Pierre Chandon & Brian Wansink

How Biased Household Inventory Estimates Distort Shopping and Storage Decisions

The authors develop a model of how consumers estimate the level of product inventory in their households. Two laboratory experiments and two field studies involving 29 product categories show that (1) consumers anchor their estimates on their average inventory and fail to adjust sufficiently; (2) adjustments follow an inelastic psychophysical power function, leading to overestimations of low levels of inventory and underestimations of high levels; and (3) adjustments are more elastic and, thus, more accurate when inventory is salient. Contrary to the assumptions of practitioners and academic modelers, these inventory *estimates*, not actual inventory levels, drive subsequent purchase incidence. Simulation results further show that biased estimates increase overstocking and spoilage among stockout-averse consumers but increase stockouts and unmet demand among overstocking-averse consumers. By predicting the magnitude, not just the direction, of estimation biases, the model and the results offer new insights into accelerating the consumption of healthy foods and improving the targeting of stockpiling-inducing sales promotions.

"I buy lots of things and then go back to the house and see the fridge is full of all the stuff I've just bought."
—Prince William, Heir of the British Throne, (*BBC News* 2003).

Biased inventory estimates can lead to overstocking and stockouts in a household. Overstocking contributes to households losing an estimated 14% of their meat, grain, fruit, and vegetable purchases (Jones et al. 2003). Household-level stockouts contribute to unmet demand, which represents a wasted opportunity for consumers and for retailers and manufacturers. For example, the success of the "Got milk?" advertising campaign is primarily attributed to its reduction of the frequency of milk stockouts in consumers' refrigerators (Manning 1999). If biased inventory estimations lead to overstocking and stockouts, it is an important issue for the welfare of budget-pressured and time-pressured consumers. It is also an important issue for managers and researchers because these biases can influence storage, consumption, and repurchase decisions and therefore affect promotional elasticity (Meyer and Assunçao 1990). Furthermore, if consumers are biased in their estimations, this would imply that the interpretation and policy recommendations of purchase quantity and timing models may be biased because these models typically assume that consumers know how much of the product is in their pantries (Ailawadi and Neslin 1998).

A recent surge of interest in estimation biases underscores the importance of understanding consumers' inventory estimation biases. Researchers have shown that reference levels, stimulus size, and salience influence consumers' estimations of product volume in a container (Krider, Raghubir, and Krishna 2001; Raghubir and Greenleaf 2006; Raghubir and Krishna 1999), numerosity (Krishna and Raghubir 1997; Pelham, Sumarta, and Myaskovsky 1994), and estimations of purchase and consumption frequency (Lee, Hu, and Toh 2000; Menon, Raghubir, and Schwarz 1995). Still, however, no research has examined the following important questions regarding consumers' inventory estimations that are of interest to consumers, marketers, and researchers: (1) How do people estimate how much product they have in inventory? (2) Are their estimations accurate, or are they systematically biased by the reference inventory, inventory size, and inventory salience? (3) How do biased inventory estimates distort storage and purchase decisions? and (4) What is the relationship between estimation biases and key category characteristics?

To answer these questions, we build on psychophysics research on magnitude estimation and develop a model of how consumers estimate inventory. This model incorporates reference, size, and salience effects. Using the model, we show in a simulation that biased inventory estimates increase overstocking and spoilage among stockout-averse consumers but increase stockouts and unmet demand

Pierre Chandon is Assistant Professor of Marketing, INSEAD, (e-mail: pierre.chandon@insead.edu). Brian Wansink is John S. Dyson Chair of Applied Economics and Management, Professor of Marketing, and Professor of Nutritional Science, Department of Applied Economics and Management, Cornell University (e-mail: Wansink@Cornell.edu). The authors thank Gilles Laurent; Collin Payne; Jack Soll; Luc Wathieu; Alex Chernev; Priya Raghubir; Aradhna Krishna; Gita Johar; Sachin Gupta; Gérard Cachon; and those who participated when the authors presented this research at INSEAD; at the University of Illinois, Urbana–Champaign; at the Marketing Science Conference, College Park, Md.; and at the Association for Consumer Research Conference, Toronto.

To read or contribute to reader and author dialogue on this article, visit http://www.marketingpower.com/jmblog.

ISSN: 0022-2429 (print), 1547-7185 (electronic)

among overstocking-averse consumers. We directly test the model's predictions in two laboratory experiments, which include manipulations of internal and external reference levels, of the actual size of the inventory, and of the inventory's salience. Two field studies involving 29 product categories further demonstrate the robustness of the model predictions. They also show that estimated inventory levels predict repurchase decisions better than actual inventory levels do and that inventory estimation biases are related to three key category characteristics: the degree of impulse buying, the ease of stockpiling, and the average promotional elasticity. In the general discussion, we show how our model and the results provide a parsimonious explanation for other, previously unexplained estimation biases. We also offer new insights for accelerating the consumption of healthy foods and improving the targeting of sales promotions

A Model of Consumers' Inventory Estimations

In this section, we develop a model of how consumers estimate the quantity of a product they have in inventory. The key features of the model are that (1) consumers anchor their estimations on internal or external reference levels and insufficiently adjust for the actual inventory level, (2) adjustments are inelastic (they increase more slowly than the true deviation from the reference level, and thus their quality worsens as the deviation increases), and (3) adjustments are more elastic (and, thus, more accurate) when inventory is perceptually salient. We first show how each prediction can be derived from psychophysical research on magnitude estimations and then show how these predictions can be integrated in a simple power model.

Reference Effects

Inventory estimations involve either judgments of numerosity (e.g., "How many eggs do I have?") or judgments of volume (e.g.,"How many ounces of coffee do I have?"). Many studies have shown that consumers anchor numerosity and volume estimations on salient internal or external reference levels and fail to adjust sufficiently for deviations from the reference level. For example, Krishna and Raghubir (1997) show that consumers' estimation of the number of dots in a line is higher when the dots are in multiple clusters (high-reference anchor condition) than when they are all together in one uninterrupted line (low-reference anchor condition). Raghubir and Krishna (1999) show that volume estimations are anchored on the elongation of the container (see also Wansink and Van Ittersum 2003). Krider, Raghubir, and Krishna (2001) show that consumers anchor area estimations on the most salient dimension, in which salience is context dependent. For example, they show that the orientation of a square influences whether the diagonal or a side is used as an anchor when estimating its surface (see also Pelham, Sumarta, and Myaskovsky 1994).

In the context of inventory estimations, we expect that the default anchor is the average category inventory level for each consumer. This is a reasonable assumption because in the absence of other information on actual inventory, the average inventory level is the best estimator of actual inventory if inventory follows a normal or uniform distribution. However, consistent with Krider, Raghubir, and Krishna's (2001) results, we expect that consumers will use external anchors if they are made contextually salient (e.g., by asking consumers to judge explicitly whether an inventory level is above or below some number). In summary, we expect that consumers anchor their inventory estimations on their average inventory, except when external reference levels are made salient (in which case these reference levels serve as anchors).

Size Effects

Recent research on anchoring effects has shown that when people have internally selected a reference as an anchor, they insufficiently adjust for the difference between the reference and the actual value of the magnitude to be estimated (Epley and Gilovich 2001). Epley and Gilovich (2004) show that people estimate the number of days it takes Mars or Neptune to orbit the sun by using the number of days it takes the Earth as an anchor (365 days). As a result, they adjust more for Neptune (mean estimated answer is 3447 days) than for Mars (mean estimated answer is 574 days) because they know that Neptune is further away from the sun than Mars, but they still fall short of the truth (which is 60,225 days for Neptune and 869 days for Mars).

Our model contributes to the literature on anchoring and adjustment by further predicting that the size of the adjustment is inelastic to the actual deviation from the anchor because estimations follow a compressive power function of actual magnitude (i.e., $EST = a \times (ACT)^b$, where EST is the estimated magnitude, ACT is the actual magnitude, and $b < 1$). In other words, the percentage change in estimations is lower than the percentage change in actual size (Stevens 1986). As a result, adjustments become less effective as the deviation between the reference level and the actual inventory level increases.

There is considerable evidence that magnitude estimations follow a compressive power function of actual magnitudes. For example, Teghtsoonian (1965) finds that the exponent of the power function is approximately .7 when estimating the size of three-dimensional objects. Frayman and Dawson (1981) examine exponents of power functions for different shapes (cubes, spheres, octahedrons, cylinders, tetrahedrons) and find that they are all approximately .6. For perceived numerosity judgments, Krueger (1982) finds a power exponent between .80 and .82. Overall, there is strong support in the literature for our prediction not only that consumers fail to adjust sufficiently for the deviation from the reference level but also that such adjustments are inelastic and therefore increase at a lower rate than the true difference between the reference level and the actual inventory level.

Salience Effects

The power exponent, which measures the elasticity of estimations to actual changes in the magnitude of the stimuli, is influenced by the perceptual salience (e.g., visual prominence) of the different dimensions of the stimulus. Krider, Raghubir, and Krishna (2001) find that the power exponent of area estimations for two-dimensional objects is greater when the salience of secondary dimensions (those that are not used as anchors) is increased. For example, because people anchor area estimations of circles on the length of their horizontal diameters, area estimations and willingness to pay for round pizzas are more sensitive to the actual size of the circle when the vertical diameter is made salient.

Building on these findings, we predict that the elasticity of adjustment improves with the perceptual salience of the product in inventory (and, thus, that the power exponent is greater when inventory is salient than when it is not). Following Krider, Raghubir, and Krishna (2001), we define salience as the ability of the product's inventory to attract attention. For example, the salience of a product's inventory increases if the product is stored in a visible place or if it is purchased or consumed frequently, because these factors increase attention to the actual inventory level. When inventory is salient, consumers are more likely to know whether the reference level they anchor on is wrong and should be adjusted. In the extreme case, consumers may have encoded salient inventory so well that they do not need to rely on the reference level at all. When inventory is not salient, however, consumers may not even know whether actual inventory is above or below the reference level and therefore will rely mostly on the reference anchor. Thus, we expect that inventory estimations are adjusted slightly when inventory is not salient and are adjusted more strongly when inventory is salient. In other words, we expect that inventory estimations are more sensitive to actual inventory and, thus, more accurate when inventory is salient than when it is not.

Modeling Biases in Inventory Estimations and Their Effects on Storage and Shopping Decisions

A Model of Household Inventory Estimation

We develop a simple mathematical model of consumers' inventory estimations that allows us to test the effects of inventory reference, size, and salience simultaneously. This model applies to situations in which consumers cannot simply look and count their inventory and, therefore, must estimate it from memory (e.g., when they are in a store). In essence, we estimate a series of power models in different conditions (e.g., when salience is low versus high) and use the model parameters to test for the hypothesized effects. The general form of these power models is as follows:

$$\text{EST} = \text{a} \times (\text{ACT})^{\text{b}}, \tag{1}$$

where EST is estimated inventory, ACT is actual inventory, and a and b are parameters estimated with regression. We use power models because they are consistent with psychophysical research on magnitude estimation and because they have four desirable properties.

The first desirable property is that the model intercept captures systematic differences between estimated and actual inventory, regardless of inventory level. Changes in the intercepts shift the power curve up or down but leave the shape of the curve unchanged. Therefore, we can test our prediction that internal or external anchors shift estimations toward the reference level by examining the value of the intercept a across anchor conditions.

Second, the power exponent b measures the elasticity of estimations and influences the shape of the power curve. If $b = 1$, estimations increase at the same rate as actual inventory. If $b < 1$, the power function is compressive, and estimations are inelastic (i.e., the percentage change in estimations is lower than the percentage change in actual inventory, or in other words, estimations increase at a slower rate than actual inventory). For example, if $b = .5$ and actual inventory increases by 50% (i.e., by a factor of 1.5), estimations increase by a factor of $(1.5)^{.5} = 1.225$ (i.e., 22.5%). If $b > 1$, the power function is expansive, and estimations are elastic (they increase at a faster rate than actual inventory). Therefore, we can test our prediction that inventory estimations are inelastic by verifying that $b < 1$.

Third, we can use the power model to compute the crossover inventory level at which estimations are accurate ($\text{ACT}^* = a^{1/(1-b)}$). Inventory levels below ACT* tend to be overestimated, and inventory levels above ACT* tend to be underestimated. We can then compare ACT* with the hypothesized reference level (e.g., a person's average perceived inventory).

Fourth, the power function treats positive and negative deviations from the crossover level similarly (i.e., the power exponent is the same, regardless of inventory level; there is no asymmetry between "gains" and "losses"). When the relevant logarithmic measures of accuracy are used (e.g., the log ratio of estimated divided by actual inventory), the magnitude of the bias is a linear function of actual inventory (measured in logs), and its slope is the same above and below the crossover level. Note, however, that the symmetry in the power model leads to asymmetry when more conventional measures of accuracy are used, such as absolute error (AE; |estimated – actual|) or percentage absolute error (PAE; |estimated – actual|/actual). A positive deviation from the crossover inventory level always leads to a smaller PAE (but to a larger AE) than a similar-size negative deviation from the crossover inventory level. For example, if $\text{EST} = 3 \times \text{ACT}^{.5}$, $\text{ACT}^* = 9$ units. When inventory is above the crossover level by 6 units (ACT = 15 units), the AE and PAE are 3.38 units and 23%, respectively. When inventory is below the crossover level by 6 units (ACT = 3 units), the AE and PAE are 2.2 units and 73%, respectively.

To test the hypothesized reference and salience effects, we estimate a different power model for each reference and salience condition. We illustrate how salience effects can be incorporated, but the approach is the same when we examine the effects of different inventory reference levels:

$$\text{EST} = \text{a} \times (\text{ACT})^{\text{b}}, \text{ when salience is low, and} \tag{2}$$

$$\text{EST} = \text{a}' \times (\text{ACT})^{\text{b}'}, \text{ when salience is high.}$$

To facilitate the statistical estimation of these power models, we linearized them as follows:[1]

[1]The linearization prevents the model from being estimated when either estimated magnitude or actual magnitude is zero. To overcome this limitation, it is possible to estimate the following nonlinear model with the Levenberg–Marquardt least square algo-

(3) $\ln(\text{EST}) = \ln(a) + b \times \ln(\text{ACT})$, when salience is low, and

$\ln(\text{EST}) = \ln(a') + b' \times \ln(\text{ACT})$, when salience is high.

To test whether a is statistically different from a′ and whether b is statistically different from b′, we estimate the following moderated regression:

$$\text{(4)} \qquad \ln(\text{EST}) = \alpha + \beta \times \text{SAL} + \delta \times \ln(\text{ACT}) + \gamma \times \ln(\text{ACT}) \times \text{SAL} + \varepsilon,$$

where SAL (salience) is a binary variable coded as –½ when salience is low and as ½ when salience is high; ε is the error term; and α, β, δ, and γ are estimated with ordinary least squares (OLS). These parameters are interpreted as follows: α measures the average intercept across the two salience conditions ($e^{\alpha} = [a \times a']^{½}$ is the geometric mean of the two intercepts); β measures the main effect of salience on the intercepts ($e^{\beta} = a'/a$); δ measures the average power exponent across the two salience conditions ($\delta = [b + b']/2$); and γ measures the effects of salience on the power exponent, that is, the interaction between salience and actual inventory ($\gamma = b' - b$). If salience improves the accuracy of estimations by reducing the degree of compression, we would expect γ to be positive and statistically different from 0.

Simulating the Effects of Biased Inventory Estimations on Storage and Shopping Decisions

Because no one expects that consumers have perfect inventory knowledge, the important issue is to determine the extent to which these biases distort consumers' storage and purchase decisions. To address this issue, we simulate the weekly category purchase decisions of consumers for a perishable product with stochastic household-level demand and examine the conditions under which biased inventory estimations lead to overstocking and, thus, to spoilage or stockouts and, ultimately, to unmet demand.

Consider a shopper who must decide whether to buy a dozen eggs in each weekly shopping trip but does not know what the exact demand will be over the next days because the three household members have an independent 50% chance of wanting an egg for breakfast. On any given day, demand can be 0 (p = 1/8), 1 (p = 3/8), 2 (p = 3/8), or 3 (p = 1/8) eggs. For the sake of simplicity, we consider two different types of households. The first segment consists of overstocking-averse households that discard their eggs quickly (e.g., after 9 days because they consume their eggs soft boiled and are afraid of bacteria). The optimal purchasing rule for overstocking-averse households is to buy a dozen eggs if their inventory on the day of their weekly shopping trip falls below 6 eggs. The second segment consists of stockout-averse households that keep their eggs for a relatively long time (e.g., 14 days, because they are consuming them hard-boiled). The optimal purchasing rule for stockout-averse households is to buy a dozen eggs if their inventory is below 12 eggs.

Our model of inventory estimation allows us to simulate how biased inventory estimates influence the magnitude of spoilage and unmet demand for households that are averse either to stockouts (with a discard threshold of 14 days and a purchasing threshold of 12 eggs) or to overstocking (discard threshold = 9 days, purchasing threshold = 6 eggs). Building on the field study results for eggs (see Field Study 2), we model inventory estimations for eggs as a compressive power function with parameters a = b = 1 (i.e., EST = ACT) when estimations are unbiased or with parameters a = 3 and b = .5 (i.e., $\text{EST} = 3 \times [\text{ACT}]^{.5}$) when estimations are biased.

For stockout-averse consumers, because of the underestimation of high inventory levels, when their estimate is 12 eggs, the actual inventory is 16 eggs (because $3 \times [16]^{.5} = 12$). As a result, stockout-averse consumers continue to repurchase a dozen egg even if they have as many as 16 eggs in inventory, whereas the optimal purchase threshold for them is 12 eggs. Figure 1 illustrates the effects of estimation biases for these households in terms of spoilage and unmet demand over the first eight weeks of the year. Over the course of one year, Table 1 shows that estimation biases cause stockout-averse consumers to purchase more (+7%, or 36 eggs), to increase the amount of spoilage strongly (+225%, or 27 eggs), and to reduce unmet demand only slightly (–33%, or 2 eggs).

For overstocking-averse consumers, because of the overestimation of low inventory levels, when their estimate is 6 eggs, the actual inventory is 4 eggs (because $3 \times [4]^{.5} = 6$). As a result, overstocking-averse consumers forgo repurchasing a dozen eggs even when they have as few as 4 eggs in inventory, whereas the optimal purchase threshold for them is 6 eggs. Figure 2 illustrates the effects of estimation biases for these households in terms of spoilage and unmet demand over the first eight weeks of the year. Over the course of one year, Table 1 shows that estimation biases cause overstocking-averse consumers to buy less (–9%, or 48 eggs), to increase unmet demand strongly (+118%, or 52 eggs), and to reduce spoilage slightly (–23%, or 12 eggs).

Overall, the simulation shows that biased inventory estimations can significantly increase both spoilage and unmet demand, depending on whether consumers are averse to stockouts or to overstocking. The underestimation of high inventory levels causes stockout-averse households to repurchase when their inventory is too high, thus leading to a significant increase in spoilage. Conversely, the overestimation of low inventory levels prevents overstocking-averse consumers from repurchasing when their inventory is too low, thus causing a significant increase in unmet demand. In addition, Table 1 shows that biased inventory estimates increase the variance of inventory and increase shopping and storage inefficiencies (the sum of units spoiled because of overstocking and those not consumed because of stockouts) for all households. In the "General Discussion" section, we show how these effects provide marketers with new opportunities for increasing consumption and for improving the targeting of sales promotions.

rithm: $\text{EST} = \alpha \times (\beta^{\text{SAL}}) \times (\text{ACT})^{(\delta \times (\gamma)^{\text{SAL}})}$, where $\alpha = (a \times a')^{½}$, $\beta = a'/a$, $\delta = (b \times b')^{½}$, and $\gamma = b'/b$. The interpretations of the α, β, δ, and γ parameters is the same as in the linearized model. In all analyses, both methods provided similar estimates, and therefore we use the simpler linearized model, except when incorporating zeroes is desirable because of a low number of observations.

FIGURE 1
How Biased Inventory Estimates Influence Daily Inventory, Spoilage, and Unmet Demand for Stockout-Averse Households (Purchase Threshold = 12 Units)

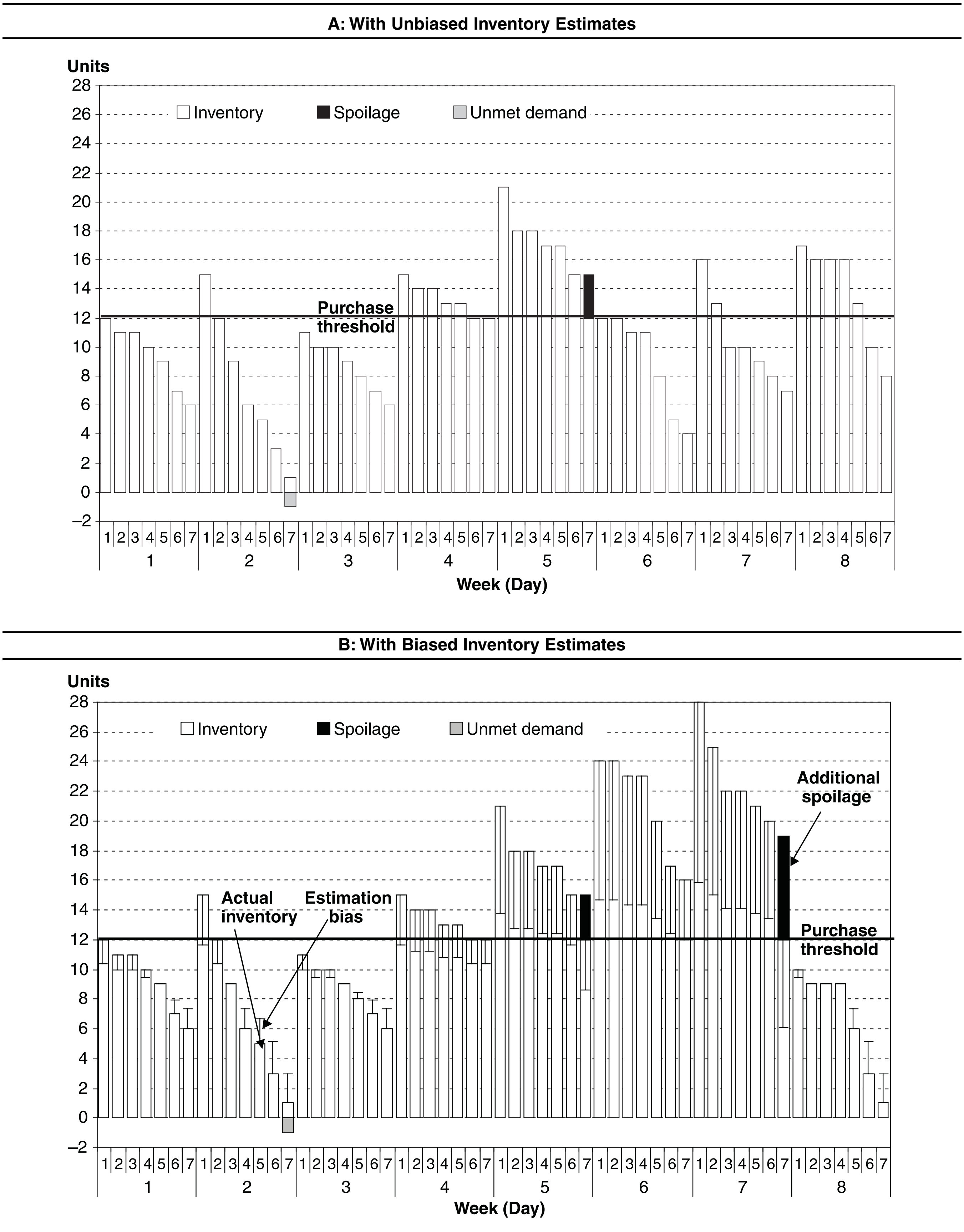


TABLE 1
Simulation Results: How Biased Inventory Estimates Influence Purchase Quantity, Spoilage, Unmet Demand, and Daily Inventory

	Stockout-Averse Households (Purchase Threshold = 12 Eggs)			Overstocking-Averse Households (Purchase Threshold = 6 Eggs)		
Inventory Estimations	Unbiased (Actual PT = 12 Eggs)	Biased (Actual PT = 16 Eggs)	Difference	Unbiased (Actual PT = 6 Eggs)	Biased (Actual PT = 4 Eggs)	Difference
Demand	539	539	0	539	539	0
Purchase quantity	552	588	+7%	552	504	–9%
Spoilage	12	39	+225%	53	41	–23%
Unmet demand	6	4	–33%	44	96	+118%
Spoilage plus unmet demand	18	43	+139%	97	137	+41%
Daily inventory (M)	11.1	14.0	+27%	7.9	6.6	–16%
Daily inventory (SD)	4.8	5.4	+13%	4.1	4.3	+4%

Notes: Spoilage is the annual number of eggs discarded. The discard threshold is 14 days for stockout-averse consumers and 9 days for overstocking-averse consumers. Unmet demand is the number of eggs that were not consumed because of stockouts. The simulation assumes that consumers purchase a dozen egg if their inventory estimate on the day of their weekly shopping trip is below the purchase threshold (PT). Because of the underestimation of high inventory levels, the actual purchase threshold for biased stockout-averse households is too high (16 versus 12 eggs). Because of the overestimation of low inventory levels, the actual purchase threshold for biased overstocking-averse households is too low (4 versus 6 eggs).

Experiment 1: How Inventory Size and External Anchors Influence Inventory Estimations

Procedure

The objective of Experiment 1 is to test how external anchors and inventory size influence inventory estimations (we test the effects of internal anchors and of inventory salience in Experiment 2). To achieve this objective, Experiment 1 used a mixed design with one within-subject factor with four levels (one, three, seven, or nine units in inventory), one between-subject factor with three levels (no external anchor, low external anchor, or high external anchor), and two replications (two different products) for each level of inventory.

Participants first examined a color picture of a pantry containing 8 target products and 13 other products in different quantities. The pantry contained one unit of 2 target products, three units of 2 other target products, seven units of 2 other target products, and nine units of 2 other target products. Participants were later asked to recall the inventory level of the 8 target products (Coca-Cola cans, Lifesavers candy, Smucker's jam, Campbell's soup, Charmin toilet tissue, Crest toothpaste, Carr's crackers, and Heinz tomato sauce). We rotated the products across the four inventory-level conditions following a Latin-square design. We selected the inventory levels of the target products on the basis of a prestudy of 21 consumers, who indicated that this range (one to nine) would straddle the average inventory of these products for almost all respondents (the typical average inventory for these products is three or four units).

The participants were 216 undergraduate students who were awarded extra credit participation points for a course they were attending. We told them that the aim of the study was to measure their liking of different types of teas. Consistent with this, we first asked them to evaluate the three brands of tea that were present in the pantry. To direct their attention toward the other products in the pantry (including the eight target products), we asked the participants to estimate the overall quality of the brands displayed and to indicate whether some of these products should be stored in a refrigerator rather than a pantry. We then asked each participant to return the first booklet containing the pantry picture. After a brief distracter task, we gave participants a second booklet containing a typical anchoring manipulation. We asked participants in the high-external-anchor condition (low-external-anchor condition) to indicate whether the number of units of each of the eight target products was above or below nine (one). We then asked them to estimate the number of units of the target products. Participants in the control condition estimated the number of units of the target products. Finally, we asked all the participants to write down how they had estimated the inventory levels for these products.

Results

We consider first the estimations made in the control (no external anchor) condition. Figure 3 shows that the mean estimated inventory was well below reality when there were seven or nine units in inventory, close to reality when there were three products in inventory, and slightly above reality when there was only one product in inventory. To test size effects in the control condition formally, we estimated the following linear regression with OLS:

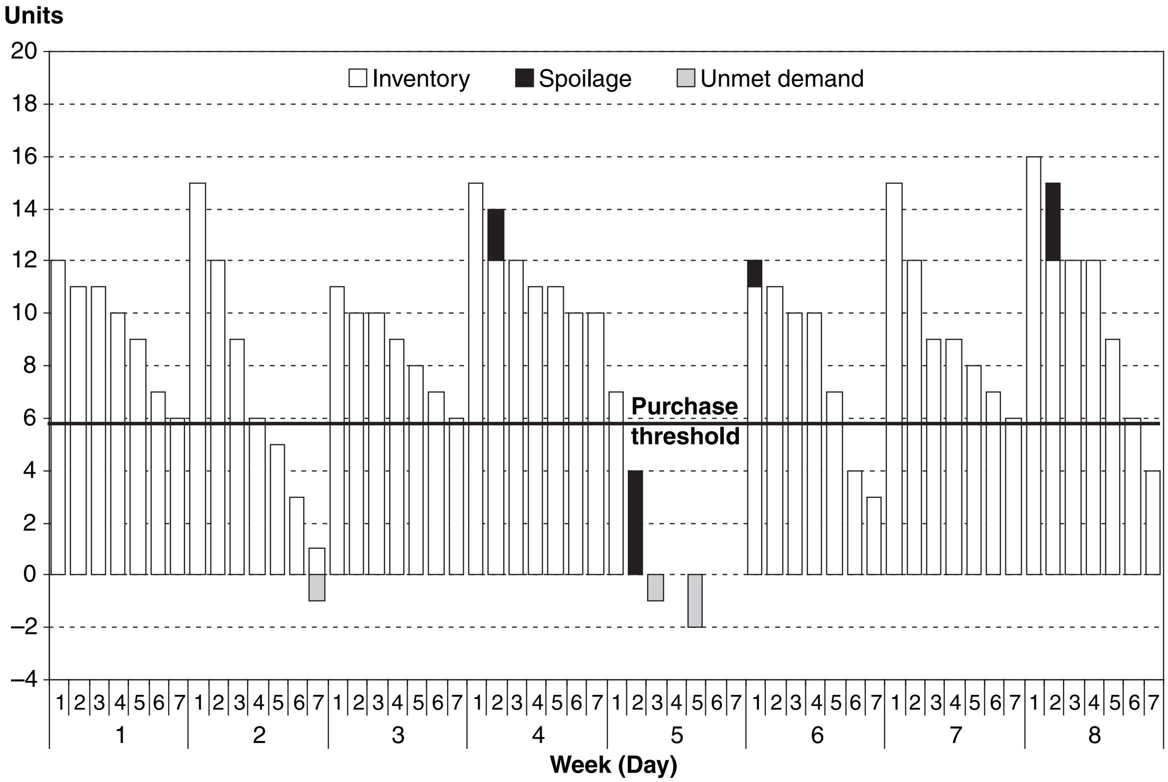
$$\ln(EST_{ij}) = \alpha + \beta \times \ln(ACT_{ij}) + \Sigma\sigma_j \times CAT_j + \varepsilon, \quad (5)$$

where EST_{ij} is the estimated inventory of product j by participant i, ACT_{ij} is the actual inventory, CAT_j are seven binary variables accounting for product differences, and ε is the error term.

As we expected, the power exponent was statistically below 1 (α = b = .43, t-test of difference from 1 = –9.1, $p <$

FIGURE 2
How Biased Inventory Estimates Influence Daily Inventory, Spoilage, and Unmet Demand for Overstocking-Averse Households (Purchase Threshold = 6 Units)

A: With Unbiased Inventory Estimates



B: With Biased Inventory Estimates

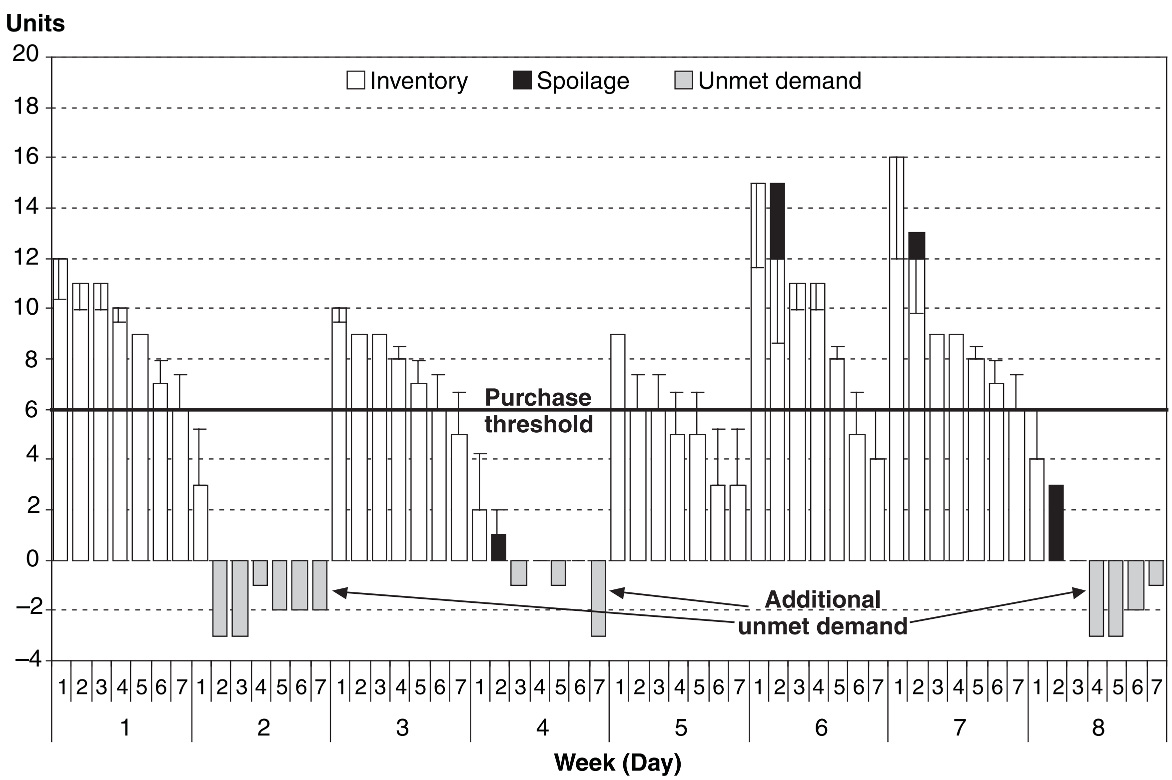
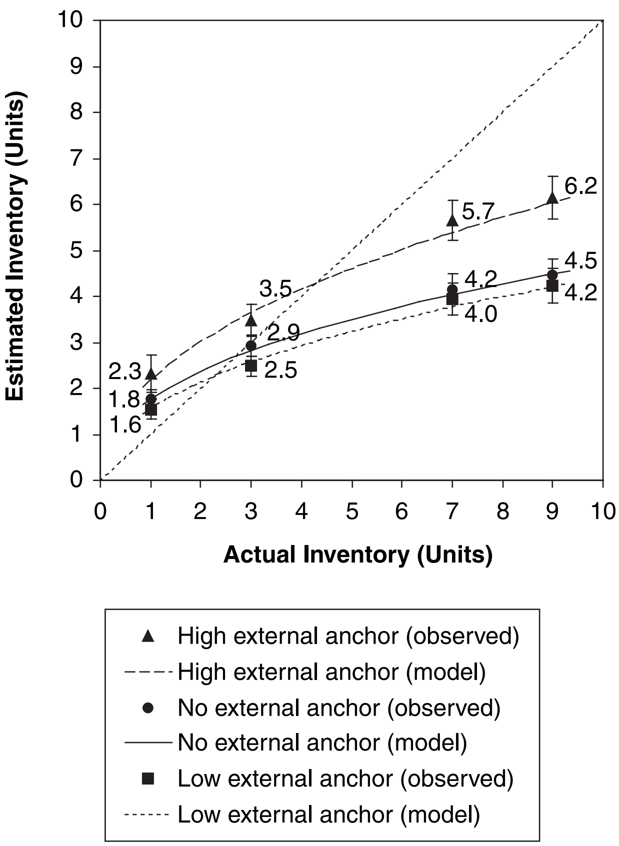


FIGURE 3
Experiment 1: Effects of External Anchors on Inventory Estimations (Geometric Means, Confidence Intervals, and Model Predictions)



.001), and the intercept was statistically above 0 ($\alpha = \ln(a) = .58$, t = 11.8, $p < .001$), which indicates that the intercept of the power model (a = 1.79) is statistically larger than 1. Five of the seven product-specific intercepts were statistically significant. In addition, we compared the fit of the power model with that of an alternative linear model ($EST_{ij} = \alpha + \beta \times ACT_{ij} + \Sigma\delta_j \times CAT_j + \varepsilon$). As we expected, the R-square was stronger for the power model ($R^2 = .36$) than for the linear model ($R^2 = .33$), and the mean average percentage error (MAPE) was lower for the power model (MAPE = .58) than for the linear model (MAPE = .74; paired t-test = 21.3, $p < .001$). In summary, as we predicted, consumer inventory estimations in the control condition follow a compressive power function of actual inventory.

We now turn to the analysis of the anchoring manipulation. As a manipulation check, we asked two coders, both of whom were unaware of the objective of the experiment, to classify consumers' retrospective protocols into three categories. The "internal anchor" category included protocols that referred to how much of the product a person usually has in inventory. The "visual memory" category included protocols that recalled the picture of the pantry. The "other" category included the remaining protocols, such as "I just guessed." Consistent with the anchoring literature, which shows that people are unaware of the effects of external anchors (Mussweiler, Strack, and Pfeiffer 2000), no protocol mentioned the anchoring manipulation. An analysis of the 166 usable protocols shows that the frequency of protocols mentioning internal anchors was higher in the control condition (M = 22.1%) than in the two external anchor conditions (M = 7.1%; $\chi^2 = 7.8$, $p < .01$). This suggests that, as we predicted, consumers were more likely to use internal anchors when external anchors were not contextually salient.

Figure 3 shows that the anchoring manipulation shifted inventory estimations toward the anchor but did not influence the relationship between estimated and actual inventory (the power curves are parallel). To test anchoring effects formally, we estimated the following regression:

$$\begin{aligned}(6)\quad \ln(EST_{ij}) &= \alpha + \beta \times \ln(ACT_{ij}) \\ &+ \delta \times EXTANCH1_i + \gamma \times EXTANCH9_i \\ &+ \lambda \times EXTANCH1_i \times \ln(ACT_{ij}) \\ &+ \theta \times EXTANCH9_i \times \ln(ACT_{ij}) + \Sigma\sigma_j \times CAT_j + \varepsilon_{ij},\end{aligned}$$

where EST_{ij} is the estimated inventory for product j by participant i, ACT_{ij} is the (geometric) mean-centered actual inventory for product j and participant i,[2] $EXTANCH1_i$ is a binary variable taking the value of ⅔ if participant i was in the low-external-anchor condition (anchor = 1) and –⅓ otherwise, $EXTANCH9_i$ is a binary variable taking the value of ⅔ if participant i was in the high-external-anchor condition (anchor = 9) and –⅓ otherwise, and CAT_j are seven binary variables accounting for product differences (j = [1, …, 7]). The simple effects of both external anchors were statistically significant and in the expected direction (for the anchor on one unit, $\delta = -.07$, t = –2.0, $p < .05$; for the anchor on nine units, $\gamma = .27$, t = 7.8, $p < .01$), indicating that inventory estimations were assimilated toward the anchors. Consistent with the model, the interaction parameters were not statistically significant ($\lambda = .03$, t = .6, $p = .52$, and $\theta = .2$, t = .4, $p = .68$), indicating that the degree of compression of inventory estimations was the same across the three conditions. Finally, the intercept was statistically different from 0 ($\alpha = 1.21$, t = 78, $p < .01$). This indicates that participants significantly overestimated their inventory when it was only one unit (in the control condition). Four of these product-specific intercepts were statistically significant.

[2]In practice, we divide actual inventory by its mean across inventory conditions (five units). In a moderated regression, mean centering the variables involved in an interaction allows us to estimate the main effects of the variables involved in an interaction when the other variables involved in the interaction are at their mean. For example, it allows us to estimate the effects of anchoring on inventory estimations when inventory level is at its average level (five units). We use geometric mean centering rather than arithmetic mean centering to be consistent with the power functional form and to prevent negative values whose logs cannot be computed.

Discussion

Experiment 1 shows that inventory estimations are assimilated toward external anchors and are adjusted for the actual inventory level through a nonlinear compressive power function. Experiment 1 also shows that the rate of adjustment for the actual size of the inventory remains constant, regardless of which reference levels serve as anchors. In other words, external anchors do not influence the exponent of the power function and thus do not influence the quality of adjustments from the reference level. Finally, the protocols provide indirect evidence that consumers rely on internal anchors when external anchors are not salient. However, because we did not measure the value of these internal anchors, we cannot test the effects of internal anchors or determine whether an average home inventory can serve as one. Experiment 2 further tests the model by examining the effects of internal anchors and by directly manipulating product salience.

Experiment 2: How Internal Anchors, Inventory Salience, and Inventory Size Influence Inventory Estimations

Procedure

Experiment 2 used the same procedure and stimuli as Experiment 1 but with three important differences. First, we did not manipulate external anchors but asked each participant to indicate the average inventory of the eight target products in their own house (the hypothesized internal anchor). Second, we manipulated the perceptual salience of the target products in three combined ways. Salient products were located on the top or middle shelf of the pantry (as opposed to the bottom shelf), were separate from other products (rather than being crowded together with them), and were given multiple facings when available in more than one unit (rather than being stacked together in an overlapping fashion). We assigned each of the eight target products to one of the eight conditions created by the two (high or low salience) × four (one, three, seven, or seven units) design. As in Experiment 1, we rotated the products across the eight inventory size and salience conditions following a Latin-square design. Third, as a means to check the effectiveness of the salience manipulation, we asked participants to rate how visible each product was in the picture.

The participants were 150 undergraduate students who were awarded extra credit participation points for a course they were attending. Manipulation checks show that the salience manipulation was successful. An analysis of variance indicated that products in the high-salience condition were rated as “more visible” (M = 6.75 on a nine-point scale, anchored by 1 = “completely disagree” and 9 = “completely agree”) than those in the low-salience condition (M = 5.89; $F(1, 1090) = 24.6$, $p < .001$). However, neither the inventory size nor the salience manipulation influenced the home inventory reported for each product ($F(3, 1090) = .58$, $p = .63$, and $F(1, 1090) = .59$, $p = .44$, respectively). This shows that the average home inventory level was not estimated from the inventory level or the salience of the product in the study.

Results

Figure 4, Panel A, shows the mean estimated inventory as a function of the inventory and salience manipulations. As we expected, increasing salience made estimations less compressive and more accurate. To examine the effects of internal anchors, we assigned participants to a high- or low-internal-anchor group on the basis of their self-reported average home inventory level for each product. Across the eight products, the average home inventory in the high-internal-anchor group was M = 8.5 units versus M = .9 units in the low-internal-anchor group. Figure 4, Panel B, shows mean estimated inventory as a function of actual inventory for both internal anchor groups. As we expected, estimations in the high-internal-anchor condition were higher than those in the low-internal-anchor condition, regardless of inventory level, but the two estimation curves remained parallel, suggesting that internal anchors (unlike salience) did not influence the elasticity of adjustments.

To test our predictions that salience reduces the degree of compression of estimations and that estimations shift toward internal reference levels, we estimated the following regression:

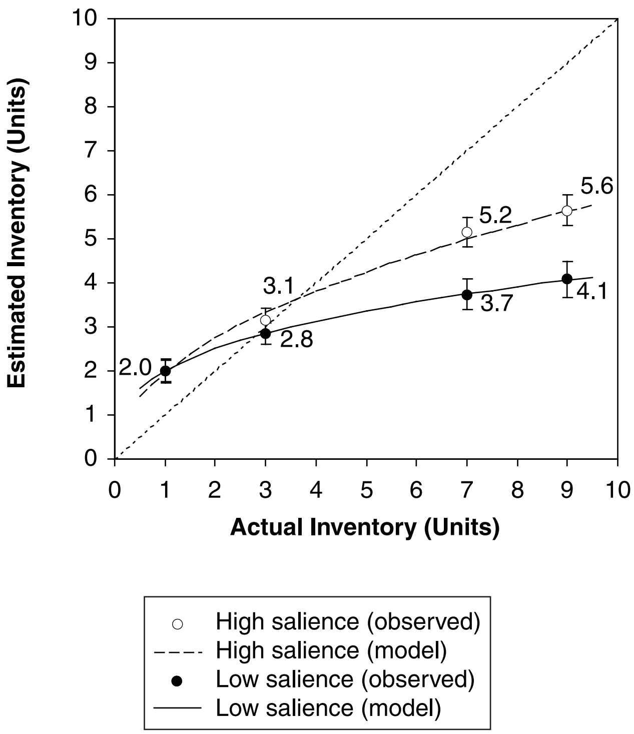
$$\text{(7)}\quad \ln(EST_{ij}) = \alpha + \beta \times \ln(ACT_{ij}) + \delta \times INTANCH_{ij} + \gamma \times SAL_{ij} + \lambda \times INTANCH_{ij} \times \ln(ACT_{ij}) + \theta \times SAL_{ij} \times \ln(ACT_{ij}) + \sum\sigma_j \times CAT_j + \varepsilon_{ij},$$

where EST_{ij} is the estimated inventory for product j by participant i, ACT_{ij} is the geometric mean-centered actual inventory for product j and participant i, $INTANCH_{ij}$ is a binary variable taking the value of ½ if the home inventory of participant i for product j is in the top 50% of the distribution for this product and –½ otherwise, SAL_{ij} is a binary variable taking the value of ½ if product j is in the high-salience condition for participant i and –½ otherwise, and CAT_j are seven binary variables accounting for product differences (j = [1, …, 7]).

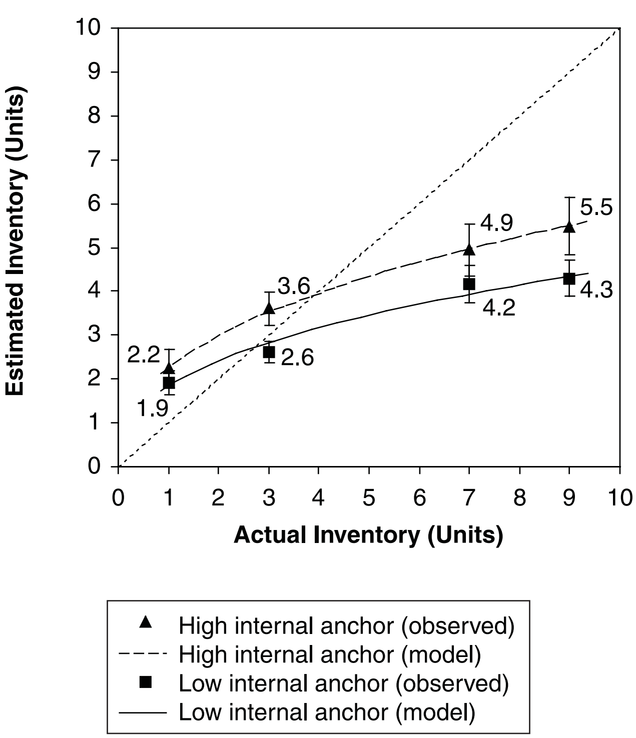
As in Experiment 1, the power exponent was statistically below 1 ($\beta = .41$, t-test of difference from 1 = –28.9, $p < .001$), indicating that the rate of adjustment increases more slowly than the actual inventory size. The coefficient capturing the simple effect of the internal anchor was positive and statistically significant ($\delta = .09$, $t = 2.3$, $p < .05$), but its interaction with the actual inventory level was not ($\lambda = .02$, $t = .5$, $p = .61$). This shows that inventory estimations were assimilated toward the average home inventory for that product but that this internal anchor did not change the rate at which estimations were adjusted for the actual inventory level. The main effect of the salience manipulation was positive and statistically significant ($\gamma = .22$, $t = 6.5$, $p < .01$), and its interaction with the actual inventory level was also positive and statistically significant ($\theta = .15$, $t = 3.7$, $p < .01$). Because of the significant interaction between the actual inventory level and salience, the effects of salience are not statistically significant when the inventory level is low (one or three units). This shows that adjustments are more sensitive when inventory is salient than

FIGURE 4
Experiment 2: Effects of Perceptual Salience and Internal Anchors on Inventory Estimations (Geometric Means, Confidence Intervals, and Model Predictions)

A: Perceptual Salience



B: Internal Anchors



when it is not. Finally, the intercept was statistically different from 0 (α = 1.23, t = 71.3, p < .01), and four of the seven product-specific intercepts were statistically significant, replicating the results of Experiment 1.

Discussion

Experiment 2 shows that internal anchors, like the external ones, shift estimations toward the reference level but do not change the rate at which estimations are adjusted for the actual inventory level. Experiment 2 also shows that the salience of the product in the pantry increases the rate of adjustment. Estimations of less salient products are almost entirely driven by the reference inventory (the power curve is almost flat). In contrast, estimations of salient products are significantly influenced by the actual inventory level (the slope of the curve is close to one).

Taken together, Experiments 1 and 2 provide strong evidence in support of the model of inventory estimations. Yet although judgment and estimation biases are often found in a laboratory setting, they can be less apparent, or even negated, in the field, in which there is less variation in inventory size, reference levels, and salience and consumers may have greater experience with the estimation task. In Field Study 1, we investigate the robustness of these effects by measuring actual and estimated inventory and product salience in six categories. We also study whether estimated inventory is a better predictor of category purchase incidence than actual inventory. In Field Study 2, we test the robustness of the model in 23 new categories, and we examine whether the degree of compression of a given category is associated with the degree of impulse buying, the ease of stockpiling, and the promotional elasticity of that category.

Field Study 1: Inventory Salience and Estimation Accuracy

Procedure

Over two periods of five days each, we intercepted 121 adult consumers in four different central Illinois supermarket parking lots as they were exiting the supermarket, and we asked them to estimate their current inventory of six product categories in exchange for \$9. After completing their estimates, participants were given a preaddressed, stamped envelope and a brief questionnaire that asked them to check their actual inventory levels for these categories when they returned home the same day. The same questionnaire also asked them to rate the visibility of each category in their homes by indicating their agreement with the following sentence: "These [category name] are stored in a very visible place." They rated their agreement on a nine-point scale, anchored by 1 = "strongly disagree" and 9 = "strongly agree." Finally, participants indicated whether they had purchased any products from these six categories during their latest shopping trip.

We chose the six categories (apples, canned tuna, hot dogs, potatoes, tea bags, and tomatoes) on the basis of a pretest involving 16 consumers, which showed that they estimated their inventory using discrete package units (as opposed to continuous measures, such as ounces) and that there was enough variance in the way different households stored them to expect salience effects (e.g., some consumers keep tomatoes in a salient container on the kitchen counter, whereas others store them in the refrigerator where they are less visible). To verify the quality of the measures of actual inventory levels, we visited 16 households the day after they sent us their pantry checks and measured the number of units for these six categories ourselves. When we excluded partial units and subsequently purchased units, pantry accuracy was well over 90% for this subsample of households. Of the 121 consumers intercepted, 90 (74.4%) returned their questionnaire within one week, and we included them in the analysis.

Results

Purchase incidence. Part of our motivation for studying biases in inventory estimations is the assumption that estimations, not actual inventory, drive important decisions, such as whether to repurchase from a given category during a supermarket shopping trip. Field Study 1 enables us to test this assumption by comparing the association between category purchase incidence on the one hand and estimated or actual inventory on the other hand.

We measured purchase incidence using a variable ($REFILL_{ij}$) that took the value of 1 if at least one purchase was made from category j by participant i during the supermarket shopping trip that occurred just before the estimation and 0 if otherwise. As we expected, the correlation between purchase incidence and estimated inventory was negative and statistically significant ($r[REFILL_{ij}, EST_{ij}] = -.10$, $p < .05$), whereas the correlation between purchase incidence and actual inventory was not statistically different from zero ($r[REFILL_{ij}, ACT_{ij}] = -.03$, $p = .52$). To test whether the predictive power of estimated and actual inventory is statistically different, we used a repeated measures analysis of variance with EST_{ij} and ACT_{ij} as the within-subjects measures and $REFILL_{ij}$ as the between-subjects factors. As we expected, the interaction between $REFILL_{ij}$ and the within-subjects factor (estimated versus actual inventory) was statistically significant ($F(1, 441) = 3.8$, $p < .05$).[3] This result implies that purchase incidence is more strongly associated with estimated inventory than with actual inventory.

Inventory estimation biases. We found that the mean estimated inventory was within 10% of the mean actual inventory level for the six categories. However, this aggregate accuracy is not a result of consumers being accurate. Rather, it is a result of underestimations compensating for overestimations. Across the six categories, only 49% of inventory estimations were accurate, whereas 28% were underestimations and 23% were overestimations. For example, Figure 5 shows that the mean inventory estimation of tea bags and tomatoes is slightly overestimated at low levels (the two lowest quartiles) but is strongly underestimated at high levels (the two highest quartiles). Figure 5 also shows that these estimates were less compressive and, thus, more accurate when tomatoes and tea bags were highly salient than when they were not.

To test the statistical significance of size and salience directly, we estimated the following regression:

$$\ln(EST_{ij}) = \alpha + \beta \times \ln(ACT_{ij}) + \delta \times SAL_{ij} + \gamma \times SAL_{ij} \times \ln(ACT_{ij}) + \sum\sigma_j \times CAT_j + \varepsilon_{ij}, \tag{8}$$

where EST_{ij} is the estimated inventory for category j by participant i, ACT_{ij} is the actual inventory for category j and participant i (it is not mean centered so as to be able to test the intercept when the actual inventory is equal to 1), SAL_{ij} is a mean-centered binary variable measuring the visibility of category j in the pantry of participant i (it is categorized using a median split, but the results are unchanged if we use the original continuous measure), and CAT_j are five binary variables accounting for product differences ($j = [1, \ldots, 5]$). As we expected, the intercept was statistically larger than zero ($\alpha = .39$, $t = 6.2$, $p < .01$), indicating that the intercept of the power model is larger than 1 ($a = e^{\alpha} = 1.48$). Consistent with the hypothesized size effects, the exponent was statistically lower than 1 ($b = \beta = .77$, t-test of difference from $1 = -7.5$, $p < .01$), indicating that inventory estimations are also compressive in the field. The main effect of salience was not statistically significant ($\delta = -.09$, $t = -.9$, $p = .36$), which means that as in Experiment 2, salience did not influence estimations when there was only one unit in inventory. Consistent with the hypothesized salience effects, the interaction between salience and actual inventory was positive and statistically significant ($\gamma = .08$, $t = 2.0$, $p < .05$). As in Experiment 2, inventory estimations were more compressive when the category was not particularly visible in the pantry, and they were more accurate when the category was visible in the pantry.

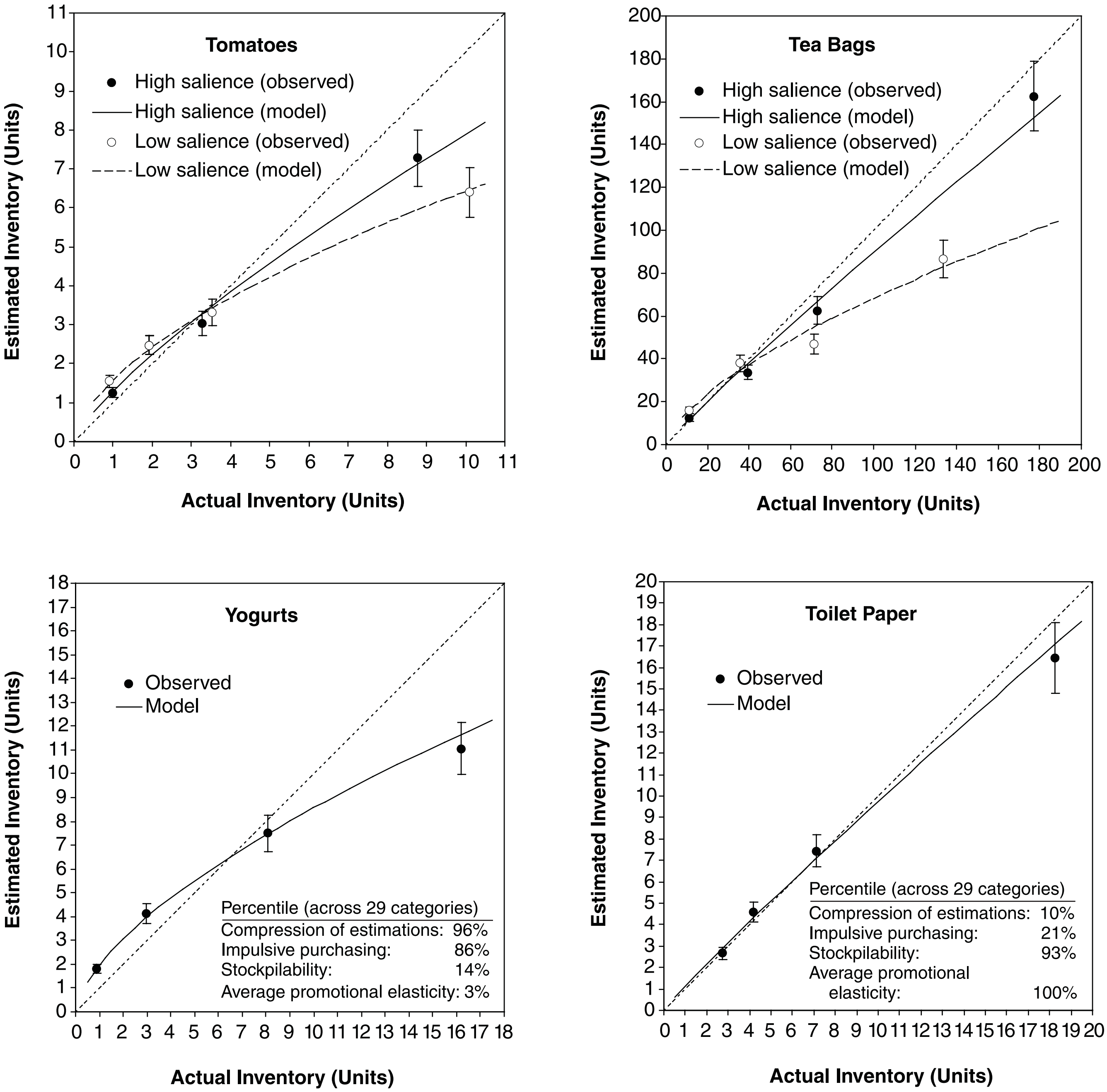
Discussion

Field Study 1 makes two contributions. First, it shows that estimated inventory is more strongly related to category purchase incidence than is actual inventory. Importantly, this result cannot be explained by mere-measurement effects, because inventory estimations were measured as consumers were exiting the store, *after* their purchases had been made. In addition, our decision to measure purchase incidence with the second questionnaire (completed at home) and not during the parking lot interview (when inventory estimations were collected) reduces the likelihood that these results are driven by self-presentation biases (i.e., that consumers reported inventory estimations consistent with their purchase decisions).

Second, Field Study 1 supports the experimental results by showing that inventory estimations follow a compressive function of actual inventory levels, even when they are made by adult consumers for products that they have just bought. In addition, it shows that consumers adjust for

[3]The main effect of the within-subjects factor was not statistically significant ($F(1, 441) = 1.6$, $p = .20$), indicating that mean estimated inventory is similar to mean actual inventory. The main effect of $REFILL_{ij}$ was also not statistically significant ($F(1, 441) = 2.0$, $p = .16$), which is not surprising, given that it measures the average effect of a nonsignificant association (between $REFILL_{ij}$ and ACT_{ij}) and of a significant but weak association (between $REFILL_{ij}$ and EST_{ij}).

FIGURE 5
Field Studies 1 and 2: Estimated and Actual Inventory for Selected Product Categories (Geometric Means, Confidence Intervals, and Model Predictions)



Notes: The degree of compression of inventory estimations of the product category is based on the estimated power exponent in the psychophysical equation EST = a × $(ACT)^b$, the degree of impulse purchasing and the stockpilability of the category are based on Narasimhan, Neslin, and Sen (1996), and the average promotional elasticity was measured by Information Resources Inc. (*P-O-P Times* 1991).

actual inventory levels more when the products are salient (stored in a visible place) than when they are less salient.

A limitation of Field Study 1 is that it did not measure the average product inventory for each household. Therefore, its findings should be interpreted as showing that respondents with lower actual inventory tend to overestimate their inventory, whereas respondents with higher inventory levels tend to underestimate them. Naturally, the average inventory across respondents may not be representative of the average inventory of either very light or very heavy buyers. However, note that this heterogeneity in average inventory level reduces the likelihood of detecting inventory biases in the population (because extreme inventory levels are actually average levels for these extreme consumers).

Field Study 1 raises three further questions: (1) Are these results generalizable to nonfood products? (2) Do average inventory levels serve as estimation anchors in the field? and (3) Is the degree of compression of inventory estimations related to managerially important category characteristics? We address the first question by conducting a large-scale field study in which we measure estimated and actual inventory for 23 new food and nonfood categories. We address the second question by incorporating data on average inventory level (rather than on current actual inventory level), which we collected in an additional survey. Finally, we address the third question by studying the association between the power exponent, which measures the degree of compression in each category, and three category characteristics, which we obtained from secondary data, namely, the degree of impulse buying in the product category, the ease of stockpiling the product, and the average promotional brand elasticity of the brands in the product category.

Field Study 2: How Estimation Biases Vary Across Products

Procedure

Field Study 2 used the same procedure as Field Study 1 to measure biases in inventory estimations. In Field Study 2, however, we measured inventory levels in ounces rather than in units for seven products (soft drinks, coffee, shampoo, mayonnaise, laundry detergent, dishwashing detergent, and ketchup) for which inventory levels are typically measured in ounces because of the large variations in package sizes. We measured inventory levels for the other 16 products (soap, canned soup, spaghetti, vacuum cleaner bags, yogurt, toothpaste, frozen meat, eggs, frozen vegetables, butter sticks, canned fruit, pasta sauce, cookies, toilet tissue, salad dressing, and breakfast cereals) in units. Another difference is that we did not measure inventory salience.

To avoid respondent fatigue, we surveyed participants on five to eight products. To verify the accuracy of the actual inventory measures, we asked a subgroup of consumers to telephone one of the researchers immediately after they had checked their actual inventory levels. We called the consumers who had not telephoned by 7:30 that evening and reminded them to check their inventory. There were no systematic differences among the results of participants who had telephoned, the results of participants who needed to be reminded, and the rest of the participants who were not contacted. To check accuracy further, we told an additional subgroup of consumers to keep their questionnaire because we would pick them up the next day. During the pickup round, we requested permission to inspect their actual inventory. With the households who agreed to the inspection (approximately 75%), there were no full-unit discrepancies with the self-reported inventory. Of the 461 consumers who participated in Field Study 2, 317 (68.7%) returned their questionnaire in a timely manner. Together with data we obtained in Field Study 1, we had a total of 2185 estimations on 29 products (an average of 75 observations per category).

To avoid common method biases, we obtained data on impulse buying, ease of stockpiling, and promotional response from two independent sources. For impulse buying and ease of stockpiling, we used the survey of 108 product categories that Narasimhan, Neslin, and Sen (1996) conducted and published in *Journal of Marketing*.[4] This survey measured impulse buying by asking 100 consumers to rate their agreement with the following two statements: "I often buy the product on a whim when I pass by it in the store," and "I typically like to buy this product when the urge strikes me." We measured ease of stockpiling by asking the same consumers to rate their agreement with the following two statements: "It is easy to store extra quantities of this product in my home," and "I like to stock up on this product when I can." For promotional elasticity, we used data from the "Infoscan Topical Marketing Report," generated by Information Resources Inc. and published in the *P-O-P Times* (1991). This report provides an estimate of the average percentage brand sales increase in response to a 15% price cut with an in-store display for 164 product categories based on the results of Information Resources Inc.'s PromotionScan model (Abraham and Lodish 1993) estimated on the checkout data of 2400 grocery stores. The category definitions of Narasimhan, Neslin, and Sen's (1996) survey and of the Infoscan Report matched ours in 25 of the 29 product categories. We used data from categories closest to the remaining four (e.g., frozen side dishes for frozen vegetables).

Results

Are these results generalizable to nonfood products? As in Field Study 1, we found that the mean estimated inventory is similar to the mean actual inventory. With four exceptions (crackers, ketchup, tomatoes, and potatoes), the mean estimations are all within 20% of the mean actual inventory for the 29 products (the 6 products surveyed in Field Study 1 and the 23 surveyed in Field Study 2). As in Field Study 1, this aggregate accuracy does not indicate that consumers are accurate at the individual level; instead, it indicates that overestimations compensate for underestimations. Across the 29 products, only 41% of inventory estimations are accurate (within one unit or one ounce of the actual inventory), whereas 31% are underestimations, and 25% are overestimations.

To test the model predictions, we estimated the power model shown in Equation 1 for each of the 29 categories. Because of the low number of observations for some categories, we used the nonlinear Levenberg–Marquardt least square algorithm, which allows us to incorporate observa-

[4]Scott Neslin, Albert Wesley Frey Professor of Marketing in the Tuck School of Business at Dartmouth College, generously provided these data.

tions with a zero-estimated or actual inventory level (we obtained similar results when estimating a linearized model on nonzero observations with OLS).

Table 2 shows that the power model fits the inventory estimation data well (the mean R^2 is 54%). All power exponents are below 1, and the variance from 1 is statistically significant at the 5% level for all 29 products. All power intercepts are above 1, and the difference is statistically significant at the 5% level for 22 of the 29 products and is statistically significant at the 10% level for 3 other products. In addition, the fit of the power model was superior to the fit of a linear model (EST = a + b × ACT) for almost all the products. Specifically, the R-square of the power model was higher than the R-square of the linear model for 23 of the 29 products, and the average fit improvement was 3% and was statistically significant. Similarly, the MAPE was lower for the power model than for the linear model for 28 of the 29 products and was statistically lower for the power model than for the linear model across all observations ($MAPE_{power} = .44$ versus $MAPE_{linear} = .61$; paired $t = 10.5$, $p < .001$). All these results support our prediction that inventory estimations for these 29 products follow a compressive power function of actual inventory level.

Using the estimated model parameters, we computed the crossover inventory level ($e^{\alpha/(1-\beta)}$) for each product. As we expected, the crossover inventory level was within the range of observed actual inventory levels for all 29 products. Excluding one outlier (tea bags, which has high inventory levels when measured in units), we found that low inventory levels tended to be overestimated but that as actual inventory reached four to six units or 24–42 ounces, estimations tended to become accurate. However, when actual inventory levels were above these average levels, they tended to be strongly underestimated.

Do average inventory levels serve as anchors in field estimations? We asked 37 adult consumers, similar to those involved in Field Studies 1 and 2, to estimate their *average* inventory level (as opposed to their current inventory level) for these 29 products. We then estimated the following constrained model:

$$EST_{ij}/AVG_j = \alpha \times (ACT_{ij}/AVG_j)^{\beta} + \varepsilon_{ij}, \tag{9}$$

TABLE 2
Field Studies 1 and 2: Category-Level Power Regression Results (Estimates, Standard Errors, Fit, and Predicted Crossover Inventory Level)

Category	Intercept		Exponent		r(EST, ACT)	N	ACT*
Apples	1.23**	(.13)	.83**	(.04)	.93	89	3.4
Butter sticks	2.03*	(.68)	.63**	(.14)	.63	51	6.8
Canned fruit	1.52**	(.25)	.71**	(.08)	.86	52	4.2
Canned soup	1.81**	(.22)	.69**	(.06)	.78	155	6.8
Canned tuna	1.61**	(.22)	.66**	(.07)	.78	98	4.1
Cereals	1.35**	(.16)	.77**	(.06)	.88	53	3.7
Coffee[a]	2.61**	(.92)	.73**	(.09)	.71	56	34.9
Cookies	1.46**	(.13)	.42**	(.06)	.75	53	2.0
Dishwashing detergent[a]	3.40*	(1.60)	.61**	(.11)	.60	55	23.1
Eggs	3.08**	(1.09)	.54**	(.14)	.61	33	11.5
Frozen meat	2.34**	(.71)	.53**	(.13)	.61	39	6.1
Frozen vegetables	1.74**	(.28)	.61**	(.08)	.80	52	4.1
Hotdogs	2.45**	(.72)	.57**	(.09)	.68	87	8.0
Ketchup[a]	3.08**	(.91)	.57**	(.08)	.69	55	13.7
Laundry detergent[a]	3.51*	(1.71)	.71**	(.09)	.67	54	75.9
Mayonnaise[a]	3.75**	(1.27)	.60**	(.09)	.73	56	27.2
Pasta sauce	1.14	(.12)	.84**	(.08)	.85	53	2.3
Potatoes	1.52	(.47)	.78**	(.10)	.74	88	6.7
Salad dressing	1.36**	(.23)	.66**	(.11)	.70	53	2.5
Shampoo[a]	4.87**	(2.13)	.55**	(.11)	.55	57	33.7
Soap	1.79**	(.21)	.68**	(.06)	.74	147	6.2
Soft drinks[a]	14.36	(11.64)	.41**	(.16)	.50	55	91.5
Spaghetti	1.43**	(.11)	.53**	(.05)	.69	159	2.1
Tea bags	2.07**	(.61)	.82**	(.05)	.83	85	56.9
Toilet paper	1.57**	(.24)	.81**	(.05)	.91	52	10.7
Tomatoes	1.35	(.29)	.69**	(.10)	.71	85	2.6
Toothpaste	1.36**	(.12)	.75**	(.06)	.89	42	3.4
Vacuum cleaning bags	1.78**	(.26)	.59**	(.06)	.66	140	4.1
Yogurt	3.29**	(.60)	.42**	(.07)	.66	131	7.8

*Statistically different from 1 at the 10% level (one-tailed).
**Statistically different from 1 at the 5% level (one-tailed).
[a]Inventory measured in ounces.
Notes: EST = estimated inventory, ACT = actual inventory, N = number of estimations, and ACT* = crossover inventory (mean estimated inventory is equal to mean actual inventory).

where EST_{ij} is the estimated inventory for product j by participant i at the time of the study, ACT_{ij} is the actual inventory, and AVG_j is the average inventory for product j measured in the additional survey. This model is a simple rewriting of the basic power model in Equation 1, in which estimated and actual inventory are expressed as a proportion of the average inventory of the product rather than as the original units. This transformation influences the intercept but leaves the power exponent and the model fit unchanged from the values shown in Table 2.

If people use the average inventory level as an anchor, inventory estimations are unbiased when actual inventory is at its average level, and therefore the mean estimated inventory is equal to the mean actual inventory level when actual inventory is at its average level. In Equation 9, this occurs only if the intercept (α) is equal to 1 (because if $ACT_{ij} = AVG_j$, $EST_{ij} = AVG_j = ACT_{ij}$). As expected, we found that the intercept of this new model was not statistically different from 1 in 25 of the 29 categories. We obtained similar results when using the mean actual inventory level, which we computed across participants in Field Studies 1 and 2 as the category average inventory rather than the mean obtained from the new survey. Overall, these results are consistent with the hypothesis that average inventory levels serve as internal anchors in the absence of salient external anchors.

Is the degree of compression of inventory estimations related to managerially important category characteristics? As we show in Table 2 and Figure 5, some products, such as yogurt, exhibit strong compression and therefore are somewhat inelastic to actual changes in inventory. For example, given that the power exponent of yogurt is .42, if its inventory increases by 50%, estimations increase by only 19% (see Figure 5). Figure 5 also shows that other products, such as toilet paper, exhibit little compression and therefore are relatively accurate at all levels of inventory. Because the power exponent of toilet paper is .81, if its inventory increases by 50%, estimations increase by almost the same percentage (39%). We expect that these category differences in the rate of adjustment to actual inventory are linked to three key category characteristics: (1) the likelihood of impulse buying, (2) the ability to stockpile, and (3) the average brand promotional elasticity in that category.

The actual inventory of products that are likely to be bought on impulse is apt to fluctuate more and in a less predictable way than the actual inventory of products whose purchases are planned. As a result, we expect that estimations will be more compressive and, therefore, less accurate for products with a high degree of impulse purchasing. In contrast, the actual inventory of products that are easy to stockpile should be easier to monitor than the inventory of products that are difficult to stockpile. We also expect that estimations will be less compressive for products that are easy to stockpile. Finally, consumers are more likely to switch to another brand or to stockpile in response to a promotional display when they have an accurate understanding of their inventory than when they do not know how much of the product is left in their inventory and therefore hesitate to buy the promoted brand (Wansink, Kent, and Hoch 1998). Consumers who have no idea about how much of the product they have in inventory are more likely to pass a promotion and follow their habitual purchasing pattern for fear of overstocking. Consequently, we expect estimations to be more compressive for categories with a high average promotional elasticity.

We tested these hypotheses by estimating the following model with the nonlinear Levenberg–Marquardt least square algorithm:

$$EST_{ij} = \alpha \times \beta_1^{IMP_j} \times \beta_2^{STO_j} \times \beta_3^{PROM_j} \times (ACT_{ij})^{\delta} \times \gamma_1^{IMP_j} \times \gamma_2^{STO_j} \times \gamma_3^{PROM_j}, \quad (10)$$

where EST_{ij} is the estimated inventory for category j by participant i, ACT_{ij} is the actual inventory for category j and participant i, IMP_j is a mean-centered binary variable measuring the degree of impulse buying in the category, STO_j is a mean-centered binary variable measuring the ease of stockpiling the product category, and $PROM_j$ is a mean-centered binary variable measuring the average promotional elasticity in the category (we dichotomized the three binary variables with a median split).

As we expected, the parameter capturing the interaction between ACT and IMP was statistically below 1 ($\gamma_1 = .93$, asymptotic 95% confidence interval [CI] = [.85, .99]), indicating a lower power exponent (i.e., less accuracy) for categories bought on impulse. Conversely, the parameter capturing the interaction between ACT and STO was statistically above 1 ($\gamma_2 = 1.15$, 95% CI = [1.04, 1.26]), indicating a higher power exponent (i.e., more accuracy) for categories that are easy to stockpile. Finally, the parameter capturing the interaction between ACT and PROM was above 1 but was statistically significant only at the 10% level ($\gamma_3 = 1.07$, 90% CI = [1.00, 1.15]), indicating a slightly higher power exponent (i.e., more accuracy) for categories with high average promotional elasticity. In summary, these analyses supported our hypotheses that inventory estimates are particularly biased for product categories that are bought on impulse, are difficult to stockpile, and have a low average promotional elasticity.

General Discussion

Consumers' biases in inventory estimations can distort storage and shopping decisions and can lead to overstocking and spoilage or to stockouts and unment demand. To understand the origin of these biases, we developed a model of how consumers estimate the amount of product they have in inventory. This model predicts that (1) unless an external reference level is salient, consumers anchor their estimations on their average inventory level and insufficiently adjust for the actual inventory level; (2) the adjustment from the average inventory follows an inelastic power function; and (3) the adjustment is more elastic and, thus, more accurate when inventory is perceptually salient. Drawing on this model, we simulated the effects of biased inventory estimates on the amount of spoilage and unmet demand for consumers who are averse to stockouts or to overstocking. We then tested the predictions of the model for 8 products in two laboratory experiments and for 29 products in two field studies.

The key results of these studies are as follows: First, contrary to the assumptions of practitioners and modelers, consumers' estimates of their inventory are better predictors of category purchase than their actual inventory. Second, although individual inventory estimations are seldom accurate, the mean estimated inventory of a product category is a valid estimation of its mean actual inventory because underestimations tend to compensate overestimations. Third, in the absence of salient external reference levels, consumers anchor their inventory estimations on their average inventory level and adjust insufficiently for the actual size of the inventory. Fourth, adjustments from the reference level follow an inelastic psychophysical function (a power function with an exponent below 1), and as a result, the quality of the adjustments deteriorates as the inventory deviates from the reference level. In short, below-average inventory levels are slightly underestimated, average inventory levels are accurately estimated, and above-average inventory levels are strongly underestimated. Fifth, inventory estimations are more elastic (more sensitive to actual changes in inventory) and, thus, more accurate when inventory is salient than when it is not. Sixth, the underestimation of high inventory levels causes stockout-averse households to repurchase when their inventory is too high, thus leading to a significant increase in spoilage. Conversely, the overestimation of low inventory levels prevents overstocking-averse consumers from repurchasing when their inventory is too low, thus causing a significant increase in unmet demand. Finally, the least elastic and, thus, least accurate estimations are those of product categories that are bought on impulse, are difficult to stockpile, and have a low promotional elasticity.

Implications for Further Research

By identifying three possible sources of biases, our model contributes to the literature on estimations, which has focused primarily on documenting biases rather than on explaining why these biases occur and what factors moderate them. For example, despite the considerable evidence that people underestimate the duration of past and future events when their duration is greater than five minutes and overestimate them when they last less than five minutes, there is no consensus about why this occurs (Lee, Hu, and Toh 2000; Roy, Christenfeld, and McKenzie 2005). Still, this pattern of results, even in this seemingly unrelated context, can be readily explained by two features of our model: (1) People anchor estimations on average duration, and (2) adjustments for deviations from the average are inelastic.

For researchers interested in magnitude estimations, our basic finding that deviations from the reference level follow a compressive power function has important implications. First, it enables them to predict the magnitude, not just the direction, of these biases. Second, it enables them to predict the average estimation bias for any duration, not just for those already measured. Third, it can lead to new testable predictions, such as the magnitude of estimation biases will become larger as the duration increases, even when the magnitude of the bias is measured as a percentage deviation from reality (for a proof of this, see Chandon and Wansink, in press). Fourth, this model can enable researchers to make predictions about the boundary conditions of duration estimation biases. For example, it would predict that duration estimations are more accurate when the event is salient and when average duration is used as an anchor than when the event is not salient or when external reference levels are used as anchors.

Our finding that inventory estimations follow a compressive power function of actual inventory is consistent with psychophysics research and with a great deal of accumulated evidence on magnitude estimation studies but is the opposite of what signal detection theory would predict. A key feature of signal detection theory is its assumption that people account for the relative costs of over- and underestimations (Green and Swets 1988). For inventory estimations, overestimations are more costly when inventory is low and when stockouts are likely to occur. In contrast, underestimations are more costly when inventory is high and when overstocking is likely. As a result, signal detection theory would predict that consumers underestimate low inventory levels (to avoid costly overestimations) and overestimate high inventory levels (to avoid costly underestimations). One explanation for our opposite findings is that in our studies, the cost–benefit payoff of the estimations was constant across inventory levels. Further research could try to reconcile the psychophysics and signal detection predictions by manipulating the costs of over- and underestimations (e.g., by measuring or manipulating the opportunity costs of having too much or too little). Another explanation is that wishful thinking leads consumers to bias their estimations optimistically in the direction of what they hope they will be rather than in the direction that will help them avoid overstocking and stockouts. Collecting process data, especially over time, may help in teasing out these explanations and in further advancing the understanding of the origins of inventory estimation biases.

Finally, our findings challenge the interpretation of the results of purchase quantity and timing models, which assume that consumers have accurate, or at least unbiased, knowledge of how much of a given product they have in inventory (Ailawadi and Neslin 1998; Bell, Chiang, and Padmanabhan 1999; Sun 2005). These models are estimated on scanner panel data, which contain no information on inventory. Therefore, they estimate inventory from a person's purchase quantity and timing data. Recall our finding that estimated inventory predicts repurchase decisions better than actual inventory. This indicates that the inventory estimated by these models may be what consumers estimate the inventory to be and not what it really is. Thus, these models measure the effects on purchases of consumers' inventory estimations rather than the effects of their actual inventory. Our finding that inventory estimations are inelastic implies that these models overestimate the effects of actual inventory on purchase decisions. Larger changes in actual inventory (and, thus, deeper price cuts) may be necessary to achieve the effect sizes reported in these studies. Biases in inventory decisions could easily be incorporated into purchase timing and quantity models. One solution would be to incorporate a power model of inventory estimation within these models. Another would be to convert the actual inventory levels predicted by these models directly into inventory estimates using the relevant parameters estimated in the field studies and shown in Table 2.

Implications for Managers and Consumers

Our model and the results offer new insights into accelerating the consumption of healthful foods. An important finding of prior research on inventory was the "stock pressure effect"; that is, inventory stockpiling accelerated the consumption of foods (Chandon and Wansink 2002). Unfortunately, it was convenient snack foods that received the greatest boost in consumption frequency when stockpiled. Given the rising trend in obesity, responsible managers of healthful foods could use our new insights into how consumers estimate inventory to influence their consumption of these foods. One way, as was done with the "Got milk?" campaign, would be to raise inventory salience. This would reduce the likelihood of wasting healthful foods, which are often more perishable than less healthful ones. It would also accelerate the consumption rate of health foods. Studies of food-intake diaries have shown that the intake frequency of fruits and yogurts increases as their perceived expiration date approaches (Wansink 2006). Such efforts can also lead people to think of a recent instance in which they consumed the food, which in turns increases consumption intentions (Wansink and Deshpandé 1994). Because inventory estimations are anchored on a person's own typical or normal level of inventory, another way of increasing a person's purchase and consumption of a healthful food would be to increase his or her perception of what should be a typical or normal level of inventory of that food. For example, knowing that the average family of four has 11.2 cans of soup in inventory could lead consumers to alter their inventory (and consumption) norm for the relatively healthful option of canned soup (Wansink and Chandon 2006).

Our model and the results also offer new insights into the targeting of sales promotions to consumer and category segments. Consider the two segments of consumers shown in Table 1 with identical demand: overstocking-averse consumers (who only repurchase when their inventory is low) and stockout-averse consumers (who repurchase at higher levels of inventory). The tendency to overestimate low inventory levels makes overstocking-averse consumers too conservative and thus limits their annual purchases to 504 units in the simulation, which is below their annual demand of 539 units. In contrast, the tendency to underestimate high inventory levels makes stockout-averse consumers too liberal and thus increases their annual purchases to 588 units, which is above their annual demand of 539 units. Which of these two segments should be targeted with sales promotions? To answer this question, we simulated the impact of a price reduction that would lower the purchase threshold by two units for both segments. We found that this would increase purchases by 12% among overstocking-averse consumers but by only 6% among stockout-averse consumers. The actual increase is likely to be even greater because these figures do not account for the consumption acceleration that higher inventory levels create. Because these overstocking-averse consumers probably include those with small households or those with small kitchens, multiunit promotions ("buy 5 get 1 free") and packaging would seem particularly appropriate because they facilitate flexible storage and reduce spoilage. Although identifying who the overstocking-averse consumers are would require specific data collection, the same reasoning would apply to category differences, suggesting benefits of targeting sales promotions to categories in which most consumers are averse to overstocking rather than to categories in which most consumers are averse to stockouts.

Finally, the simulation results show that retailers and manufacturers would increase their sales if they helped overstocking-averse consumers improve the accuracy of their estimations. This is because it would prevent these consumers from overestimating their low inventory levels and from mistakenly postponing their purchases. In contrast, improving the accuracy of stockout-averse consumers would reduce their purchases because it would prevent them from making the mistake of repurchasing when they still have a lot of product left in inventory.

What could be done to improve the accuracy of inventory estimations? The lack of self-knowledge about estimation strategies revealed by the protocol data and the robustness of the biases exhibited in the field studies suggest that consumers do not learn much from experience, even though they often run out of stock and waste overstocked products. (This, in itself, raises the issue as to whether consumers may be softening the negative consequences of estimation errors by adapting their consumption.)

Our model and our findings suggest that inventory estimations could be improved by helping consumers recognize when their estimations are likely to be particularly inaccurate, such as when their inventory is lower or higher than usual. In these cases, our findings suggest that consumers should increase the extremity of their estimations to compensate for the inelasticity of their intuitive estimates. Marketers could also improve the accuracy of inventory estimations by raising the perceptual salience of inventory levels. This could be done by changing where and how a product is stored, by usage-related advertising, or by transparent package designs that facilitate the monitoring of inventory.

REFERENCES

Abraham, Magid M. and Leonard M. Lodish (1993), "An Implemented System for Improving Promotion Productivity Using Store Scanner Data," *Marketing Science*, 12 (3), 248–69.

Ailawadi, Kusum and Scott A. Neslin (1998), "The Effect of Promotion on Consumption: Buying More and Consuming It Faster," *Journal of Marketing Research*, 35 (August), 390–98.

BBC News (2003), "William's Praise for St. Andrews," (May 30).

Bell, David R., Jeongwen Chiang, and V. Padmanabhan (1999), "The Decomposition of Promotional Response: An Empirical Generalization," *Marketing Science*, 18 (4), 504–526.

Chandon, Pierre and Brian Wansink (2002), "When Are Stockpiled Products Consumed Faster? A Convenience–Salience Framework of Postpurchase Consumption Incidence and Quantity," *Journal of Marketing Research*, 39 (August), 321–35.

——— and ——— (in press), "Is Obesity Caused by Calorie Underestimation? A Psychophysical Model of Fast-Food Meal Size Estimation," *Journal of Marketing Research*, forthcoming

Epley, Nicholas and Thomas Gilovich (2001), "Putting Adjustment Back in the Anchoring and Adjustment Heuristic," *Psychological Science*, 12 (5), 391–96.

——— and ——— (2004), "Are Adjustments Insufficient?" *Personality and Social Psychology Bulletin*, 30 (4), 447–60.

Frayman, Bruce J. and William E. Dawson (1981), "The Effect of Object Shape and Mode of Presentation on Judgments of Apparent Volume," *Perception and Psychophysics*, 29 (1), 56–62.

Green, David M. and John A. Swets (1988), *Signal Detection Theory and Psychophysics*. New York: Peninsula.

Jones, Timothy, Andrew Bockhorst, Brian McKee, and Aida Ndiaye (2003), "Percentage of Food Loss in the Household," report to the United States Department of Agriculture, Economics Research Service from the Bureau of Applied Research in Anthropology, University of Arizona.

Krider, Robert E., Priya Raghubir, and Aradhna Krishna (2001), "Pizzas: Pi or Square? Psychophysical Biases in Area Comparisons," *Marketing Science*, 20 (4), 405–425.

Krishna, Aradhna and Priya Raghubir (1997), "The Effect of Line Configuration on Perceived Numerosity of Dotted Lines," *Memory & Cognition*, 25 (4), 492–507.

Krueger, Lester E. (1982), "Single Judgments of Numerosity," *Perception & Psychophysics*, 31 (2), 175–82.

Lee, Eunkyu, Michael Y. Hu, and Rex S. Toh (2000), "Are Consumer Survey Results Distorted? Systematic Impact of Behavioral Frequency and Duration on Survey Response Errors," *Journal of Marketing Research*, 37 (February), 125–33.

Manning, Jeff (1999), *Got Milk? The Book*. Westminster, MD: Random House.

Menon, Geeta, Priya Raghubir, and Norbert Schwarz (1995), "Behavioral Frequency Judgments: An Accessibility-Diagnosticity Framework," *Journal of Consumer Research*, 22 (2), 212–28.

Meyer, Robert J. and Joao Assunçao (1990), "The Optimality of Consumer Stockpiling Strategies," *Marketing Science*, 9 (1), 18–41.

Mussweiler, Thomas, Fritz Strack, and Tim Pfeiffer (2000), "Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility," *Personality and Social Psychology Bulletin*, 26 (9), 1142–50.

Narasimhan, Chakravarthi, Scott A. Neslin, and Subrata Sen (1996), "Promotional Elasticities and Category Characteristics," *Journal of Marketing*, 60 (April), 17–30.

Pelham, Brett W., Tin Tin Sumarta, and Laura Myaskovsky (1994), "The Easy Path from Many to Much: The Numerosity Heuristic," *Cognitive Psychology*, 26 (April), 103–133.

P-O-P Times (1991), "Latest IRI Data Confirm the Effectiveness of P-O-P," (July–August), 37–42.

Raghubir, Priya and Eric A. Greenleaf (2006), "Ratios in Proportion: What Should the Shape of the Package Be?" *Journal of Marketing*, 70 (April), 95–107.

——— and Aradhna Krishna (1999), "Vital Dimensions in Volume Perception: Can the Eye Fool the Stomach?" *Journal of Marketing Research*, 36 (August), 313–26.

Roy, Michael M., Nicholas J.S. Christenfeld, and Craig R.M. McKenzie (2005), "Underestimating the Duration of Future Events: Memory Incorrectly Used or Memory Bias?" *Psychological Bulletin*, 131 (5), 738–56.

Stevens, Stanley Smith (1986), *Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects*. Oxford: Transaction Books.

Sun, Baohong (2005), "The Promotion Effect on Endogenous Consumption," *Marketing Science*, 24 (3), 430–43.

Teghtsoonian, Martha (1965), "The Judgment of Size," *American Journal of Psychology*, 78 (3), 392–402.

Wansink, Brian (2006), *Mindless Eating: Why We Eat More Than We Think*. New York: Bantam-Dell.

——— and Pierre Chandon (2006), "Can 'Low Fat' Nutrition Labels Lead to Obesity?" *Journal of Marketing Research*, 43 (November), forthcoming.

——— and Rohit Deshpandé (1994), "Out of Sight, Out of Mind: Pantry Stockpiling and Brand-Usage Frequency," *Marketing Letters*, 5 (1), 91–100.

———, Robert J. Kent, and Stephen J. Hoch (1998), "An Anchoring and Adjustment Model of Purchase Quantity Decisions," *Journal of Marketing Research*, 35 (February), 71–81.

——— and Koert Van Ittersum (2003), "Bottoms Up! The Influence of Elongation on Pouring and Consumption Volume," *Journal of Consumer Research*, 30 (3), 455–63.