prices for cars whose mileage falls below 10,000 miles (Lacetera, Pope and Sydnor, 2012). When consumers are behavioral, changing the available information affects consumers' decisions and therefore their welfare. Grubb (2014) shows that bill-shock regulation that forces firms to notify consumers about overages can improve welfare. Moreover, the way in which information is presented can also matter. Luca and Smith

(2013) show that the way in which *U.S. News and World Report* lists its college rankings affects the number of applications. If cognitive biases can influence major life decisions, they can also impact consumers' minor purchasing decisions.

A number of studies provide methods to identify and to recover inattention to product attributes. Abaluck and Compiani (2020) provide a method to test for inattention using the cross derivatives of the choice probabilities. Their method is not applicable in my context because it assumes that consumers ignore the hidden attribute when searching. Brown and Jeon (2020) provide a method for recovering consumers' information processing strategies grounded in a rational inattention framework. For their method to be tractable, they place restrictions on the prior distribution of product utilities and hence the information processing strategy. In constrast, my model recovers inattention without functional form assumptions, but unlike Brown and Jeon (2020), my model does not explain how consumers process information and become inattentive.

II. Data

To analyze downsizing, I use the Nielsen Retail Scanner data and the Nielsen Consumer Panel data from the Kilts Center at the University of Chicago. The Retail Scanner data provides point-of-sale data for around 35,000 stores in the United States. The Scanner data covers over 4 million consumer package goods. The Consumer Panel provides a micro-level panel of consumer purchases. It tracks between 40,000 and 60,000 households. I estimate my structural model using the aggregate, store-level data for the period from 2014 to 2016. I focus on this period to avoid complications that might arise if size preferences shift over time. While my analysis is at the store-level, I also use the individual-level purchase data from 2012 to 2016 to inform the modelling.

I focus on McCormick's downsizing efforts since I cannot identify the private-labels that downsized during this period. This is because Nielsen masks the identity of private-label brands to preserve the anonymity of its data partners.

III. Industry Background

Pepper is a staple seasoning with the majority of households purchasing it at least once in the five-year span from 2012 to 2016.[2] While most consumers will purchase pepper at some point, they do so infrequently. In any given year, only around 30% of consumers buy pepper. Many go several years before purchasing pepper again. The long interpurchase times are due to pepper's high storability. As black pepper is a dried fruit, it does not spoil, but instead loses its pungancy over time (Feucht, 2019). When stored properly, whole peppercorns will retain their flavor for three to four years and ground peppercorns will retain their flavor for one to two years (Feucht, 2019).

Because all black pepper comes from the same flowering vine, pepper products are very similar in most respects. They differ slightly in terms of quality and taste which stem from differences

[2] Author's calculation based on a balanced panel from the Nielsen Consumer Panel data.

in soil, climate, and processing method. The largest differences are in branding and packaging. From 2014 to 2016, there were 133 different brands available at stores in the my data. Given the similarity between products, many consumers opt for cheaper store brands. Store brands capture around 40% of the market during this period. In constrast, the typical name brand is a small and regional with a market share that is less than 0.1%. Among name brands, McCormick stands out with its 40% market share; no other name brand exceeds 7%. The brand's owner McCormick & Co. dominates the industry, owning three of the top five selling name brands in McCormick, 5th Season, and Spice Classics. Through its various brands and private labels, the company controls around 70% of the market. The next largest firm B&G Foods, the producer of the brands Tone's and Durkee, accounts for approximately 9% of the market.

In addition to the large number of brands, the industry features a wide array of sizes. In the consumer panel data, products range from 0.4-ounce bags to 32-ounce containers with many sizes in between (Figure A5). Examining the histogram of the sizes purchased from 2014 to 2016, the most-frequently purchased sizes were two and four ounces, which correspond to the standard sizes of small and medium tins, respectively. Most stores in the scanner data offer these two sizes along with many others. The typical store offers 15 different sizes of black pepper at any given time (Figure A6). Some stores offer as many as 30 distinct sizes and others as few as a single size. Although most stores offer more than ten sizes, a noticable percentage of stores offer a limited variety, having fewer than four distinct sizes at a given point in time. Differences in the available sizes across stores force consumers to substitute to similarly sized products and directly reveals consumer substitution patterns, which in turn allows me to separate size preferences from inattention.

IV. Downsizing in the Black Pepper Industry

Downsizing in the black pepper market came in response to rising commodity costs. From 2009 to 2014, wholesale pepper prices were increasing due to growing demand in emerging markets (Figure A1). With prices trending upward, a poor harvest in 2014 caused wholesale prices to spike (Figure A1). Over the course of 2014, the wholesale price of black pepper increased by over 30%. Brands responded to this sudden cost increase in different ways. Most brands chose to increase their product prices. Others, like McCormick & Co, changed their product sizes (Table 1). McCormick downsized a wide range of products. It decreased the weight of the pepper in its tins by 25% and the weight of the pepper in its grinders by 19% in February 2015 (Figure A2). A federal court noted that McCormick also asked the private-labeled brands that it manufacturers to reduce their fill levels (*In Re: McCormick & Co.*, 2019). Most agreed to the new smaller sizes (*In Re: McCormick & Co.*, 2019). Initially, McCormick downsized its products by reducing fill levels while keeping packaging the same size. In the middle of 2016, the company adjusted its package sizes to reflect the reduced weight (*In Re: McCormick & Co.*, 2019). This change is not observable in the data since it did not affect the product codes.

McCormick states that it engaged in downsizing to preserve product quality while avoiding large

Table 1—: Downsized Products

	Product	Original Size (Oz)	New Size (Oz)	Size Decrease (Oz)	
1.	McCormick Large Tin	8	6	2	25%
2.	McCormick Medium Tin	4	3	1	25%
3.	McCormick Medium Grinder	3.1	2.5	0.6	19%
4.	McCormick Small Tin	2	1.5	0.5	25%
5.	McCormick Small Grinder	1.24	1	0.24	19%

Note: The list includes only name brand products. The products are ordered from largest to smallest size change.

price increases (Hughlett, 2016). Several consumer lawsuits allege that McCormick sought to exploit consumer inattention using downsizing. The plaintiffs cite an internal email in which a McCormick employee writes, "keep in mind consumers do not really know the fill level right now" (*In Re: McCormick & Co.*, 2019). No matter what McCormick's intentions were, downsizing can still take advantage of inattention.

V. Reduced-Form Analysis

A. Inattention

I start by considering how downsizing impacts market shares. After an initial decline as stores transition from the original products to downsized one, the market share of the downsized products in Nielsen stores actually increases (Figure A4a). This increase occurs despite a large increase in unit prices for the downsized products. The average price per ounce of the downsized products increases by approximately $0.50, whereas the average price per ounce of the nondownsized products increases by approximately $0.25 (Figure A4b). This relatively large increase in unit prices for the downsized products should have caused consumers to substitute away from McCormick to other brands. The fact that this does not occur and that the market share of the downsized products increases suggests that consumers either prefer smaller product sizes or are inattentive to unit prices.

To explore whether consumers' insensitivity to unit price increases stems from inattention, I regress the log of quarterly units sold on price, size, indicators for the downsized products, and various controls. Table 2 reports the results. All of the specifications include year-quarter fixed effects. Specification (1) also includes product fixed effects; whereas, specifications (2) and (3) have brand fixed effects. Since products change size only once or not at all, the size coefficient is not identifiable after the inclusion of product fixed effects and downsizing dummies and is therefore missing from (1).

If consumers are attentive, any changes in market shares or product sales should come from the change in size, all else equal. Thus, after controlling for size and other factors, the downsizing indicators should be indistinguishable from zero. These indicators are actually positive and significant in every specification (Table 2), which suggests that consumers prefer the downsized products' new smaller sizes. However, the positive coefficient on product size suggests the opposite, namely that consumers prefer larger sizes (See (2) and (3) in Table 2). Inattention to size changes would account for this apparent contradiction between consumers' size preferences and their purchasing behavior. Inattentive consumers do not notice the decrease in package size and instead focus on package prices. To inattentive consumers, the downsized products are attractive since their package prices remain the same while nondownsized products see a corresponding increase.

Table 2—: Reduced-Form Evidence of Inattention

	(1)	(2)	(3)
Price	−0.074	−0.118	−0.124
	(0.001)	(0.034)	(0.034)
Size		0.080	0.236
		(0.021)	(0.062)
Size Squared			−0.009
			(0.003)
Downsized 1	0.562	0.692	0.739
	(0.015)	(0.086)	(0.084)
Downsized 2	0.623	0.655	0.738
	(0.006)	(0.092)	(0.085)
Downsized 3	0.647	0.653	0.709
	(0.010)	(0.097)	(0.092)
Downsized 4	0.642	0.631	0.686
	(0.007)	(0.141)	(0.133)
Downsized 5	1.302	1.267	1.292
	(0.006)	(0.097)	(0.094)
Product FE	Y		
Brand FE		Y	Y
Year-Quarter FE	Y	Y	Y

Note: Based on scanner data from 2014 to 2016. The dependent variable is the log of units sold by quarter. The variables *Downsized 1-5* are indicators for the five downsized products. The columns report clustered standard errors in parentheses. I cluster the standard errors by product and year-quarter.

Even if consumers exhibit inattention, this inattention may prove transitory. As consumers use a downsized product, they may notice the product's smaller size and adjust their future purchasing behavior. To explore if learning matters, I examine how downsizing affects product sales over over time. In the previous regressions, I replace the indicator for the downsized products by a single indicator for downsizing, but allow its coefficient to vary by the quarter since the downsizing event. If learning takes place, the magnitude of the downsizing coefficient should decline over time. Initially, the coefficient on the downsizing term is positive with a magnitude of 0.789. Even six quarters after downsizing, the coefficient has a similar magnitude at 0.773 (Table A1). This suggests that consumers do not learn about size changes or that learning takes place over a much longer time period. Given the long interpurchase times for pepper, it may take several years to see the effects of learning. Because of this, my model does not incorporate learning, but instead assumes that consumers are either attentive or inattentive to a change in product size.

B. *Inventory Dynamics*

In the section that follows, I consider a static demand model that incorporates inattention. The model abstracts away from consumers' dynamic inventory decisions. If consumers do engage in stockpiling behavior and time their purchases when prices are low, static demand estimation will produce biased estimates of consumer preferences (Hendel and Nevo, 2006). Stockpiling is a concern here due to pepper's high storability.

The consumer purchase data suggests that stockpiling is uncommon. Examining household purchases from 2014 to 2016, most consumers purchase pepper at regular prices rather than at promoted prices (Figure A7). Only 12% of purchases involved a coupon or another promotion. [3] This suggests that consumers purchase pepper when they need more, not when prices are low.

The large number of purchases that occur at regular prices also reflects the infrequency of promotions. Stores rarely discount pepper products. Considering stores with available promotion data reveals that the typical pepper product was on sale for one week during the entire two-year period from 2014 to 2016. Even when consumers purchase products on sale, most restrict themselves to a single package (Figure A7). More than 80% of purchases that occur during promotions are for a single unit. In general, multi-purchases are rare with only 8% of purchases involving more than one unit and less than 2% involving more than two units. Overall, pepper purchases seem to reflect ordinary consumption decisions.

VI. Product Choice under Inattention

A. *Model*

In period t, M_{kt} consumers visit store k looking to buy pepper. Each consumer selects one product from the available pepper products J_{kt} or selects the no-purchase option 0. Consumer i's *actual* utility from purchasing product j is:

$$U^a_{ijkt} = x_{jkt}\beta + \gamma_i z_{jkt} - \alpha p_{jkt} + \xi_{jkt} + \epsilon_{ijkt} \tag{1}$$

where x_{jkt} is a set of observable characteristics; p_{jkt} is the price; z_{jkt} is the current net weight; ξ_{jkt} is the unobserved product attributes; and ϵ_{jkt} is a random shock. The utility of the outside option

[3] I construct the indicator for a promotion as a combination of the variables *coupon_value*, *deal_flag_uc*, and *is_promotion*.

is:

$$U^a_{i0kt} = 0 + \epsilon_{i0kt} \tag{2}$$

A portion of consumers may be inattentive and misevaluate the weights of the downsized products. Some may misevaluate sizes because they remember the old weight, but not notice the reduction. These consumers simply assume that the net weight has not changed since their last purchase. Others may evaluate product weights based on package sizes and mistakenly conclude that the downsized products have the same weight as rival products since they have the same package size.[4] Regardless why inattention occurs, inattentive consumers use the original product size, whereas attentive consumer use the current product size.

The model allows consumers to be inattentive to some products, but not others. With L downsized products, there are 2^L combinations of downsized products to which a consumer can be inattentive. In the black pepper industry, there are five downsized products and hence 32 possible combinations. Let τ_{ij} be an indicator for whether consumer i is inattentive to downsized product j. An inattentive consumer evaluates downsized product j using its original weight and *perceives* his utility from j as:

$$\begin{aligned} U^p_{ijkt} &= x_{jkt}\beta + \gamma_i z_{jk0} - \alpha p_{jkt} + \xi_{jkt} + \epsilon_{ijkt} \\ &= x_{jkt}\beta + \gamma_i z_{jkt} + \gamma_i\big(z_{jk0} - z_{jkt}\big) - \alpha p_{jkt} + \xi_{jkt} + \epsilon_{ijkt} \\ &= U^a_{ijtk} + \tau_{ij} \cdot \gamma_i\big(z_{jk0} - z_{jkt}\big) \end{aligned} \tag{3}$$

where z_{jk0} is the original weight before downsizing. In my context, product sizes change once or not at all. If the size changes in period t', $z_{jkt} = z_{jk0}$ for all periods $t < t'$. In contrast to actual utility, perceived utility depends both on the current and original sizes. Inattention drives a wedge between the perceived and actual utility for downsized product j equal to $\gamma_i\big(z_{jk0} - z_{jkt}\big)$. Thus, inattention causes the size change to enter perceived utility as an additional product characteristic with a product specific random coefficient. I denote the cumulative distribution function of the random coefficients as $G(\boldsymbol{\tau_i}, \gamma_i)$ where $\boldsymbol{\tau_i} = (\tau_{i1}, \tau_{i2}, \tau_{i3}, \tau_{i4}, \tau_{i5})$.

Because a consumer's type is not observable, perceived utility has a latent structure. A latent class consists of a combination of downsized products for which a consumer is inattentive. Consider the five downsized products ordered from smallest to largest in terms of the absolute size change. Consumers that belong to the latent class $\{1, 2, 4\}$ do not notice the change in size of products 1, 2 and 4. There are 32 possible latent classes. The joint distribution $G(\boldsymbol{\tau}, \gamma)$ dictates the probability of observing any one type and hence the latent structure.

[4]In this case, consumers misevaluate product size only when rival products occupy a large enough shelf space. Since the shelf space devoted to a product is not observable in the Nielsen data, I cannot model inattention stemming from a reference size.

The model accomodates many types of inattention. Complete attention corresponds to the case where all consumers belong to the class, $\emptyset$. In contrast, complete inattention corresponds to the case where all consumers belong to the class $\{1,2,3,4,5\}$. Another possibility is that consumers notice changes above a certain threshold (e.g. Han, Gupta and Lehmann, 2001). In this case, consumers who notice small changes in size must notice larger ones. Under threshold perception, consumers must fall into one of six classes $\big\{\emptyset,\{1\},\{1,2\},\{1,2,3\},\{1,2,3,4\},\{1,2,3,4,5\}\big\}$. The model can also capture differences in the visual saliency of the size change stemming from differences in the packaging (Milosavljevic et al., 2012). Unlike pepper tins, pepper grinders use a translucent material, allowing consumers to judge the fill level. Consumers who notice the size decrease for grinders, but not tins would belong to the class $\{1, 2, 4\}$. Thus, different assumptions about the type of inattention place different restrictions on the possible classes. The modeling structure is flexible enough to find any of these outcomes as well as others.

Assuming that the random taste shock ϵ is drawn i.i.d. from a Type I extreme value distribution, the store share for product j at store k in period t conditional on the random coefficients is:

$$s_{jkt}(\boldsymbol{\beta},\alpha,\boldsymbol{\tau}_i,\gamma_i) = \frac{\exp\big\{x_{jkt}\beta - \alpha p_{jkt} + \xi_{jkt} + \gamma_i z_{jkt} + \tau_{ij}\cdot\gamma_i\big(z_{jk0} - z_{jkt}\big)\big\}}{1+\sum_{l\in J_{kt}}\exp\big\{x_{lkt}\beta - \alpha p_{lkt} + \xi_{lkt} + \gamma_i z_{lkt} + \tau_{il}\cdot\gamma_i\big(z_{lk0} - z_{lkt}\big)\big\}} \tag{4}$$

Integrating over the joint distribution of the random coefficients, the unconditional store share for j in period t is:

$$s_{jkt} = \int s_{jkt}(\boldsymbol{\beta},\alpha,\boldsymbol{\tau_i},\gamma_i)dG(\boldsymbol{\tau_i},\gamma_i) \tag{5}$$

and the expected demand for product j in period t at store k is then:

$$Q_{jkt} = s_{jkt}M_{kt} \tag{6}$$

The model ignores store choice. This abstraction is reasonable as consumers select a store based on a basket of products rather just than pepper (Thomassen et al., 2017).[5] As a result, pepper prices are likely not an important determinant of store choice.

B. Identification

When estimating demand, price endogeneity is concern. Unmeasured factors, such as coupon availability, advertising, and shelf space, affect demand and hence prices. The omission of these factors will produce a biased estimate of the price coefficient α. To address the issue of endogeneity, I instrument for price using supply-side instruments.

Exogenous variation that shifts relative market shares identifies the distribution of γ. By relaxing

[5]In the consumer panel data, every household purchases pepper with another product.

the independence of irrelevant alternatives assumption inherent in the standard logit model, the random coefficient γ allows relative market shares to vary with the number of similar products and therefore governs the extent to which consumers substitute between similarly sized products. In my case, the main driver of consumer substitution is differences in product offerings across stores before downsizing. Throughout this period, stores offered a wide array of products and sizes. Different product offerings induce consumers to make different choices and thus reveals consumers' substitution patterns.

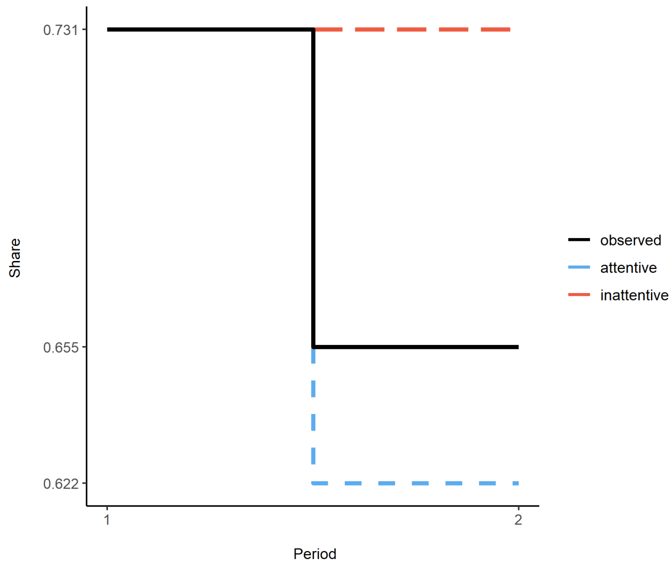
Without existing variation to identify γ, separating inattention from size preferences would be difficult. Consumers may choose downsized products because they prefer the new, smaller size. Alternatively, they may prefer larger sizes, but choose downsized products because they are inattentive. Thus, attentive consumers who prefer smaller packages make the same choices as inattentive consumers who prefer larger ones.

One way to separate inattention from size preferences would be to consider stores that offered both the original and downsized products at the same time. As stores transitioned to the downsized product, some continued to offer the original product to clear out their existing stock. During this transition period, these stores offered the original and downsized versions at the same time, and in some cases, offered the products side by side (e.g. Figure A3). If consumers are inattentive, they would not have noticed the difference between the versions and would have purchased them at the same frequency. Unfortunately, there are too few instances to use this as a way to identify the inattention parameters $\boldsymbol{\tau}$. Instead, I rely on deviations from consumers' size preferences to identify the distribution of the inattention parameters.

When a product's size decreases, its share should change in line with consumers' size preferences. Inattention will dampen this response. Thus, a smaller than expected change in market shares or no change at all would indicate inattention. This argument assumes that size preferences are time-invariant. If size preferences change over time, a shift in size preferences in favor of smaller sizes would also explain a smaller than expected decline after downsizing. The time-invariance of size preferences is thus the main identifying assumption.

As an illustrative example, consider a store that offers a single product over two periods $t = 1, 2$. In both periods, consumers can select the product or an outside option. The utility of the product is $u_{1t} = \gamma z_t + \epsilon_{1t}$ where z_t is the product's size in period t and ϵ_{1t} is a random shock drawn from a Type I extreme value distribution and the utility of the outside option is $u_{0t} = 0 + \epsilon_{0t}$. Initially, the product's size is 2 and its market share is $s_1 = \frac{e}{1+e} \approx 0.731$. The product's size and market share imply a value of γ equal to $\frac{1}{2}$. Before period 2, the size of product declines from 2 to 1. If consumers are fully attentive, the product's market share in period 2 will decrease in line with consumers' size preferences to $s_2 = \frac{e^{0.5}}{1+e^{0.5}} \approx 0.622$ (the red line). However, if consumers are fully inattentive, they will evaluate the product's utility using the original size $z_1 = 2$ and the product share would remain constant (the blue line). In reality, the observed product share (the black line) reflects a combination of attentive and inattentive consumers. The observed market share is

$s_2 = \alpha \frac{e^{0.5}}{1+e^{0.5}} + (1-\alpha)\frac{e}{1+e}$ where α is the fraction of inattentive consumers. The greater the fraction



of attentive consumers the closer the observed trend is to the expected trend (a smaller vertical distance between the black and blue lines). Because of this, the difference between the observed change in the product's share and the expected change if consumers were attentive identifies the percentage of inattentive consumers. This comparison is possible only because the size preferences γ are observable from the initial period and do not change over time. If γ is unknown or changes over time, I could not determine the product's share under complete attention.

C. Estimation Framework

Standard approaches to estimating demand assume a parametric form for the distribution of the random coefficients, which places unwanted restrictions on the latent structure. For greater flexibility, I use semi-nonparametric approach that recovers the model parameters in two stages. In the first stage, I estimate the fixed parameters $(\boldsymbol{\beta}, \alpha, \xi)$ using the standard logit. In the second stage, I set the fixed coefficients to their values from the first stage and then estimate the distribution of the random coefficients nonparametrically using the fixed grid estimator from Fox, Kim, Ryan, and Bajari (2011), hereafter FKRB, and its generalization from Heiss, Hetzenecker, and Osterhaus (2021), hereafter HHO.

The first-stage treats the random coefficients $(\gamma, \boldsymbol{\tau})$ as fixed constants. In essence, I am replacing the random coefficients with their means $(\overline{\gamma}, \overline{\boldsymbol{\tau}})$. Given this, the market shares satisfy:

$$\log \frac{s_{jkt}}{s_{0kt}} = x_{jkt}\beta + \overline{\gamma} z_{jkt} + \overline{\tau_j} \cdot \overline{\gamma}\big(z_{jk0} - z_{jkt}\big) - \alpha p_{jkt} + \xi_{jkt} \tag{7}$$

The model parameters are estimable using linear IV. Instruments are necessary since prices p_{jkt}

are correlated with unobserved product characteristics ξ_{jkt}. For the instruments, I use the differentiation instruments from Gandhi and Houde (2019) and the BLP instruments (Berry, Levinsohn and Pakes, 1995).

In the second-stage, I do not make any functional form assumptions on the distribution of the random coefficients. Instead, I approximate the true distribution using a mixture of point masses. This fixed grid estimator divides the support of $G(\boldsymbol{\tau}, \gamma)$ into R fixed grid points. Each grid point r has a corresponding probability weight θ_r. This fixed grid estimator approximates the true shares as:

$$s_{jkt} \approx \sum_{r=0}^{R} \theta_r s_{jkt}(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \widehat{\xi}, \boldsymbol{\tau_r}, \gamma_r) \tag{8}$$

where $(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \widehat{\xi})$ are the estimates from the first stage. The conditional shares at each grid point come from (4). The fixed grid estimator transforms the estimation of the shares into a linear probability model with the shares at each grid point as the explanatory variables. To ensure the weights form a valid probability distribution, FKRB require the weights to be nonnegative and sum to one. The weights are the solution to the minimization problem:

$$\begin{aligned} \widehat{\theta}^{\text{FKRB}} = \arg\min_{\theta} \quad & \frac{1}{N}\sum_{t=1}^{T}\sum_{k=1}^{K_t}\sum_{j=0}^{J_{kt}}\left(s_{jkt} - \sum_{r=1}^{R}\theta_r s_{jkt}(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \widehat{\xi}, \tau_r, \gamma_r)\right)^2 \\ \text{s.t.} \quad & \theta_r \geq 0, \quad r = 1, ..., R \\ & \sum_{r=1}^{R}\theta_r = 1 \end{aligned} \tag{9}$$

Given the estimated probability weights, the estimated cumulative distribution function is:

$$\widehat{G}(\boldsymbol{\tau}, \gamma) = \sum_{r=1}^{R} \widehat{\theta}_r \, \mathbb{1}\left[\boldsymbol{\tau_r} \leq \boldsymbol{\tau}, \gamma_r \leq \gamma\right] \tag{10}$$

and the percentage of consumers who are inattentive to product j is:

$$\widehat{\phi_j} = \sum_{r=1}^{R} \widehat{\theta}_r \, \mathbb{1}\left[\tau_{rj} = 1\right] \tag{11}$$

Heiss, Hetzenecker and Osterhaus (2021) note that the FKRB estimator tends to generate very few positive weights. This sparsity stems from the fact that FKRB estimator is Nonnegative LASSO (NNL) for a specific shrinkage parameter. The connection to NNL becomes apparent when the minimization problem is rewritten to exclude the Rth grid point. Using the fact that the $R+1$th probability weight satisfies $\theta_r = 1 - \sum_{r=1}^{R-1} \theta_r$, the minimization problem becomes

$$\begin{aligned} \widehat{\theta}^{\text{FKRB}} = \arg\min_{\theta} \quad & \frac{1}{N} \sum_{t=1}^{T} \sum_{k=1}^{K_t} \sum_{j=0}^{J_{kt}} \left(\tilde{s}_{jkt} - \sum_{r=1}^{R-1} \theta_r \tilde{s}_{jkt}(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \widehat{\xi}, \boldsymbol{\tau_r}, \gamma_r) \right)^2 \\ \text{s.t.} \quad & \theta_r \geq 0, \quad r = 1, ..., R \\ & \sum_{r=1}^{R-1} \theta_r \leq 1 \end{aligned} \tag{12}$$

where $\tilde{s}_{jkt} = s_{jkt} - \theta_r s_{jkt}(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \widehat{\xi}, \boldsymbol{\tau_R}, \gamma_R)$ and $\tilde{s}_{jkt}(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \widehat{\xi}, \boldsymbol{\tau_r}, \gamma_r) = s_{jkt}(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \widehat{\xi}, \boldsymbol{\tau_r}, \gamma_r) - \theta_r s_{jkt}(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \widehat{\xi}, \boldsymbol{\tau_R}, \gamma_R)$. This is the NNL problem for a specific tuning parameter. The FKBR estimators acts like NNL and shrinks some of the probability weights to zero when the first $R-1$ probability weights are greater than 1. Similar to NNL, the FKRB estimator will not yield a unique solution when the grid shares are strongly correlated (Heiss, Hetzenecker and Osterhaus, 2021). With a dense grid, the FKRB estimator will give only a few of the correlated points positive weights and set the rest to zero.

To avoid the sparsity and selection problems associated with the FKRB estimator, I use a generalized version from Heiss, Hetzenecker and Osterhaus (2021). HHO add an ℓ_2 constraint to (12). The minimization problem becomes:

$$\begin{aligned} \widehat{\theta}^{\text{HHO}} = \arg\min_{\theta} \quad & \frac{1}{N} \sum_{t=1}^{T} \sum_{k=1}^{K_t} \sum_{j=0}^{J_{kt}} \left(\tilde{s}_{jkt} - \sum_{r=0}^{R-1} \theta_r \tilde{s}_{jkt}(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}, \boldsymbol{\tau_r}, \gamma_r) \right)^2 \\ \text{s.t.} \quad & \theta_r \geq 0, \quad r = 1, ..., R-1 \\ & \sum_{r=1}^{R-1} \theta_r \leq 1, \quad \sum_{r=1}^{R-1} \theta_r^2 \leq u \end{aligned} \tag{13}$$

The quadratic constraint transforms the problem into the equivalent of nonnegative elastic net (Wu and Yang, 2014). This constraint prevents any one coefficient from being too large and ensures that closeby grid points receive positive weight, resulting in a smoother distribution (Heiss, Hetzenecker and Osterhaus, 2021). When the quadratic constraint does not bind, the HHO estimator reduces to the FKRB estimator. Because of this, the elastic net estimator will approximate the true distribution at least as well as the FKRB estimator.

Ultimately, the performance of the HHO estimator rests on the choice of the grid points and the choice of the tuning parameter u. These choices are nontrivial. To obtain the best possible approximation, the grid values should span the true support, which is unknown. In addition, while having more grid points may provide greater flexibility, it also results in more highly correlated points that result in selection problems. To deal with these issues, I take a data-driven approach to selecting the grid and tuning parameters. Given a set of candidate u-values $\{u_1, \ldots, u_K\}$ and grids $\{\mathcal{B}_1, \ldots, \mathcal{B}_L\}$, I choose the combination that minimizes the 5-fold cross-validated mean squared error.[6] The grids differ in their support for γ as τ_j can only take the values 0 or 1. I center all of the values for γ around its first-stage estimate. After finding the best grid and u-value, I resolve (13) to obtain the final estimates. The cross-validation is the most computationally intensive step since it requires the model to be fit five times for set of parameter values. To reduce computation time, I perform the cross-validation on a random subset of the data.

The two-stage approach has several advantages over alternatives such as Berry, Levinsohn and Pakes (1995) or FKBR. In constrast to BLP, the fixed grid estimator in the second stage provides nonparametric estimates for the distribution of the random coefficients. Like most nonparametric methods, the fixed grid estimators suffer from the curse of dimensionality. Using the standard logit to recover the fixed parameters greatly reduces the dimensionality of the problem and overcomes a shortcoming of FKBR. Another advantage of the two-stage procedure is that it can accomodate random measurement error, unlike BLP (Fox et al., 2011). BLP assumes that the observed shares are the true expected shares and even a small degree of measurement error produces large bias in the parameter estimates (Fox et al., 2011; Berry, Linton and Pakes, 2004). Accomodating measurement error is important since observed store shares are a noisy signal of the expected shares since few consumers purchase pepper each month.

D. Inference

To quantify the uncertainty around my estimates, I use subsampling (Politis, Romano and Wolf, 1999). Subsampling involves drawing blocks of size b from the original dataset. It then uses the variation of the estimate across these subsamples to construct confidence intervals and standard errors, which adjust for the smaller sample size. I generate subsamples of size b and then estimate the two-stage procedure on each. Rather than use every possible subsample of this size, I randomly select 500 subsets to reduce computation time.

FKBR (2011) show that for their estimator, subsampling is pointwise consistent in constrast to the standard bootstrap, which is inconsistent at points near the ends of the support (Andrews, 1999). Subsampling, however, is not uniformily consistent and may result in undercoverage for values the ends of the support (Andrews and Guggenberger, 2010). In general, the coverage of subsampled confidence intervals can be sensitive to the choice of the subsample size b. Following HHO, I choose the subsample size according to the rule of thumb from Jentsch and Leucht (2016) as $b = (N)^{2/3} = 40,383$ where N is the number of observations.

[6]Heiss, Hetzenecker and Osterhaus (2021) recommend choosing the value of u based on the one standard error rule (Hastie, Tibshirani and Friedman, 2009) rather than choosing based on the lowest mean squared error. The rule states that between two models with similar cross-validated errors, the researcher should choose the simpler one. For the HHO estimator, this means selecting the lowest value of u whose error rate is within one standard error of the lowest error rate. Due to the small variation across folds, I find that the two criteria tend to yield the same u-value.

VII. Results

A. The Degree of Inattention

Since the number of potential pepper customers M_{kt} is unobservable, I proxy for it using total sales in the pepper product category. In addition to black pepper, this product category includes various seasoning blends and chile powders. I assume that the number of potential pepper customers at a store is equal to the total sales of all such products at that store. I do not use the number of vistors to a store since some consumers never buy pepper.

With the shares defined, I apply my two-stage procedure to the pepper data. The coefficients from the first stage have the correct, expected signs (Table A2). Notably, the price coefficient is negative indicating that consumers prefer less expensive products and the size coefficient is positive indicating that consumers prefer larger sizes on average. These coefficients imply a mean own-price elasticity of -4.03 and a mean own-size elasticity of 1.48. The relative magnitudes of the elasticities suggest that in the absence of inattention, consumers would be almost three times more responsive to an increase in price than to a decrease in product size. This differential sensitivity to price and size suggests that downsizing can be an effective strategy for raising unit prices even when consumers are fully informed.

Different sensitivities to price and size alone do not indicate inattention; however, the significance of the downsizing terms does. All of the downsizing coefficients in the first stage are positive, implying that consumers prefer the smaller sizes. At the same time, the size coefficient indicates that consumers prefer larger sizes. Inattention reconciles this apparent contradiction.

Setting the fixed parameters equal to their first-stage values, I estimate the distribution of the size and inattention parameters using the HHO estimator with the grid and u-value selected by cross-validation. The optimal grid for γ is an equally spaced sequence of 25 points ranging from -0.017 to 0.983. The endpoints of this grid correspond to a window of 0.5 around the first stage estimate. The u-value with the lowest cross-validated mean squared error is 0.34, which implies a fairly high degree of smoothing.

The second-stage estimates indicate that consumers exhibit heterogenous size preferences. The marginal distribution of γ is approximately symmetric and bell-shaped with a mode around 0.5, which is close to the value of first-stage estimate (Figure A8). Additionally, the distribution has virtually no mass outside the interval from 0.4 to 0.7.

If consumers are fully attentive, the marginal distribution of $\boldsymbol{\tau_i}$ should be degenerate and consist of a single mass point at $\boldsymbol{\tau_i} = \mathbf{0}$. To the contrary, the marginal distribution of inattention parameters has positive mass at points where the coefficients are one, implying some degree of inattention. Subsetting to γ-values with non-zero mass,[7] I calculate the probability of being inattentive to a product conditional on the size preference γ_i (Figure A9). With the expection of the medium tin, the conditional probability of being inattentive to a product is around 0.5 and is roughly constant across

[7] A mass greater than 10^{-4}

values of γ_i. This result implies that consumers who are sensitive to product sizes are not more attentive than those who care little for product size. For the medium tin, the probability of being inattentive increase from 0.5 to over 0.75 as γ increases, implying that size-sensitive consumers are more likely to be inattentive. Integrating over the values of γ_i gives the unconditional probability of being inattentive. I find that approximately 50% of consumers are inattentive to any given product (Table 3). In addition, the level of inattention varies only slightly with the packaging material and the size of the reduction. For instance, the same percentage of consumers are innattentive to the small tin with its opaque packaging as to the small grinder with its transparent packaging. Slightly more consumers fail to notice the change in the medium tin. This may stem from the greater availability of medium tins. If consumer mistakenly believe that products with the same package size have the same net weight, having more rival products with the original size would reinforce consumers' mistaken beliefs and result in greater inattention.

Table 3—: Degree of Inattentiveness by Product

	HHO
Product	Percent
Large Tin	0.513
	(0.0001)
Medium Tin	0.636
	(0.0083)
Medium Grinder	0.512
	(0.0000)
Small Tin	0.519
	(0.0005)
Small Grinder	0.519
	(0.0002)
Any Inattention	0.978
	(0.0005)

Note: Subsampled standard errors in parantheses.

Consumers who are inattentive to one product may be inattentive to others. Since the estimation procedure recovers the full joint distribution, it is able to recover the correlation structure. My results suggest that inattention is not correlated across products (Table 4). The largest correlation between products is 0.024. Since inattention is uncorrelated across products, the set of consumers who are inattentive to one product is different than the set of consumers who are inattentive to another product. As a result, the fraction of consumers who exhibit inattention is larger than the fraction who are inattentive to any one product. Considering all grids values where at least one inattention indicator is equal to one, I find that 97.8% of consumers fail to notice at least one size change. Thus, almost every consumer is inattentive.

Table 4—: Correlation Matrix

		τ_1	τ_2	τ_3	τ_4	τ_5
Large Tin	τ_1	1.000				
Medium Tin	τ_2	0.017	1.000			
Medium Grinder	τ_3	0.024	0.018	1.000		
Smaller Tin	τ_4	0.023	0.015	0.024	1.000	
Smaller Grinder	τ_5	0.023	0.015	0.024	0.023	1.000

Note: Calculated from the HHO estimates.

While the model recovers the degree of consumer inattention, it does capture the underlying mechanisms that drive inattention. I test whether various mechanisms explain the observed pattern of inattention by considering what fraction of consumers belong to latent classes consistent with a particular mechanism (Table 5). The model indicates that 91.5% consumers exhibit some type of partial inattention, so most consumers fall between the extremes of complete attention and complete inattention. Additionally, the estimates rule out threshold perception as the reason for this partial inattention. Under threshold perception, consumers must belong to one of the following latent

Table 5—: Types of Inattention

	HHO
Type of Inattention	Percent
Partial Inattention	0.915 (0.0001)
Threshold Perception	0.325 (0.0001)
Visibility of the Fill Line	0.219 (0.0004)

Note: Subsampled standard deviations in parantheses.

classes $\big\{\emptyset, \{1\}, \{1,2\}, \{1,2,3\}, \{1,2,3,4\}, \{1,2,3,4,5\}\big\}$. Summing the probability of all grid points corresponding to these classes, I find that only 32.5% of consumers exhibit inattention consistent with threshold perception. Finally, I examine whether inattention stems from the inability to see the fill line. If the visibility of the fill line is a major reason for inattention, most consumers should be inattentive to tins and not grinders. However, a mere 21.9% of consumers are inattentive to just tins. These results suggest that inattention may take many forms.

B. Robustness Check

One problem with the two-stage approach is that bias in the first-stage estimates can produce bias in the second stage. Due to its inflexibility, the standard logit may fail to recover the means of the random coefficients and hence generate biased estimates of the fixed parameters. Because of this, I experiment with using BLP in the first stage. Houde and Myers (2019) recommend using BLP since its flexibility should allow it to recover the means of the random coefficients. Using BLP with the random coefficients specified as independent normals did not significantly impact the second-stage results.

VIII. The Impact of Inattention

A. Pricing

By distorting product utilities, inattention affects demand and hence prices. Removing inattention may cause stores to adjust prices and other product features. I assume that stores adjust only price after the removal of inattention. The new prices p^a_{kt} represent the prices that the stores would set if $\boldsymbol{\tau_i}$ was set to zero. I recover these counterfactual prices from the demand-side estimates by making assumptions on the supply-side model.

Because consumers choose stores based on basket of goods, pepper prices are unlikely to impact consumers' choice of stores. Consquently, stores will act as local monopolists when pricing pepper. Under this assumption, the store k's profits in period t is:

$$\pi_{kt} = \sum_{j \in J_{kt}} \left[p_{jkt} - w_{jkt} - mc^s_{jkt} \right] s_{jkt}(p_{kt}) \tag{14}$$

where w_{jkt} is the wholesale price of product j and mc^s_{jkt} is the store's marginal cost of product j. Differentiating with respect to prices, the first-order conditions are:

$$s_{jkt} + \sum_{j \in J_{kt}} \left[p_{jkt} - w_{jkt} - mc^s_{jkt} \right] \frac{\partial s_{jkt}}{\partial p_{jkt}} = 0 \tag{15}$$

I do not consider changes in wholesale prices. Such pricing behavior is consistent with a number of models in which manufacturers set the wholesale margin to zero. A zero wholesale margin can arise from the use of a nonlinear pricing or substantial retailer bargaining power Villas-Boas (2007). The first-order conditions now becomes:

$$s_{jkt} + \sum_{j \in J_{kt}} \left[p_{jkt} - mc^m_{jkt} - mc^s_{jkt} \right] \frac{\partial s_{jkt}}{\partial p_{jkt}} = 0 \tag{16}$$

where mc^m_{jkt} is the manufacturer's marginal cost of product j. Stacking the first-order conditions and rearranging terms, the optimal prices satisfy:

$$p_{kt} + \Delta_{kt}^{-1} s_{kt}(p_{kt}) = mc^m_{kt} + mc^s_{kt} \tag{17}$$

where Δ_{kt} is a matrix with entry (m, n) equal to $\frac{\partial s_{mkt}}{\partial p_{nkt}}$ if store k sells products m and n during period t and zero otherwise. Because changes in attention affect prices through demand and not through marginal costs, store and manufacturer marginal costs remain the same after the removal of inattention.

When consumers are fully attentive, product demands $s^a_{kt}(p^a_{kt})$ do not depend on the size change. In addition, the response matrix Δ^a_{kt} now depends on the new demand with entry (m, n) equal to $\frac{\partial s^a_{mkt}}{\partial p_{nkt}}$. Given that marginal costs remain the same, the counterfactual prices p^a_{kt} satisfy:

$$p^a_{kt} + \Delta_{a,kt}^{-1} s^a_{kt}(p^a_{kt}) = p_{kt} + \Delta_{kt}^{-1} s_{kt}(p_{kt}) = mc^m_{kt} + mc^s_{kt} \tag{18}$$

This equation defines the counterfactual prices as an implicit function of the demand-side parameters, eliminating the need to estimate supply.

In theory, the removal of inattention represents a quality decrease for the downsized products since newly attentive consumers now find these products less attractive than before. The decrease in demand for the downsized products should result in lower prices for the downsized products and higher prices for the nondownsized ones.

Using the demand estimates, I solve equation (18) for the counterfactual prices using a Newton method. To reduce the computational burden, I consider a sample of 2000 stores that offer downsized products. On average, the prices of the downsized product decrease and the prices of the nondownsized products increase. However, most of the price changes are small and close to zero. Moreover, no price change exceeds \$0.04. The fact that prices remain approximately the same indicates that the change in demand is small, which makes sense given that the outside option represents a large shares at most stores.

B. Consumer Welfare

Inattentive consumers purchase the downsized products under the belief that their sizes have not changed. After purchasing, some of these consumers may experience discontent when they discover the smaller size. Post-purchase discontent does not necessarily translate into welfare losses. For inattention to reduce consumer welfare, inattention must cause the consumer to choose a product that they otherwise would not have chosen.

Consumers can choose the wrong product since they base their purchase decision on perceived utility rather than on actual utility. For instance, an inattentive consumer at store k in period t chooses the product that maximizes perceived utility $j^* = \underset{1,\ldots,J_{kt}}{\arg\max}\, U^p_{ijkt}$ instead of the one that maximizes actual utility is $m^* = \underset{1,\ldots,J_{kt}}{\arg\max}\ U^a_{ijkt}$. By choosing the incorrect product, the consumer experiences a loss in utility of:

$$\mathcal{W} = U^a_{im^*kt} - U^a_{ij^*kt} \tag{19}$$

Note that j^* and m^* depend on the random parameters $(\tau_{ij}, \gamma_i, \epsilon_{ijkt})$. Taking the expectation over these parameters gives the average welfare loss from imperfect knowledge:

$$\Delta CS = \frac{\mathbb{E}\big[\mathcal{W}\big]}{\alpha} = \frac{\mathbb{E}\big[U^a_{im^*kt}\big]}{\alpha} - \frac{\mathbb{E}\big[U^a_{ij^*kt}\big]}{\alpha} \tag{20}$$

As Train (2015) shows, this loss is equal to:

$$\begin{aligned}\Delta CS = \frac{1}{\alpha}\int\Bigg[&\log\Bigg(1+\sum_{J_{kt}}\exp\big\{x_{jkt}\beta - \alpha p_{jkt} + \xi_{jkt} + \gamma_i z_{jkt}\big\}\Bigg)\\ &-\log\Bigg(1+\sum_{J_{kt}}\exp\big\{x_{jkt}\beta - \alpha p_{jkt} + \xi_{jkt} + \gamma_i z_{jkt} + \tau_{ij}\cdot\gamma_i\big(z_{jk0} - z_{jkt}\big)\big\}\Bigg)\\ &-\sum_{\ell\in\{1,2,3,4,5\}} s_{\ell kt}(\boldsymbol{\tau},\gamma)\tau_{i\ell}\cdot\gamma_i\big(z_{\ell k0} - z_{\ell kt}\big)\Bigg]dG(\boldsymbol{\tau_i},\gamma_i)\end{aligned} \tag{21}$$

The first term is the standard log-sum formula based on actual utility. The log-sum formula is the closed form for the expectation from making the choice. The second term is the log-sum formula based on perceived utility and the final term is the average difference between actual and perceived utility. The final term includes only the downsized products since there is no difference between actual and perceived utilities for the nondownsized products.

Since prices changes very little after the removal of inattention, I calculate the welfare loss with respect to the existing prices p_{kt}. I find that inattention reduces a consumer's welfare by \$0.10 on average (Figure A10). The welfare loss ranges from around \$0.00 to \$0.75 and has a right-skewed distribution. The welfare loss varies across stores due to differences in product availability and pricing. The availability of the downsized products matters since if a store does not stock a particular downsized product, consumers cannot mistakenly choose that product. I also consider the welfare loss in relative terms as a percentage of the share-weighted average price at a given store during a given time period. The loss from inattention represents about 3% of the total product price (Figure A11). In general, the relative loss is small with most values falling between 0.0% and 5.0%.

On its own, welfare loss is relatively small, which is not surprising given that there are five downsized products and that consumers are almost three times more sensitive to price than size. However, taken across all downsized products, even a small welfare loss can add up. On a typical shopping trip, consumers will interact with dozens of downsized products ranging from chips to soap. In the aggregate, the loss due to inattention could be substantial.

Although inattention reduces welfare, this does not mean that downsizing is worse than its alternative, an increase in package prices. As Yonezawa and Richards (2016) show, firms may respond to a rival's decision to downsize by lowering their prices. These decreases may offset losses from inattention.

IX. Conclusion

The practice of product downsizing occurs across a wide range of products and represents one strategy that firms use to increase unit prices. When consumers underuse size information or ignore unit prices, downsizing represents a hidden price increase. I utilize a downsizing event in the black pepper industry to determine whether consumers are inattentive to decreases in product size. The large amount of existing size variation in this industry allows me to recover the degree to which consumers are inattentive.

To study how consumers respond to downsizing, I build a demand model that incorporates inattention to size changes and apply it to scanner data. In the model, inattentive consumers misperceive the sizes of the downsized products and as a result, they evaluate them based on their original sizes, instead of the actual sizes. Because of this, the size change enters utility as an additional product characteristic with a random coefficient, whose distribution I recover nonparametrically using a fixed grid estimator.

Remarkably, I find that nearly 98% of consumers are partially inattentive, failing to notice at least one size change and approximately 50% of consumers fail to notice the change in any given product. For black pepper, the degree of inattention does not depend on the magnitude of the size change or the visibility of the fill line. The estimated preferences' suggest that even if consumers were fully aware, they would be more responsive to price than product size.

Inattention reduces consumer welfare mainly by distorting consumers' product choices. I show that the removal of inattention does not have a major impact on stores prices. Under the existing pricing structure, inattention results in an average welfare loss of around $0.10, which represents 3% of the average product price, implying that downsizing is an effective obfuscation strategy.

My results suggest that statements of net weight do not prevent consumers from misperceiving product sizes. In fact, the vast majority of consumers appear to ignore such statements. Existing laws against deceptive packaging appear insufficient to address the myriad of ways in which firms downsize their products. The courts do not appear ready to extend existing laws to protect against the misreading of labels. In a case dealing with downsized Tylenol bottles, the court ruled that "the suggestion that such laws should cover [the plaintiffs'] failure to read an unambiguous tablet-count does not pass the proverbial laugh test" (*Fermin v. Pfizer*, 2016). Given the gap in existing packaging and consumer protection laws, additional laws requiring manufacturers to indicate fill levels or to announce downsizing could increase consumer attention to size changes and improve consumer welfare.

REFERENCES

Abaluck, Jason, and Giovanni Compiani. 2020. "A Method to Estimate Discrete Choice Models that is Robust to Consumer Search." Working Papers 26849, NBER. https://doi.org/10.3386/w26849.

Allcott, Hunt. 2013. "The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market." *American Economic Journal: Economic Policy*, 5(3): 30–66. https://doi.org/10.1257/pol.5.3.30.

Andrews, Donald W. K. 1999. "Estimation When a Parameter is on a Boundary." *Econometrica*, 67(6): 1341–1383. https://doi.org/10.1111/1468-0262.00082.

Andrews, Donald WK, and Patrik Guggenberger. 2010. "Asymptotic Size and a Problem with Subsampling and with the m out of n Bootstrap." *Econometric Theory*, 26(2): 426–468. https://doi.org/10.1017/S0266466609100051.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890. https://doi.org/10.2307/2171802.

Berry, Steve, Oliver B. Linton, and Ariel Pakes. 2004. "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems." *The Review of Economic Studies*, 71(3): 613–654.

Brown, Jennifer, Tanjim Hossain, and John Morgan. 2010. "Shrouded Attributes and Information Suppression: Evidence from the Field." *The Quarterly Journal of Economics*, 125(2): 859–876. https://doi.org/10.1162/qjec.2010.125.2.859.

Brown, Zach, and Jihye Jeon. 2020. "Endogenous Information and Simiplifying Insurance Choice." Working Paper.

Cakir, Metin, and Joseph V. Balagtas. 2014. "Consumer Response to Package Downsizing: Evidence from the Chicago Ice Cream Market." *Journal of Retailing*, 90(1): 1 – 12. https://doi.org/10.1016/j.jretai.2013.06.002.

Chandon, Pierre, and Nailya Ordabayeva. 2009. "Supersize in One Dimension, Downsize in Three Dimensions: Effects of Spatial Dimensionality on Size Perceptions and Preferences." *Journal of Marketing Research*, 46(6): 739–753. https://doi.org/10.1509/jmkr.46.6.739_JMR6C.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145–77. https://doi.org/10.1257/aer.99.4.1145.

Clerides, Sofronis, and Pascal Courty. 2017. "Sales, Quantity Surcharge, and Consumer Inattention." *Review of Economics and Statistics*, 99: 357–370. https://doi.org/10.1162/REST_a_00562.

Complaint at 7. n.d.. *Watkins Inc. v. McCormick & Co., Inc.* (8th Cir. 2015) (No. 15-2688).

***Fermin v. Pfizer, Inc.,*.** 15 CV 2133 (SJ)(ST)(E.D.N.Y. Oct. 14, 2016).

Feucht, Andrea. 2019. "How Long Do Spices Last?" https://www.mccormick.com/articles/mccormick/how-long-do-spices-last.

Fox, Jeremy T., Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari. 2011. "A simple estimator for the distribution of random coefficients." *Quantitative Economics*, 2(3): 381–418. https://onlinelibrary.wiley.com/doi/abs/10.3982/QE49.

Gandhi, Amit, and Jean-Francois Houde. 2019. "Heterogeneous (Mis-)Perceptions of Energy Costs: Implications for Measurement and Policy Design." Working Papers 26375, NBER. https://doi.org/10.3386/w26375.

Gourville, John, and Jonathan J. Koehler. 2004. "Downsizing Price Increases: A Greater Sensitivity to Price than Quantity in Consumer Markets." HBS Marketing Research Paper No. 04-01.

Grubb, Michael D. 2014. "Consumer Inattention and Bill-Shock Regulation." *The Review of Economic Studies*, 82(1): 219–257. https://doi.org/10.1093/restud/rdu024.

Gupta, Omprakash K., Sudhir Tandon, Sukumar Debnath, and Anna S. Rominger. 2007. "Package Downsizing: Is It Ethical?" *AI & SOCIETY*, 21(3): 239–250. https://doi.org/10.1007/s00146-006-0056-3.

Han, Sangman, Sunil Gupta, and Donald R. Lehmann. 2001. "Consumer Price Sensitivity and Price Thresholds." *Journal of Retailing*, 77(4): 435–456. https://doi.org/10.1016/S0022-4359(01)00057-4.

Heiss, Florian, Stephan Hetzenecker, and Maximilian Osterhaus. 2021. "Nonparametric estimation of the random coefficients model: An elastic net approach." *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.11.010.

Hendel, Igal, and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, 74(6): 1637–1673. https://doi.org/10.1111/j.1468-0262.2006.00721.x.

Houde, Sebastien, and Erica Myers. 2019. "Heterogeneous (Mis-)Perceptions of Energy Costs: Implications for Measurement and Policy Design." Working Papers 25722, NBER. https://doi.org/10.3386/w25722.

Hughlett, Mike. n.d.. "McCormick Says It Is Not Deceiving Customers with Product Labeling." *Star Tribune* (Minneapolis), June 10, 2015. https://www.startribune.com/mccormick-says-it-is-not-deceiving-customers-with-product-labeling/306832581/?refresh=true.

Imai, Satoshi, and Tsutomu Watanabe. 2014. "Product Downsizing and Hidden Price Increases: Evidence from Japan's Deflationary Period." *Asian Economic Policy Review*, 9(1): 69–89. https://doi.org/10.1111/aepr.12047.

***In Re: McCormick & Co., Inc., Pepper Prod. Mktg. & Sales Practices Litig.*,.** 15-1825 D.D.C. 10-14 (D.D.C. 2019).

Jentsch, Carsten, and Anne Leucht. 2016. "Bootstrapping sample quantiles of discrete data." *Annals of the Institute of Statistical Mathematics*, 68(3): 491–539. https://doi.org/10.1007/s10463-015-0503-3.

Joo, Joonwhi. 2018. "Quantity-Surcharged Larger Package Sales as Rationally Inattentive Consumers' Choice." Working Paper.

Kachersky, Luke. 2011. "Reduce Content or Raise Price? The Impact of Persuasion Knowledge and Unit Price Increase Tactics on Retailer and Product Brand Attitudes." *Journal of Retailing*, 87(4): 479 – 488. https://doi.org/10.1016/j.jretai.2011.08.001.

Krishna, Aradhna. 2006. "Interaction of Senses: The Effect of Vision versus Touch on the Elongation Bias." *Journal of Consumer Research*, 32(4): 557–566. https://doi.org/10.1086/500486.

Lacetera, Nicola, Devin G Pope, and Justin R Sydnor. 2012. "Heuristic thinking and limited attention in the car market." *American Economic Review*, 102(5): 2206–36. https://doi.org/10.1257/aer.102.5.2206.

Lennard, Dave, Vincent-Wayne Mitchell, Peter McGoldrick, and Erica Betts. 2001. "Why Consumers Under-use Food Quantity Indicators." *The International Review of Retail, Distribution and Consumer Research*, 11(2): 177–199. http://www.tandfonline.com/doi/abs/10.1080/09593960122918.

Luca, Michael, and Jonathan Smith. 2013. "Salience in Quality Disclosure: Evidence from the U.S. News College Rankings." *Journal of Economics & Management Strategy*, 22(1): 58–77. https://doi.org/10.1111/jems.12003.

Milosavljevic, Milica, Vidhya Navalpakkam, Christof Koch, and Antonio Rangel. 2012. "Relative Visual Saliency Differences Induce Sizable Bias in Consumer Choice." *Journal of Consumer Psychology*, 22(1): 67–74. https://doi.org/10.1016/j.jcps.2011.10.002.

Ochirova, Nadezhda. 2017. "The Impact of Shrinkflation on CPIH, UK: January 2012 to June 2017." `https://www.ons.gov.uk/economy/inflationandpriceindices/articles/theimpactofshrinkflationoncpihuk/january2012tojune2017`.

Ordabayeva, Nailya, and Pierre Chandon. 2013. "Predicting and Managing Consumers' Package Size Impressions." *Journal of Marketing*, 77(5): 123–137. `https://doi.org/10.1509/jm.12.0228`.

Politis, Dimitris N., Joseph P. Romano, and Michael Wolf. 1999. New York:Springer.

Thomassen, Øyvind, Howard Smith, Stephan Seiler, and Pasquale Schiraldi. 2017. "Multi-category Competition and Market Power: A Model of Supermarket Pricing." *American Economic Review*, 107(8): 2308–51. `https://doi.org/10.1257/aer.20160055`.

Train, Kenneth. 2015. "Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples." *Journal of Choice Modelling*, 16: 15–22. `https://doi.org/10.1016/j.jocm.2015.09.003`.

Villas-Boas, Sofia Berto. 2007. "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data." *The Review of Economic Studies*, 74(2): 625–652. `https://doi.org/10.1111/j.1467-937X.2007.00433.x`.

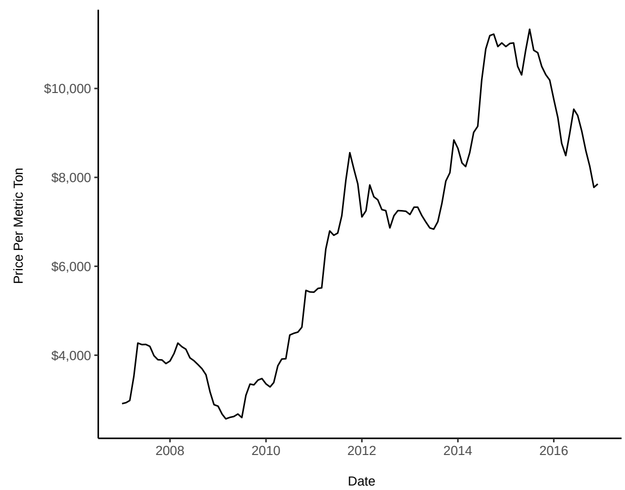
Wilkins, Stephen, Carina Beckenuyte, and Muhammad Mohsin Butt. 2016. "Consumers' Behavioural Intentions After Experiencing Deception or Cognitive Dissonance Caused by Deceptive Packaging, Package Downsizing or Slack Filling." *European Journal of Marketing*, 50(1/2): 213–235. `https://doi.org/10.1108/EJM-01-2014-0036`.

Wu, Lan, and Yuehan Yang. 2014. "Nonnegative Elastic Net and application in index tracking." *Applied Mathematics and Computation*, 227: 541–552. `https://doi.org/10.1016/j.amc.2013.11.049`.

Yonezawa, Koichi, and Timothy Richards. 2016. "Competitive Package Size Decisions." *Journal of Retailing*, 92(4): 445–469. `https://doi.org/10.1037/10.1016/j.jretai.2016.06.001`.

Figures and Tables

Figure A1. : Spot Price of Black Pepper in New York



Source: Pepper Statistical Yearbook 2018, International Pepper Community

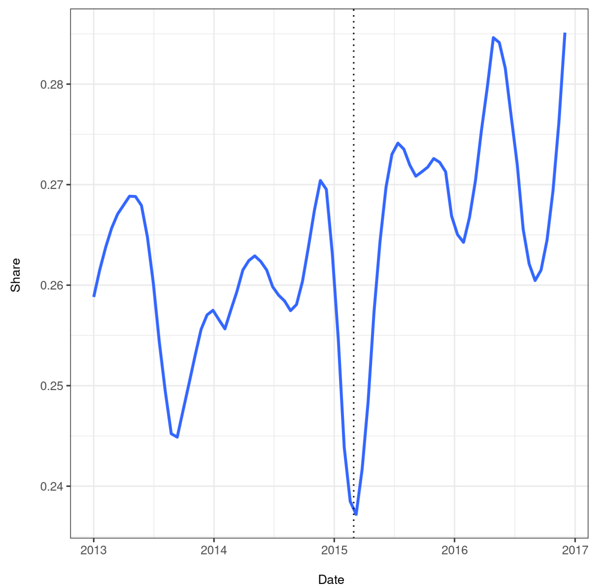
Figure A2. : Comparison of Medium Tins

Note: From left to right: Watkins's 4 oz tin, McCormick's old 4 oz tin, McCormick's new downsized 3 oz tin
Source: *Watkins v. McCormick* (2015, p. 7)

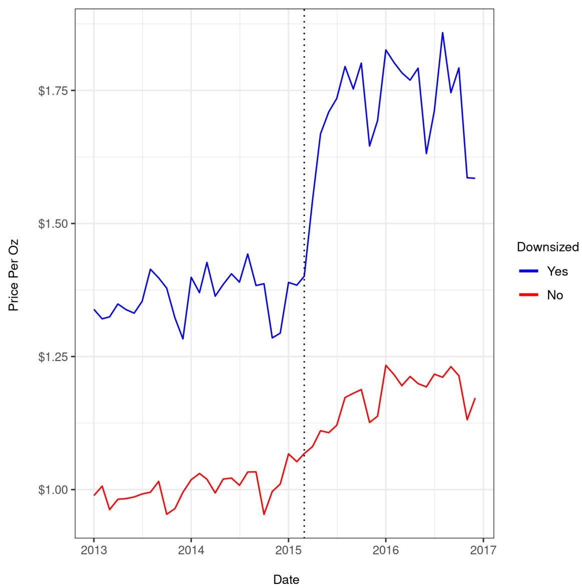
Figure A3. : Original and Downsized Versions Side by Side

Note: The downsized tin is on the right.

Source: *Watkins v. McCormick* (2015, p. 10)



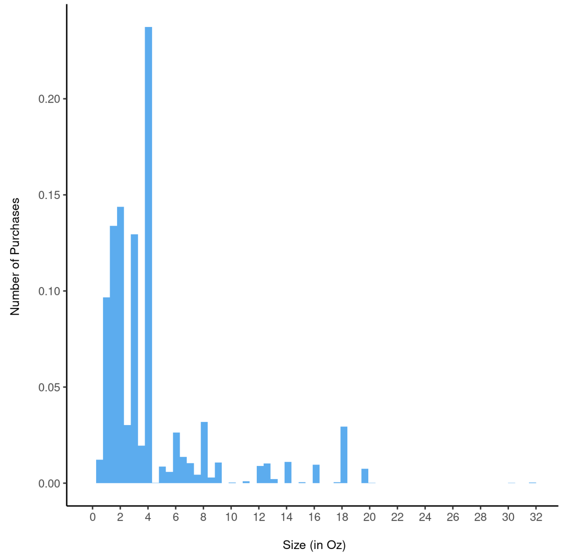
(a) Market Share of Downsized Products



(b) Comparison of Unit Prices

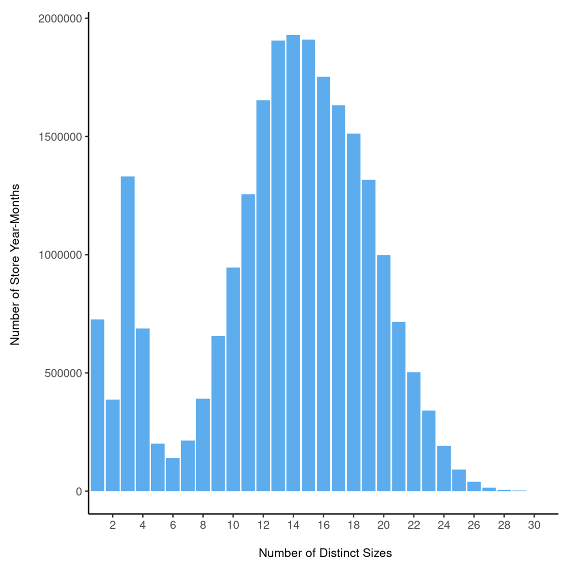
Note: The dashed line represents the date when McCormick started to ship its downsized products. The market share is for stores in the Nielsen Retailer Scanner data. The average price per ounce is weighted average by units solds.

Figure A5. : The Distribution of Sizes Purchased



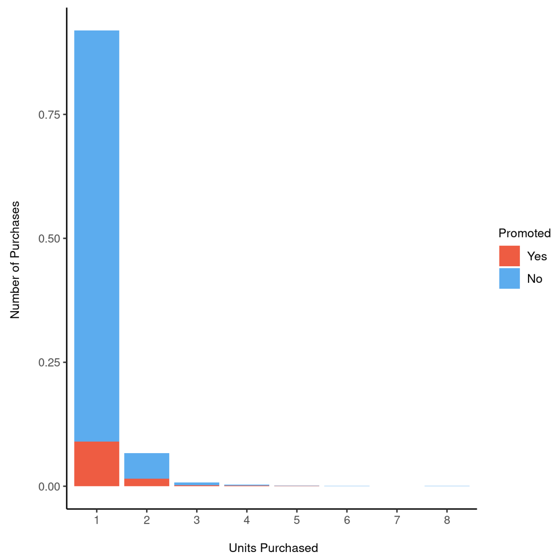
Note: Based on household purchases from 2014 to 2016.

Figure A6. : The Distribution of Distinct Sizes Offered



Note: Based on the scanner data from 2014 to 2016.

Figure A7. : The Distribution of Units Purchased



Note: Based on household purchases from 2012 to 2016.

Table A1—: Reduced-Form Evidence Against Learning

	(1)	(2)	(3)
Price	−0.074	−0.119	−0.124
	(0.018)	(0.034)	(0.034)
Size		0.081	0.240
		(0.021)	(0.062)
Size Squard			−0.009
			(0.003)
Downsized (0 Quarters After)	0.789	0.801	0.869
	(0.179)	(0.156)	(0.146)
Downsized (1 Quarter After)	1.107	1.098	1.160
	(0.200)	(0.162)	(0.151)
Downsized (2 Quarters After)	0.985	0.988	1.045
	(0.207)	(0.163)	(0.153)
Downsized (3 Quarters After)	0.903	0.912	0.965
	(0.228)	(0.177)	(0.168)
Downsized (4 Quarters After)	0.966	0.971	1.023
	(0.230)	(0.177)	(0.169)
Downsized (5 Quarters After)	0.845	0.826	0.880
	(0.236)	(0.181)	(0.173)
Downsized (6 Quarters After)	0.773	0.769	0.823
	(0.226)	(0.171)	(0.163)
Product FE	Y		
Brand FE		Y	Y
Year-Quarter FE	Y	Y	Y

Note: Based on scanner data from 2014 to 2016. The dependent variable is the log of quarterly units sold. The variables Downsizing (t Quarters After) are indicators for the downsized products for each quarter after the initial downsizing. The columns report clustered standard errors in parentheses. I cluster the standard errors by product and year-quarter.

Table A2—: First Stage

Variable	Estimate
constant	−0.0479
	(0.0021)
price	−1.0392
	(0.0008)
size	0.4831
	(0.0005)
whole	−0.1762
	(0.0017)
white	0.0883
	(0.0028)
large tin	0.5601
	(0.0072)
medium tin	0.7522
	(0.0021)
medium grinder	1.0696
	(0.0074)
small tin	−0.2183
	(0.0031)
small grinder	0.3771
	(0.0025)
size change: large tin	0.7937
	(0.0066)
size change: medium tin	0.6875
	(0.0026)
size change: medium grinder	1.3033
	(0.0151)
size change: small tin	1.0888
	(0.0075)
size change: small grinder	0.3447
	(0.0123)

Note: Robust standard errors in parantheses.

Figure A8. : The Marginal PMF of γ_i

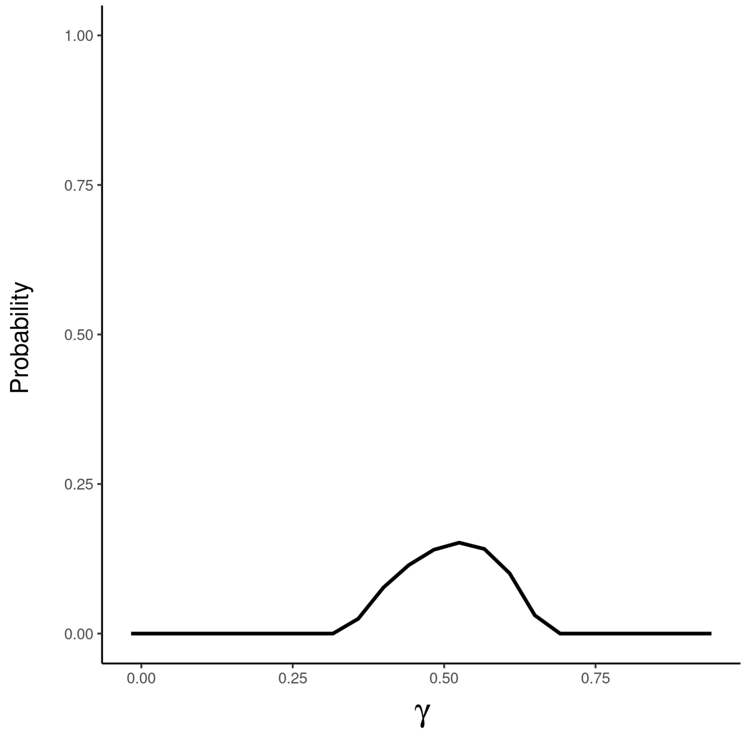
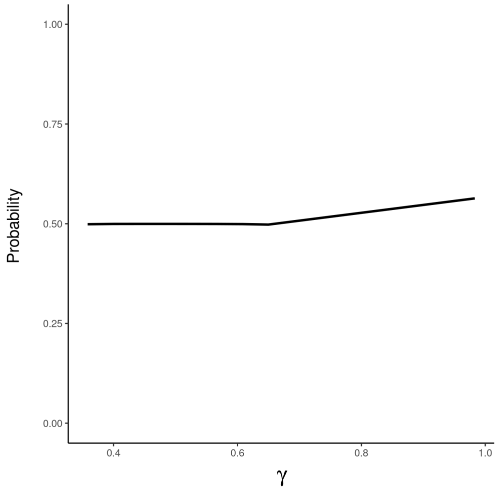
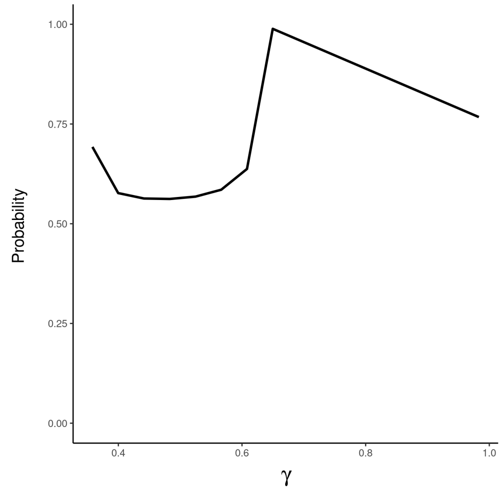


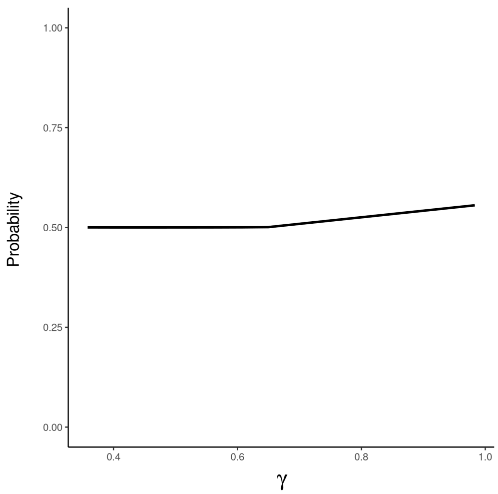
Figure A9. : The Conditional Probability of Being Inattentive $Pr(\tau_{ij} = 1|\gamma_i)$



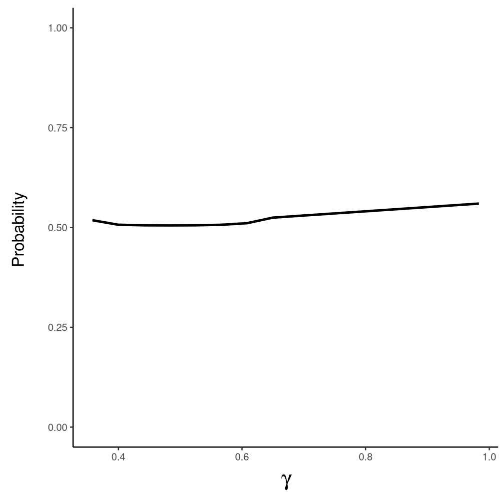
(a) Large Tin



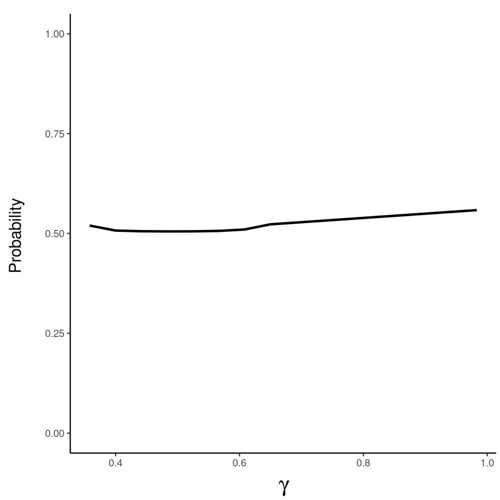
(b) Medium Tin



(c) Medium Grinder

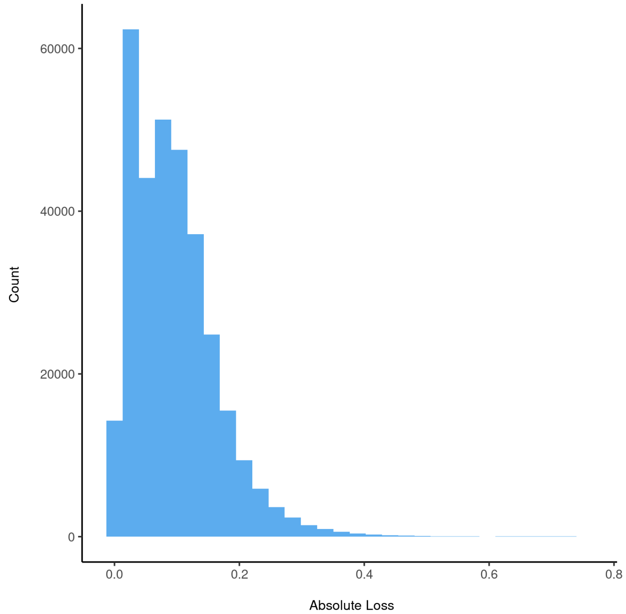


(d) Small Tin



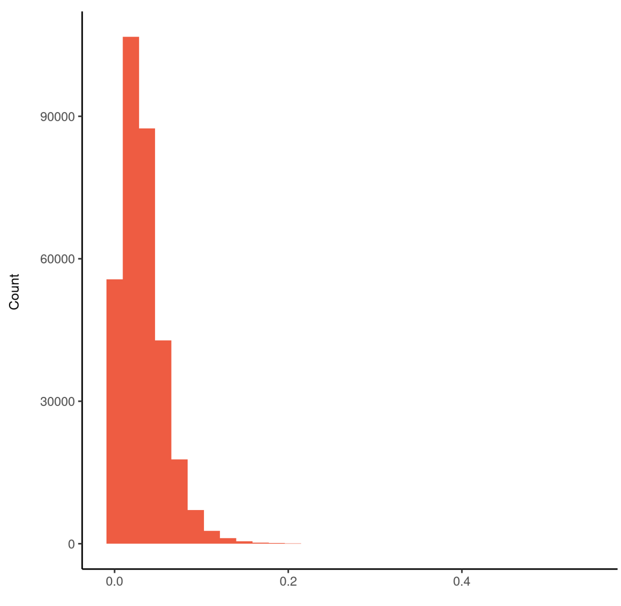
(e) Small Grinder

Figure A10. : Absolute Welfare Loss from Inattention



Note: The figures shows the change in consumer surplus at a given store in a particular time period in dollar terms.

Figure A11. : Relative Welfare Loss from Inattention



Note: The figures shows the change in consumer surplus at a given store in a particular time period relative to the share-weighted store average price.