

Testing Behavioral Hypotheses
Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior

SAM K. HUI

ERIC T. BRADLOW

PETER S. FADER

February 2009

*Sam K. Hui is an Assistant Professor of Marketing at the Stern School of Business of New York University, Eric T. Bradlow is the K. P. Chao Professor, Professor of Marketing, Statistics, and Education, and Co-Director of the Wharton Interactive Media Initiative and Peter S. Fader is the Frances and Pei-Yuan Chia Professor, Professor of Marketing, and Co-Director of the Wharton Interactive Media Initiative; both at the Wharton School of the University of Pennsylvania. Corresponding author: Sam K. Hui. Email: khui@stern.nyu.edu. The authors are grateful for the data and assistance provided by TNS-Sorensen and, in particular, the feedback and encouragement from Herb Sorensen.

Abstract

We examine three sets of established behavioral hypotheses about consumers' in-store shopping behavior (the effect of perceived time pressure, licensing, and the social presence of other shoppers) using field data on shopping paths and linked purchases obtained from an actual grocery store. We incorporate these behavioral hypotheses within an individual-level probability model to examine their empirical support via shoppers' in-store visit, shop, and buy decisions. Our results provide field evidence for the following empirical regularities. First, as consumers spend more time in the store, they become more purposeful in their trip — they are less likely to spend time on exploration, and are more likely to shop and buy. Second, consistent with “licensing” behavior (Khan and Dhar 2006), after purchasing virtue categories, consumers are more likely to shop at locations that carry vice categories. Third, the social presence of other shoppers attracts consumers towards a zone in the store (Argo et al. 2005), but reduces consumers' tendency to shop in that zone (Harrell et al. 1980). Implications of this research for store layout decisions, due to an improved understanding of consumer in-store path behavior, are briefly discussed.

Studying consumers' in-store behavior is an important topic for academic researchers and industry practitioners alike. Researchers are particularly interested in better understanding the factors that drive the dynamics of a consumer's shopping trip. For instance, how does a consumer's in-store behavior evolve (i) as she spends more time in the store, (ii) as she buys certain types of products, and (iii) as she reacts to the presence of other shoppers around her? The answers to these questions may lead to important managerial implications regarding the design of retail space and product placement, issues that are of key interest to practitioners.

In this paper, we study three situational factors that behavioral researchers have found to influence consumer's in-store decision making.¹ The first factor is *time pressure*. Dhar and Nowlis (1999) study how choice-deferral decisions (i.e., selecting a "no choice" option) are influenced by time pressure. Suri and Monroe (2003) extend this framework and find that even perceived time pressure can influence consumer behavior. The second factor is the composition of the *shopping basket*. Khan and Dhar (2006) find "licensing" effects in consumer choice, where the purchase of "virtue" categories improves a consumer's self-concept, which in turn increases the likelihood of a "vice" purchase by providing the consumer a "license" to do so. The third factor is the *presence of other shoppers*. Argo et al. (2005) investigate how the "mere presence" of other shoppers can affect consumers; Harrell et al. (1980) find that perceived crowding reduces shopping and purchase intentions.

With the notable exception of Argo et al. (2005), who also conduct field tests of their hypotheses, the aforementioned studies were mainly conducted in laboratory settings. This article enhances the external validity of these behavioral theories by providing a field test using data from an actual supermarket. We develop our hypotheses by integrating the above three

¹ We acknowledge that, beyond the three factors study here, previous research has also identified additional factors that influence in-store shopping, e.g., store knowledge (Park et al. 1989) and ceiling height (Meyers-Levy and Zhu 2007), among many others.

separate streams of research (time pressure, licensing, social influence of other shoppers), and assess their empirical support using an individual-level probability model. We control for unobserved heterogeneity using dynamic latent variables (Park and Bradlow 2005) within a hierarchical Bayesian framework (Rossi et al. 2006). We then estimate our model using PathTracker[®] data (Hui et al. 2008b; Sorensen 2003), which record (using Radio Frequency Identification) each shopper's path throughout a store and link it to traditional point-of-sale scanner data for the items purchased. Thus, through our model, we are able to examine whether these behavioral hypotheses are supported by field data.

Using the aforementioned model-based approach, we contribute to the prior literature on the three situational factors (time pressure, licensing, social influence of other shoppers) by looking at how each behavioral hypothesis differentially, if at all, affects each aspect (visit, shop, buy) of consumers' in-store decisions. This allows us to provide a richer behavioral description of the in-store shopping process. For instance, the social presence of other shoppers may attract a consumer to visit a zone; and once she gets there, she may become more or less likely to shop and buy products. In the same vein, we also study the effect of time pressure and licensing on visit, shop, buy behavior, using a set of three hypotheses for each situational factor. To the best of our knowledge, this integrated approach has never been proposed in the previous literature.

In addition to this substantive contribution, this article also develops a new methodology to analyze PathTracker[®] data, which can be applied to other path-related data in general (Hui et al. 2008a). While the previous literature on in-store path data has focused on exploratory analyses using clustering techniques (Larson et al. 2005) and comparison to optimal search algorithms (Hui et al. 2008b), this article is the first to develop an integrated probability model

that allow one to fully describe all aspects (visit, shop, and buy) of a grocery shopping path; this integrative nature of our model allows us to embed and test different behavioral hypotheses.

We organize the remainder of this article as follows: The next section integrates the previous literature on shoppers' behavior, providing us a framework and a set of field-testable behavioral hypotheses. Then we develop a probability model of shopping behavior that takes into account all the aforementioned theories. We then describe the field data used to estimate our model and conclude with a discussion of our results and managerial implications based on the behavioral findings.

THEORY AND HYPOTHESES

In this section, we develop our hypotheses through a review of relevant behavioral literature that provides insight into consumers' in-store behavior. To derive the relevant hypotheses, we divide a grocery path into a series of three exhaustive, sequential, and inter-related decisions (visit, shop, and buy), then examine how the three types of situational factors (perceived time pressure, licensing, and social influence of other shoppers) influence each of these decisions. That is, we consider the possibility that the situational factors may influence someone's shopping path in the store, but not what they buy. Or, as another example, it may be that the situational factors increase browsing (low probability of being in a shopping state) but also increase buying when the consumer is in a shopping state. Our research allows us to decompose these effects into their separate components.

Overview of the shopper's decision process

We divide a grocery trip into a series of visit, shop, and buy decisions, each of which is driven by latent attractions of categories and zones (defined in detail in our model section). An overview of the shopper’s decision process is depicted in Figure 1. We recognize that this is a paramorphic representation of the consumer decision process, albeit one that addresses each *observable* step of a shopper moving through a store.

[Insert Figure 1 about here]

We divide each shopping path into a number of zone transitions, which we refer to as “steps.” A new step is initiated each time the shopper leaves one zone and goes to another, until she reaches checkout. At step t , we denote the zone that shopper i is located as x_{it} . At $t=1$, the shopper is located at the store entrance. From there, the shopper first makes a *visit* decision: she decides which zone she is going to visit next. If that zone is the checkout, the trip ends. Otherwise, she makes a *shop* decision: she decides whether she is in “shopping mode” at her current zone, or whether she is only “passing through” on her way to a different zone. We denote this shop decision (at step t) by an (latent) indicator variable H_{it} , which takes the value 1 if the consumer is in shopping mode, and 0 otherwise. We note that it is possible that the consumer is in shopping mode ($H_{it} = 1$), but decides not to buy anything.

Depending on whether she shops or not, the shopper may stay in the zone for a different duration; presumably, the shopper tends to stay longer if she is shopping than simply passing through. Let S_{it} denotes the number of RFID “blinks” (five-second intervals as recorded by the PathTracker[®] software) that shopper i stays at her current zone during step t .

Next, if she decides to shop, she needs to make a *buy* decision: she decides which product categories, if any, to purchase in that zone. We denote her category purchase-incidence decision by the vector \vec{B}_{it} , where $B_{ikt} = 1$ if shopper i buys from category k at step t , and 0 otherwise. If

shopper i does not shop at step t ($H_{it} = 0$), she is only walking through the zone on her way to other zones, thus she does not make any buy decisions ($B_{ikt} = 0$ for all k).

Finally, the latent category attractions are updated to take into account the behavior observed in the preceding zone(s). After attractions are updated, the shopper then decides which zone to visit next, and the decision process in Figure 1 begins again.

We now consider how the three situational (behavioral) factors affect each of the visit, shop, and buy decisions. In addition, we will also utilize our model to assess the extent to which consumers exhibit planning-ahead behavior during their in-store shopping trip.

Perceived time pressure

The first situational factor is perceived time pressure. Assuming a mental accounting perspective (Thaler 1999), a consumer may enter the store with a “shopping time budget” in mind. As she spends more time in the store, the time allotted to grocery shopping is depleted, and she may start to feel time pressure when making visit, shop, and buy decisions. This is in the same spirit as Suri and Monroe (2003), who explored the influence of perceived time pressure, defined as a perceived limitation of the time available to consider information or make decisions, on consumers’ judgments of prices and products.

We hypothesize that under perceived time pressure, consumers will adapt by changing their shopping strategies. With limited time available, a consumer’s trip becomes more purposeful: they may engage in less exploratory shopping (Harrell et al. 1980) and instead focus on visiting and shopping at zones which carry categories they plan to buy. Thus, we hypothesize the effect of perceived time pressure on visit and shop behavior as follows.

H1a: (*Time Pressure-Visit*) As a consumer spends more time in the store, she becomes less likely to explore the store. That is, the checkout becomes relatively more attractive over time.

H1b: (*Time Pressure-Shop*) As a consumer spends more time in the store, she becomes more likely to be in a shopping mode when in a particular zone.

When a consumer is shopping in a zone, she has to decide what products to buy, or to make a “no-choice” decision and not purchase anything there. Dhar and Nowlis (1999) study the effect of time pressure on choice deferral; they find that when time to make a decision is limited, consumers may simplify their decision strategy and become less likely to select a no-choice option. Consistent with the previous literature, we hypothesize that under perceived time pressure, consumers are more likely to buy products in a zone (given that they are shopping there).

H1c: (*Time Pressure-Buy*) As a consumer spends more time in the store, she becomes more likely to buy in a zone.

Licensing

The second situational factor we consider is the composition of the shopping basket that a consumer assembles during his/her trip. “Licensing” (Khan and Dhar 2006), in the in-store shopping setting, refers to the idea that purchasing “virtue” items (e.g., vegetables, organic food) boosts a consumers’ self-concept, thus reducing the negative self-attributions associated with the purchase of “vice” categories (e.g., beer, ice cream). Following the same logic, buying vice categories has the opposite effect: it reduces the consumer’s self-concept and increases the negative self-attribution associated with additional purchases from vice categories. Thus, within our model, we hypothesize that at any moment during the trip, the extent of the licensing effect is governed by the current virtue/vice *balance* of the shopping basket at that moment. We expect that if the current basket has a positive virtue/vice balance (i.e., contains more virtue categories

than vice categories), licensing effect should be present, and the consumer become more likely to visit, shop, and buy from zones that contains more vice categories. Our formal definition of how we determined which categories are vice/virtue is discussed in the data/empirical section of the paper.

Formally, we hypothesize that

H2a: (*Licensing-Visit*) If the current shopping basket contains more virtue categories than vice categories, a consumer is more likely to visit zones that carry more vice categories.

H2b: (*Licensing-Shop*) If the current shopping basket contains more virtue categories than vice categories, a consumer is more likely to be in shopping mode at zones that carry more vice categories.

H2c: (*Licensing-Buy*) If the current shopping basket contains more virtue categories than vice categories, a consumer is more likely to buy at zones that carry more vice categories.

Social influence of other shoppers

The third situational factor is the social impact derived from other shoppers' presence in the store. To quantify the strength of social influence, social impact theory (Latane 1981) suggests that the extent of social impact should increase as a function of the size of the social presence (i.e., the number of other shoppers in the zone) and proximity (i.e., the size of the zone). Thus, we operationalize the strength of social impact by the *density* (number of shoppers per unit area) of other shoppers in a zone. Shopper density is time-varying and can be easily be extracted from our PathTracker[®] data.

The previous literature suggests that the social presence of other shoppers affects the three aspects of shopping (visit, shop, and buy) differently. Argo et al. (2005) find that shoppers

have a fundamental human motivation to “belong” (i.e., they desire interpersonal attachment (Baumeister and Leary 1995)). Visiting zones where other shoppers are present can create an initial level of social attachment, thus eliciting positive emotional response. Harrell et al. (1980) find that shoppers tend to conform to the traffic pattern of other shoppers. Further, Becker (1991) suggests that shoppers may be able to infer the “quality” of a zone (e.g., the presence of promotion) from the revealed visit behavior of other shoppers. Putting this together, we expect that shoppers are more likely to visit zones where the density of other shoppers is high. This is stated in Hypothesis H3a below.

H3a: (*Social Influence-Visit*) Consumers are more likely to visit zones where the density of other shoppers is high.

Once a shopper moves into a zone, the social presence of other shoppers also influences shopping and buying decisions (Harrell and Hunt 1976a,b). Harrell et al. (1980) suggest that under the conditions of crowding, shoppers may enact a set of behavioral adaptation strategies. More specifically, shoppers may adapt by delaying unnecessary purchases, exhibiting less exploratory behavior, and reducing their tendency to shop in the crowded zones. Thus, consistent with the previous literature, we hypothesizes that:

H3b: (*Social Influence-Shop*) Consumers are less likely to be in shopping mode in zones where the density of other shoppers is high.

H3c: (*Social Influence-Buy*) Consumers are less likely to buy at zones where the density of other shoppers is high.

Planning-ahead propensities

In addition to the three aforementioned situational factors, we also allow consumers to exhibit some extent of planning-ahead/forward-looking behavior in their shopping path,

consistent with the observation in Hui et al. (2008b). That is, when a consumer decides which zone to visit next, she considers not only the product categories in the focal zone, but also the location of the focal zone relative to other zones that she wants to visit later (within the same trip). As will be explained in detail in the model section, our model controls for planning ahead propensities. As a result, our model also allows us to empirically assess the degree of planning-ahead behavior that shoppers engage in. This will be discussed in more details in the results section.

The mixed effects of H1-H3, together with the accommodation of planning-ahead tendencies, highlight the value and importance of using our multidimensional (visit, shop, buy) framework. For instance, attempts to specify (and test) a simpler hypothesis linking shopper density and purchasing directly would be incomplete and potentially misleading. In order to examine this richer set of hypotheses, we now focus on developing our statistical model that will tie everything together in an integrated manner.

MODEL DEVELOPMENT

To test the aforementioned behavioral hypotheses, we develop an integrated individual-level probability model to capture each consumer's entire shopping path and purchase behavior. Given that our data are observational in nature, a well-specified model is necessary to control for heterogeneity across individuals and account for other baseline effects (e.g., the inherent difference between attractions of different categories and locations, each shopper's different planning-ahead tendencies, etc.). Thus, our model allows us to control for other confounding factors across individual observations (Freedman 2005), which, in turn, facilitates the testing of our focal hypotheses using our observation data (described in the next section).

We begin by defining category attractions and the derived zone attractions, then how a shopper's three decisions (visit, shop, buy) are modeled as a function of these constructs.

Category/zone attractions and baseline visit propensities

We define a vector of latent variables $\vec{a}_{it} = (a_{i1t}, a_{i2t}, \dots, a_{iKt})'$, where a_{ikt} denotes the attraction of category k for shopper i at step t . These *category attractions* directly drive the model of purchase behavior (and indirectly visitation and shop as described next) – categories with higher attractions to the shopper are assumed to be more likely to be purchased.

We then compute *zone attractions* based on the aggregation of category attractions of the product categories it contains. These zone attractions enter the model of shop and visit behavior, discussed later. The zone attraction for zone j for shopper i at step t is defined as:

$$A_{ijt} = \log \left(\sum_{k \in C(j)} \exp(a_{ikt}) \right) \quad (1)$$

where $C(j)$ denotes the set of product categories available at zone j . This specification is similar to the “inclusive value” notion that is commonly used in nested-logit models (McFadden 1981). In our framework, the zone can be viewed as a “nest” that contains several product categories.

As we discussed earlier, category attractions may not be constant over time. Thus, we allow them (and hence the derived zone attractions) to evolve depending on the shopper's visitation and purchase behavior up to step t . We capture the evolution of attractions as follows:

$$\begin{aligned} a_{ik(t+1)} &= a_{ikt} + \Delta_{ib} B_{ikt} + \Delta_{is} I\{k \in C(x_{it})\} & (k \neq \text{checkout}) \\ a_{ik(t+1)} &= a_{ikt} + \omega_v S_{it} & (k = \text{checkout}) \end{aligned} \quad (2)$$

For regular (non-checkout) product categories, we posit that after the shopper visits zone x_{it} , the attraction of the categories contained there will change by an amount indicated by Δ_{is} . If Δ_{is} is negative, the attraction of a product category decreases after a shopper visits the zone that

contains it. If category k is purchased at step t ($B_{ikt} = 1$), then the attraction for category j will further change by an amount indicated by Δ_{ib} . For the “checkout category,” ω_v measures the change in attraction to the checkout category based on the time that a consumer has already spent in the store. Thus, if ω_v is positive, the attraction of the checkout category increases as the shopper spends more time in the store; as a result, it reduces the tendency for shoppers to explore the store and instead gravitates a shopper towards the exit (as we will see in the model of visit). H1a can now be restated in terms of model parameter ω_v :

H1a: (*Time Pressure-Visit*) $\omega_v > 0$.

Model of visit

We begin by denoting the set of zones that are adjacent to the shopper’s current zone x_{it} as $M(x_{it})$. This represents the set of zones that the shopper can choose to visit in her next step.² Thus, the shopper’s visit choice can be viewed as a “choose-1-out-of- n ” choice problem, with n being the number of zones in $M(x_{it})$. To capture this zone-choice decision, we define a latent visit utility u_{ijt}^{VISIT} associated with the j -th zone as follows:

$$u_{ijt}^{VISIT} = Z_j + \psi_v R_{it} W_j + \gamma_v \rho_{ijt} + \kappa_t G_{ijt} + \varepsilon_{ijt}^{VISIT} \quad (3)$$

where Z_j denotes a zone-level baseline visit propensity, $\varepsilon_{ijt}^{VISIT}$ denotes error terms assumed to be i.i.d. extreme-value distributed. We assume that the shopper visits zone j in the next step if u_{ijt}^{VISIT} is larger than the latent utility of any of the other zones in the current choice set $M(x_{it})$, identical to the assumption in typical discrete choice models.

² We note that in our data, it is always possible to reach adjacent zones in one blink (five seconds). This was not chosen arbitrarily, but rather plays a key role in the zone definitions as described in the data section of the paper.

The second term $\psi_v R_{it} W_j$ represents the effect of licensing on visit behavior. R_{it} is an indicator variable that denotes the current “virtue-vice balance” of the shopping basket; it takes the value 1 if the current shopping basket contains more virtue categories than vice categories, and 0 otherwise³. W_j measures the “vice-ness” of the composition of zone j , and is defined as the number of vice categories in zone j divided by the total number of categories in zone j . ψ_v measures the directionality and magnitude of licensing effects on visit behavior. A positive ψ_v indicates that when a consumer has a “virtue” shopping basket, she will be more likely to visit zones with more vice categories. Thus, we restate H2a as follows:

H2a: (*Licensing-Visit*) $\psi_v > 0$.

The third term $\gamma_v \rho_{ijt}$ captures the social-influence effect of other shoppers. ρ_{ijt} denotes the (standardized⁴) density of other shoppers at zone j at step t (for shopper i). γ_v measures the effect of the social influence of other shoppers on the visit behavior of shopper i . A positive γ_v means that shopper i is more likely to visit zones that other shoppers are present. We restate H3a as follows:

H3a: (*Social Influence-Visit*) $\gamma_v > 0$.

The fourth term $\kappa_i G_{ijt}$ accounts for potential planning ahead behavior that consumers may exhibit. When planning ahead which zone to visit next, the shopper’s choice may involve a tradeoff between two aspects: (i) the intrinsic attraction of the adjoining zone, and (ii) by going

³ We tested alternative vice-virtue balance cutoffs besides the 50% measure that we use here. Our results are quite robust to other values and are available upon request.

⁴ To keep the relative density comparable across zones, the standardization is done by subtracting the mean and dividing by the standard deviation of zone densities across the entire store.

to the adjoining zone, whether she will be closer to other zones of high attraction. We capture this tradeoff by defining G_{ijt} as the time-varying attraction of zone j (A_{ijt} as in Equation 1) plus a weighted sum of the attractions of all other zones. The weight associated with zone j' is inversely proportional to the “distance” between zone j' and the focal zone j . Specifically,

$$G_{ijt} = A_{ijt} + \sum_{j' \neq j} \frac{A_{ij't}}{(1 + d_{jj'})^{\lambda_i}} \quad (\lambda_i \geq 0) \quad (4)$$

where $d_{jj'}$ denotes the length of the shortest path between zone j and zone j' . λ_i is a parameter that governs how shopper i trades off immediate utility with the more planning-ahead concern of reaching high attraction regions later on in her trip. For instance, $\lambda_i = \infty$ means that shopper i is myopic, i.e., only concerned about the attractiveness of what is immediately ahead when making the visitation choice. Thus, the estimate of λ_i allows us to assess the degree of planning-ahead behavior that consumer i exhibits. This is similar in spirit to work of Camerer et al. (2004) that looks at the degree of look-ahead behavior of subjects in experimental games.

From Equation (3) and (4), we can derive the likelihood regarding the shopper’s visit decision at step $t+1$:

$$P(x_{i(t+1)} = j, j \in M(x_{it})) = \frac{\exp \left[Z_j + \kappa_i \left(A_{ijt} + \sum_{j' \neq j} \frac{A_{ij't}}{(1 + d_{jj'})^{\lambda_i}} \right) + \psi_v R_{it} W_j + \omega_v \rho_{ijt} \right]}{\sum_{l \in M(x_{it})} \exp \left[Z_l + \kappa_i \left(A_{ilt} + \sum_{j' \neq l} \frac{A_{ij't}}{(1 + d_{lj'})^{\lambda_i}} \right) + \psi_v R_{it} W_l + \omega_v \rho_{ilt} \right]} \quad (5)$$

Model of shop

After arriving at a zone, the shopper may decide to shop in the current zone, in which case $H_{it}=1$ as defined earlier. We assume that the consumer shops if her latent “shop utility” is positive. Shop utility u_{ijt}^{SHOP} is defined as follows:

$$u_{ijt}^{SHOP} = \alpha_{is} + \beta_{is} A_{LX_{it}} + \omega_s T_{it} + \psi_s R_{it} W_j + \gamma_s \rho_{ijt} + \eta_j + \varepsilon_{it}^{SHOP} \quad (6)$$

where $\alpha_{is} + \beta_{is} A_{LX_{it}}$ denotes a linear function of the current zone attraction; α_{is} and β_{is} capture shopper i 's baseline shopping propensity and the extent to which her visit-to-shop behavior is correlated with latent attractions, respectively. η_j is a zone-specific random effect and ε_{it}^{SHOP} denotes random error assumed to be i.i.d. extreme value distributed.

The third term $\omega_s T_{it}$ captures the effect of time pressure on shop behavior. T_{it} denotes the total in-store time up to step t . The sign (and magnitude) of ω_s thus allows us to measure how perceived time pressure affects shop behavior. If ω_s is positive, it indicates that the shopper is more likely to shop at a zone after spending more time in the store. We therefore restate H1b as follows:

H1b: (*Time Pressure-Shop*) $\omega_s > 0$.

The fourth ($\psi_s R_{it} W_j$) and fifth ($\gamma_s \rho_{ijt}$) terms play similar roles as they do in the model of visit. A positive ψ_s means that when a consumer's shopping basket is relatively virtuous, she is more likely to shop at zones that contain more vice categories, as we hypothesized in H2b. A negative γ_s indicates that a shopper is less likely to shop at a zone if it contains a high density of other shoppers. Thus, we restate H2b and H3b as follows:

H2b: (*Licensing-Shop*) $\psi_s > 0$.

H3b: (*Social Influence-Shop*) $\gamma_s > 0$.

From Equation (6) we can derive the likelihood of a shop decision, given model parameters, as follows:

$$P(H_{it} = 1) = P(u_{ijt}^{SHOP} > 0) = \frac{e^{\alpha_{is} + \beta_{is} A_{ijt} + \omega_s T_{it} + \psi_s R_{it} W_j + \gamma_s \rho_{ijt} + \eta_{sit}}}{1 + e^{\alpha_{is} + \beta_{is} A_{ijt} + \omega_s T_{it} + \psi_s R_{it} W_j + \gamma_s \rho_{ijt} + \eta_{sit}}}. \quad (7)$$

Since a shopper is likely to spend more time in a zone if she is shopping there than if she is just passing through, we model the stay time (in each zone) using a pair of geometric distributions with different means depending on whether $H_{it} = 0$ or $H_{it} = 1$. Formally,

$$[S_{it} | H_{it} = 1] \sim \text{geometric}(\tau_{x_{it}}^{SHOP}) \quad (8)$$

$$[S_{it} | H_{it} = 0] \sim \text{geometric}(\tau_{x_{it}}^{PASS}) \quad (9)$$

For each zone, we assume that $\tau_j^{PASS} > \tau_j^{SHOP}$ (i.e., a shopper on average spends longer time in a zone if she is shopping). Thus we specify:

$$\text{logit}(\tau_j^{PASS}) = \text{logit}(\tau_j^{SHOP}) + \delta_j \quad (\delta_j > 0) \quad \text{for all } j. \quad (10)$$

Model of purchase

As mentioned earlier, we assume that a purchase in a zone is possible only if the consumer is shopping there ($H_{it} = 1$). If $H_{it} = 1$, the shopper buys from category k if it is available in her current zone ($k \in C(x_{it})$) and its ‘‘buy utility’’ u_{ikt}^{BUY} is positive. We specify u_{ikt}^{BUY} as follows:

$$u_{ikt}^{BUY} = \alpha_{ib} + \beta_{ib} a_{ikt} + \omega_b T_{it} + \psi_b R_{it} I_k^{vice} + \gamma_b \rho_{ijt} + \varepsilon_{ikt}^{BUY} \quad k \in C(x_{it}) \quad (11)$$

where α_{ib} and β_{ib} captures the shopper i 's baseline buying propensity and the extent to which shop-to-buy behavior is correlated with the latent attractions, respectively. I_k^{vice} is an indicator

variable that equals 1 if category k is a vice category, and 0 otherwise. The error terms ε_{ikt}^{BUY} are assumed i.i.d and extreme-value distributed.

Similar to its role in the models of visit and shop, the third term $\omega_b T_{it}$ captures the effect of time pressure on purchase behavior. We expect that ω_b is positive, i.e., the shopper is more likely to buy after spending more time in the store. The fourth term $\psi_b R_{it} I_k^{vice}$ captures the effect of licensing on the purchase of vice categories; a positive ψ_b indicates that if a shopper currently has a “virtuous” basket, she is more likely to purchase vice categories. Finally, the term $\gamma_b \rho_{ijt}$ captures the effect of social influence on the buy decision; we expect γ_b to be negative, i.e., a shopper is less likely to buy at zone that has a high density of other shoppers. To summarize, we have:

H1c: (*Time Pressure-Buy*) $\omega_b > 0$

H2c: (*Licensing-Buy*) $\psi_b > 0$

H3c: (*Social Influence-Buy*) $\gamma_b < 0$

From Equation (11), the likelihood for purchase behavior can be written as:

$$P(B_{ikt} = 1 | H_{it} = 1) = P(u_{ikt}^{BUY} > 0) = \frac{e^{\alpha_{ib} + \beta_{ib} a_{ikt} + \omega_b T_{it} + \psi_b R_{it} I_k^{vice} + \gamma_b \rho_{ijt}}}{1 + e^{\alpha_{ib} + \beta_{ib} a_{ikt} + \omega_b T_{it} + \psi_b R_{it} I_k^{vice} + \gamma_b \rho_{ijt}}} \text{ if } k \in C(x_{it}),$$

$$= 0 \text{ otherwise} \quad (12)$$

$$P(B_{ikt} = 0 | H_{it} = 0) = 1 \text{ for all } k. \quad (13)$$

Finally, to obtain the likelihood of a path, we multiply together the likelihood of each of the processes in Figure 1, i.e., visit, shop, and buy, for each step. The overall likelihood of the data can then be calculated by multiplying the likelihoods across all paths.

DATA

We estimate our model on data collected using the PathTracker[®] system, installed in a large supermarket in the Eastern United States. The system consists of a set of RFID tags and antennae: A small RFID tag is affixed under each shopping cart, and emits a uniquely coded signal every five seconds (“blinks”); this signal is then picked up by antennae around the perimeter of the store to locate the cart (Sorensen 2003). Purchase records (in terms of product UPC’s) were obtained from scanner data and matched to the paths, resulting in a complete record of a shopping trip. Thus, the structure of our data is similar to that collected by Burke (1993), who tracked shoppers in Marsh Supermarkets.

During our data collection period from March 14, 2004 to April 3, 2004, a total of 13486 raw trip *segments* were recorded by the PathTracker[®] system. This represents the in-store locations of all shopping carts that are recorded by RFID during the data collection period, and allow us to compute, at each given time, the number of shopping carts at each store zone. We then divide the number of carts at each zone by each zone’s area, to serve as a proxy for the density of shoppers at each location.

RFID is a relatively new data collection technology and does have certain caveats, however. First, shoppers who do not use shopping carts are not tracked. Thus, the measure of shopper density is not exact, but assumed to be a reasonable proxy for the actual density. Second, the PathTracker[®] system is unable to perfectly identify the start and end of every trip, thus many of the trips identified in the raw dataset represent only a segment of a complete grocery trip, and we remove them from our analysis: Of all the trips recorded, we have 1226 that start at the entrance and end at the checkout, corresponding to completed grocery trips. Further, some of these trips are not matched correctly with the associated purchases or have inconsistent purchase

records (i.e., a product is not visited during the trip, but a purchase is recorded). Keeping only the trips that correspond to complete shopping trips and have accurate purchase records, we end up with a dataset that contains 1051 paths (and associated purchase records). This dataset will be used to estimate our model, but all trip segments are used to compute shopper density.

We should note that while our final dataset with 1051 paths is only a small subset of all of the trips in the original dataset, a Bayesian statistical inference conditional on the smaller sample is still valid, as long as the data collection and preparation procedure is “ignorable” with respect to our model parameters (Gelman et al. 2003, p.201), a valid assumption here.⁵ Thus, we proceed to make statistical inferences on our model parameters conditional on our dataset of 1051 paths.

Data Discretization

Since our model, as discussed earlier, is a discrete choice model (McFadden 1981), the raw data need to be “discretized” to limit the number of possible locations (i.e., choice options). Similar to the approach used in Burke (1993), we divided the grocery store into 96 zones of comparable sizes, as shown in Figure 2. The location(s) of each product category across the 96 zones, along with its percent penetration (fraction of the 1051 shopping baskets containing the category), are shown in Table 1. Table 1 also classifies each category into vice, virtue, or neither. This classification was made by three independent judges; when raters disagreed (less than 5% of the time), they reached consensus through discussion.

[Insert Figure 2 and Table 1 about here]

⁵ Given that the missing data process in our case (i.e., the process that generates the incomplete trip segments) is due to the technicalities of the RFID system, the parameters governing the missing data process are independent of the parameters that govern the data generating process (i.e., our model parameters). This ensures that the condition of “distinct parameters” (Gelman et al. 2003, p.204) is satisfied, hence the data collection procedure is ignorable (see Gelman et al. 2003).

We then converted the discretized store into a mathematical graph, as shown in Figure 3. This graph defines, at each location, the set of zones that a shopper can reach in her next step (i.e., the set $M(x_{it})$ for the model of visit, Equation (5)). An implicit assumption in Figure 3 is that a pair of zones can be reached in one blink if and only if they are connected by an edge; this assumption has been empirically verified with our data and provides further validation of our zone definition.

[Insert Figure 3 about here]

Having discretized the store into 96 zones, we discretize each shopping path by mapping each (x,y) coordinate on a path at each blink to its corresponding zone. We then compute several summary statistics that describe consumers' visit, shop, and buy behavior.

Summary statistics for visit

We compute the total number of steps (i.e., zone transitions) that a shopper takes during the shopping trip, and the overall zone-to-zone transition probabilities. The histogram for the total number of steps is shown in Figure 4. In our dataset, the mean number of steps taken is 98.8 while the median is 90.0. The transitions that occur with highest frequency out of each zone are shown by the solid directed arrows in Figure 5, while the light shaded arrows indicate all possible movements.

[Insert Figures 4 and 5 about here]

Note from Figure 5 that there is a general tendency to “back-track” once a shopper enters an aisle; i.e., after a shopper enters an aisle, she is more likely to head out rather than traversing through it. This interesting observation is consistent with the common “excursion” and lack of aisle-traversal behavior documented in Larson et al. (2005) and Sorensen (2003).

Summary statistics for shop

We compute (i) the total amount of time (in minutes) that a shopper spent in the grocery store, and (ii) the average amount of time that shoppers spent in each zone in the store. The histogram for total in-store time is shown in Figure 6. In our dataset, shoppers on average spend 48.6 minutes in the store; the median in-store time is 43.8 minutes. The average amount of time shoppers spent in each zone (in minutes) is shown in Figure 7.

[Insert Figures 6 and 7 about here]

Summary statistics for purchase

We compute (i) the total number of categories that a shopper purchased during his/her trip, and (ii) the % purchase incidence (penetration) for each product category. The histogram of the total number of categories purchased is shown in Figure 8 (the leftmost bar represent trips with 1-2 category purchases). In our dataset, shoppers purchase, on average, from 6.7 categories.

[Insert Figure 8 about here]

RESULTS

Model validation

The posterior distribution of the hyperparameters that govern the individual-level parameters are summarized in Table 2. These estimates provide some face validity to our model. First, both μ_{β_s} and μ_{β_b} are positive, indicating that attractions are positively correlated with both visit-to-shop and shop-to-purchase decisions. Second, the estimates for both μ_{Δ_s} and μ_{Δ_b} are negative, suggesting that the attraction of a zone tends to decrease after a consumer visits the zone and/or purchases the product categories that it carries. Third, the reasonably large estimates of κ (mean of $\log(\kappa)$ is -1.32) suggests that purchase behavior is indeed interrelated with

visitation patterns, as expected, which indicates that an integrated model of path and purchase is necessary.

[Insert Table 2 about here]

The posterior means for the baseline attractions of the 10 highest-attractiveness categories are summarized in Table 3. Since purchase incidence is driven, in large part, by category attraction, we expect that category attractions should be positively correlated with simple purchase incidence statistics. Indeed, we find that the correlation between category attractions and purchase incidence is positive and highly significant ($r = 0.63$; $p < 0.001$). The product category that has the highest attraction is Fruit, with a posterior mean attraction of 2.83. This is well-aligned with the observation that Fruit also has the highest observed purchase incidence (53.8%).

[Insert Table 3 about here]

Next, we look at the zone-level parameters. The posterior means of τ_j^{SHOP} and Z_j for each zone are displayed using a choropleth map (Banerjee et al. 2004) in Figures 9 and Figure 10, respectively. As expected, zones with low τ_j^{SHOP} (hence a long mean shopping time) generally correspond to zones where shoppers spend longer time. The correlation between τ_j^{SHOP} and average observed time spent in the zone is negative and significant ($r = -0.39$; $p < .001$). In addition, the zones with high Z_j correspond to zones that are visited more often: the correlation between Z_j and observed zone penetration is positive and significant ($r=0.37$; $p < .001$).

[Insert Figures 9 and 10 about here]

Hypothesis testing

We now turn to the parameter estimates in Table 4, which correspond to the testing of the three sets of behavioral hypotheses H1-H3.

[Insert Table 4 about here]

For the hypotheses dealing with the effects of (perceived) time pressure (H1a, H1b, H1c), we found support for our predicted effects. We proposed that as the shopper spends more time in the store, she depletes her “shopping time budget,” and gradually increases her perceived time pressure. As a result, the shopper adapts by becoming less exploratory and more purposeful as the trip progresses. Consistent with our hypothesis H1a, the estimate for ω_v is positive ($m=.008$; $p<.05$), indicating that the attraction of the checkout does increase during the trip, thus reducing the tendency for shoppers to explore the store and instead gravitate towards checkout. Further, the estimates of ω_s ($m=0.0012$; $p<.05$) and ω_b ($m=0.0005$; $p<.05$) are both positive, indicating that the consumers are more likely to shop and buy as the trip progresses and (perceived) time pressure intensifies. This supports the behavioral adaptation strategy proposed in H1b and H1c.

Next, we move on to the set of behavioral hypotheses (H2a, H2b, H2c) that captures licensing effects on visit, shop, and buy behavior. Our data provide only limited support for licensing behavior. First, the estimate for ψ_v is not significantly different from 0 ($m=-.021$; n.s.), which means that we do not find licensing behavior to affect visit decisions. Second, consistent with H2b, the estimate for ψ_s is positive, but only marginally significant ($m=.142$; $p<.1$); this indicates a weak effect of licensing on shop behavior. When visiting a zone, consumers who have a shopping basket that contains more virtue than vice are slightly more likely to shop there if the zone contains more vice categories. Third, the estimate for ψ_b is not significantly different from 0, which means that we do not find licensing behavior on the buy decision, conditional on a shop decision being made. However, note that due to nested nature of our shop/buy model (see

Equation 7 and Equation 12), the increased likelihood of a shop conversion at zones with vice categories indirectly increases the marginal likelihood of purchasing a vice category.⁶ Thus, taken together, our field data provides some weak evidence for the licensing effect on the shopping (direct) and purchasing (indirect) of vice categories, but not on consumers' visit decisions. We discuss in the conclusions section why we may have observed only limited support for licensing effects in our study.

We now turn to the set of hypotheses that captures the social influence of other shoppers on a consumer's visit, shop, and buy decision (H3a, H3b, and H3c). We find that, consistent with H3a, γ_v is positive and significant ($m=0.012$; $p<.05$); i.e., consumers are more likely to visit zones that contain a higher density of other shoppers. Consistent with H1a, the presence of other shoppers generally attracts a consumer to visit a store zone. Once a consumer is attracted into a store zone, however, she is less likely to shop there when the density of other shoppers is high (i.e., γ_s is negative and marginally significant ($m=-0.034$; $p<.1$)). This finding is consistent with the literature on crowding (Harrell et al. (1980)). The estimate for γ_b is not significantly different from zero; thus, the presence of other shoppers in a store zone does not have a significant effect on consumers' buying behavior, once they have entered a "shopping" mode.

Finally, we assess the extent to which consumers exhibit planning-ahead behavior when formulating their in-store paths. We find that, consistent with our model assumptions, λ is small and finite, with a posterior mean of 0.442 and a 95% posterior interval of (0.423, 0.463). As we discussed earlier, a small estimate of λ indicates the existence of in-store planning-ahead behavior, which is consistent with the finding in Hui et al. (2008b) that many grocery shoppers plan ahead during their in-store trips.

⁶ To see this, note that $P(\text{buy}) = P(\text{shop}) * P(\text{buy} | \text{shop})$. Thus, the marginal likelihood of purchase, $P(\text{buy})$, increases if $P(\text{shop})$ increases even if the $P(\text{buy}|\text{shop})$ stays unchanged.

GENERAL DISCUSSION

In this article, we examine three sets of established behavioral hypotheses about consumers' in-store shopping behavior (the effect of (perceived) time pressure, licensing, and the social presence of other shoppers) using field data from an actual grocery store. We develop an individual-level probability model that incorporates the effects of those behavioral hypotheses on shoppers' in-store visit, shop, and buy decisions. Using latent category attractions and zone attractions, our model integrates three aspects of grocery shopping: (1) where shoppers visit and their zone-to-zone transitions, (2) how long they stay and shop in each zone, and (3) what product categories they purchase.

Our results provide consistent directional support for the aforementioned behavioral hypotheses, although the strength of these effects varies. First, as consumers spend more time in the store, they become more purposeful in their trip — they are less likely to spend time on exploration, and are more likely to shop and buy. Second, we also find (weak) support for licensing behavior (Khan and Dhar 2006). After purchasing virtue categories, consumers are more likely to shop at locations that carry vice categories. Licensing, however, does not significantly affect visit decisions. Third, the social presence of other shoppers attracts consumers towards a zone in the store, but reduces consumers' tendency to shop at that zone. Finally, we also provide some evidence that consumers exhibit planning-ahead behavior during their in-store shopping trip.

It is worthwhile to point out a few limitations of our study. First, as we discussed earlier, the PathTracker[®] system tracks only shoppers who utilize shopping carts, but not those who carry shopping baskets. Thus, our results may not be fully generalizable to shoppers who are performing “fill-in” trips. Despite this shortcoming, we believe that our field study is still a

major step forward in enhancing the external validity of the focal behavioral hypotheses, which have been previously tested almost exclusively in lab environments.

In addition, our operationalization of “virtue” and “vice” products is defined at the product category level; thus, we are unable to further differentiate between relative vice and virtue SKUs within a product category (for example, a diet product, a relative virtue, within a carbonated drink category, a vice category). This, and other reasons, may partially explain the relatively weak licensing effects observed from our results.

In addition to testing behavioral theories, our study also may lead to important managerial implications regarding the design of store layout, similar to the way urban planners use sophisticated models to design urban spaces to avoid crowding conditions (www.crowddynamics.com). Crowding (or more generally the social influence of other shoppers considered here) in the store environment is a two-edged sword: while it attracts shoppers to a zone to “check it out,” it also reduces shopping tendency once the shopper enters that zone. How to design store layout to achieve the “optimal” level of crowding is an important topic for retailers, but also a very difficult and computationally intensive problem. Our model offers a potential solution to solve this problem. Given a different store layout, retailers may simulate path and purchase data from our model and optimize the design against specific criterion (e.g., the penetration of a certain category, gross margin). This allows retailers to experiment with different store layouts economically.

Looking forward, this research can be extended in many directions through the collection and analysis of additional data. First, researchers can consider combining shopping path data with surveys collected before or after the shopping trip. For instance, one can ask consumers to state their shopping goals (Lee and Ariely 2006) before entering the store, and study how the

propensity of unplanned purchase (Inman et al. 2007) is related to their path behavior. By asking consumers to state their purchase goals before the start of their trip and using that as a control variable, the influence of social interaction can be tested more unambiguously. That is, we can tell whether consumers just happen to visit the same zone at similar time-of-day, or whether social effects are genuine.

Second, researchers may consider a cross-store study. The PathTracker[®] system is being installed in an increasing number of supermarkets (and other types of retail stores) around the world. A cross-store study will likely introduce more variation in store layouts and thus reduce the confounding between category and zone attractions. Further, we may study how store characteristics (e.g., square footage, number of aisles) are related to consumers' shop/purchase behavior. For instance, Meyers-Levy and Zhu (2007) demonstrated how ceiling height affect consumers' information processing and with store varying layout information; a cross-store study can be used to test their hypothesis.

In summary, we believe that this research is an important step in the continuing line of research papers that tightly link behavioral theories to statistical models for field data, in the spirit of studies such as Hardie et al. (1993) and Schweidel et al. (2006). Our hope is that this interplay between careful theory development and rigorous statistical testing can provide external validation to what may start out as laboratory-based findings, but also provide new empirical insights that can lead to the development of new theories to be subsequently tested under cleaner laboratory conditions.

REFERENCES

- Argo, Jennifer J., Darren W. Dahl, and Rajesh V. Manchanda (2005), "The Influence of a Mere Social Presence in a Retail Context," *Journal of Consumer Research*, 32, 207-212.
- Banerjee, Sudipto, Bradley P. Carlin, and Alan E. Gelfand (2004), *Hierarchical Modeling and Analysis of Spatial Data*, Chapman and Hall.
- Baumeister, Roy F., and Mark R. Leary (1995), "The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation," *Psychological Bulletin*, 117(3), 497-529.
- Becker (1991), "A Note on Restaurant Pricing and Other Examples of Social Influence on Price," *Journal of Political Economy*, 99, 1109-1116.
- Bradlow, Eric T., and David C. Schmittlein (2000), "The Little Engines That Could: Modeling the Performance of World Wide Web Search Engines," *Marketing Science*, 19(1), 43-62.
- Burke, Raymond R. (1993), "Marsh Supermarkets, Inc. (A): The Marsh Super Study," *Harvard Business School Case*, #9-594-042.
- Camerer, C., T. Ho, and J. Chong (2004), "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119(3), 861-898.
- Dhar, Ravi, and Stephen M. Nowlis (1999), "The Effect of Time Pressure on Consumer Choice Deferral," *Journal of Consumer Research*, 25(4), 369-384.
- Freedman, David A. (2005), *Statistical Models: Theory and Practice*, Cambridge University Press.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (2003), *Bayesian Data Analysis, 2nd Edition*, Chapman & Hall.

- Hardie, Bruce, Eric Johnson, and Peter Fader (1993), "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," *Marketing Science*, 12 (Fall), 378-394.
- Harrell, G. D. and M. D. Hutt (1976a), "Buyer Behavioral Under Conditions of Crowding: An Initial Framework," in *Advances in Consumer Research*, Vol. 3, B. B. Anderson Ed., Cincinnati, Ohio, Association for Consumer Research.
- Harrell, G. D., and M. D. Hutt (1976b), "Crowding in Retail Stores," *M.S.U. Business Topics* (Winter), 33-39.
- Harrell, Gilbert, Michael D. Hutt, and James C. Anderson (1980), "Path Analysis of Buyer Behavior Under Conditions of Crowding," *Journal of Marketing Research*, 17, 45-51.
- Hui, Sam K., Peter S. Fader, and Eric T. Bradlow (2008a), "Path Data in Marketing: An Integrative Framework and Prospectus for Model-Building," *Marketing Science*, forthcoming.
- Hui, Sam K., Peter S. Fader, and Eric T. Bradlow (2008b), "The Traveling Salesman Goes Shopping: The Systematic Inefficiencies of Grocery Paths," *Marketing Science*, forthcoming.
- Hruschka, Harald, Martin Lukanowicz, and Christian Buchta (1999), "Cross-Category Sales Promotion Effects," *Journal of Retailing and Consumer Services*, 6, 99-105.
- Khan, Uzma, and Ravi Dhar (2006), "Licensing Effect in Consumer Choice," *Journal of Marketing Research*, 43, 259-266.
- Larson, Jeffrey S., Eric T. Bradlow and Peter S. Fader (2005), "An Exploratory Look at Supermarket Shopping Paths," *International Journal of Research in Marketing*, 22, 395-414.

- Latane, Bibb (1981), "The Psychology of Social Impact," *American Psychologist*, 36(4), 343-356.
- McFadden, D. L. (1981), *Structural Analysis of Discrete Data with Econometric Applications*. MIT press.
- Meyers-Levy, Joan, and Rui (Juliet) Zhu (2007), "The Influence of Ceiling Height: The Effect of Priming on the Type of Processing that People Use," *Journal of Consumer Research*, 34 (August), 174-186.
- Park, C. Whan, Easwar S. Iyer, and Daniel C. Smith (1989), "The Effect of Situational Factors in In-Store Grocery Shopping Behavior: The Role of Store Environment and Time Available for Shopping," *Journal of Consumer Research*, 15 (March), 422-433.
- Rossi, Peter E., Greg M. Allenby, and Robert McCulloch (2006), *Bayesian Statistics and Marketing*, Wiley.
- Schweidel, D., E. T. Bradlow, and P. Williams (2006), "A Feature-Based Approach to Assessing Advertising Similarity," *Journal of Marketing Research*, 43(2), 237-243.
- Sorensen, Herb (2003), "The Science of Shopping," *Marketing Research*, 15(3), 30-35.
- Suri, Rajneesh, and Kent B. Monroe (2003), "The Effects of Time Constraints on Consumers' Judgments of Price and Products," *Journal of Consumer Research*, 30(1), 92-104.
- Thaler, Richard H. (1999), "Mental Accounting Matters," *Journal of Behavioral Decision Making*, 12, 183-206.

APPENDIX:

HIERARCHICAL BAYES FRAMEWORK AND MCMC PROCEDURE

Since consumers may have heterogeneous category preferences, shopping characteristics, and planning-ahead propensities, we embed our individual-model within a Hierarchical Bayesian framework. Each consumer has a different set of parameters that are assumed to be drawn from a common distribution, thus allowing us to borrow strength across customers to calibrate our model. To ensure model identifiability, a simulation experiment was conducted (and yielded excellent parameter and summary statistics recovery); details are available upon request.

The parameter vector for the i -th consumer, $(\bar{a}_{i0}, \kappa_i, \alpha_{is}, \alpha_{ib}, \beta_{is}, \beta_{ib}, \Delta_{is}, \Delta_{ib}, \lambda_i)'$, is assumed to be drawn from a set of common prior distributions. In the discussion below, we first specify the prior for the initial attraction vector \bar{a}_{i0} , then the prior for the rest of the parameters.

For the attraction vector, we specify

$$\bar{a}_{i0} \sim N(\bar{\mu}_A, \Sigma_A). \quad (\text{A1})$$

The variance-covariance matrix Σ_A allows us to borrow strength across categories by taking into account category complementarities. In particular, the (k, k') -th entry of Σ_A corresponds to the degree of complementarity between category k and category k' . For example, if category k and k' are complements, given that a person has purchased category k , we might expect that category k' is more likely to be purchased in the same trip as well. In this case, one may expect that the entry $\Sigma_{A(k, k')}$ will be large and positive. In general, Σ_A could be an unrestricted $N \times N$ matrix, with N being the number of categories. To reduce the number of

parameters, we impose a 2-dimensional factor analytic structure on Σ_A .⁷ Other studies that use a similar approach to capture dependence structures across categories include Hruschka et al. (1999). Formally, let $z_j = (z_{k1}, z_{k2})$ be the “spatial position” of the k -th category. We model Σ_A as

$$\Sigma_{A[k,k]} = \sigma^2 \exp(-\|z_k - z_{k'}\|) \quad (\text{A2})$$

where $\|z_k - z_{k'}\| = \sqrt{(z_{k1} - z_{k'1})^2 + (z_{k2} - z_{k'2})^2}$.

For model identification, the variance parameter σ^2 is set equal to 1. The variance hyperparameters and the “positions” $\vec{z} = (z_1, z_2, \dots, z_N)$ are given independent standard Gaussian diffuse priors $N(0, 100^2)$ and are jointly estimated with other parameters in our model. Following Bradlow and Schmittlein (2000), we set the first category at the origin, the second category on the x-axis, and the third category on the y-axis to control for shift, rotation around origin, and reflection about the x-axis respectively.

The other individual-level parameters (after suitable transformations) are assumed to follow standard multivariate Gaussian hyperpriors:

$$(\log(\kappa_i), \alpha_{is}, \beta_{is}, \alpha_{ib}, \beta_{ib}, \Delta_{is}, \Delta_{ib}, \log(\lambda_i))' \sim MVN(\vec{\mu}_I, \Sigma_I). \quad (\text{A3})$$

Similarly, zone-level parameters $(Z_j, \tau_j^{PASS}, \delta_j)$ for each zone are assumed to be drawn from a common multivariate Gaussian distribution:

$$(Z_j \quad \text{logit}(\tau_j^{PASS}) \quad \log(\delta_j))' \sim MVN(\mu_{ZONE}, \Sigma_{ZONE}). \quad (\text{A4})$$

For model identification, the mean hyperparameter associated with Z_j is set to 0.

⁷ Our model can be generalized to include a D-dimensional map. In particular, we fit the model using D=2 and D=3; both model fits and parameter estimates are very similar. Thus, we restrict our attention to the D=2 case for ease of computation.

To complete our Hierarchical Bayesian model specification, we specify a set of weakly informative, conjugate priors for all hyperparameters in our model. We now briefly outline the MCMC procedure used to draw samples from their posterior distributions.

In each iteration, we draw from the full conditional distribution of each parameter in the following order.

- I. Individual-level attractions (\vec{a}_{i0})
- II. Individual parameters ($\log(\kappa_i), \alpha_{is}, \beta_{is}, \alpha_{ib}, \beta_{ib}, \Delta_{is}, \Delta_{ib}, \log(\lambda_i)$)
- III. Zone-level parameters ($Z_j, \text{logit}(\tau_j^{PASS}), \log(\delta_j)$)
- IV. The location parameters z_j 's for cross-category correlation
- V. Hyperparameters for individual-level parameters ($\vec{\mu}_A, \Sigma_A, \vec{\mu}_I, \Sigma_I$)
- VI. Hyperparameters for zone-level parameters ($\mu_{ZONE}, \Sigma_{ZONE}$)

For Steps I, II, III, and IV, each parameter is sampled one at a time from its full conditional distribution. A Metropolis-Hastings algorithm with a Gaussian random walk proposal distribution is used to draw from the full conditional distribution of the focal parameter. The scale of the Gaussian distribution is adjusted to obtain an acceptance ratio of around 50% (Gelman et al. 2003). Acceptance ratios are continuously monitored over the iterations to ensure that our posterior samples have good mixing properties.

For Steps V and VI, the full conditional distribution of the hyperparameters can be drawn using standard close-form computation of the multivariate normal distribution with conjugate prior (see Gelman et al. 2003, p.78).

We run the MCMC algorithm for 2000 draws. The first 1000 draws are discarded as burn-in sample (Gelman et al. 2003), and the last 1000 draws are kept as draws from the posterior distribution. Posterior means (along with 95% posterior intervals) are reported.

Category Name	Zones	%buy
Fruit ^f	2,4	53.8%
Vegetables ^f	3,4,5	50.4%
Butter/Cheese/Cream ^v	38,39,82,83	38.0%
Carbonated Beverages ^v	16,21,22,23	24.2%
Salty Snacks ^v	62,63,64,92	23.2%
Cookies/Crackers ^v	18,44,45,46,47,93	22.6%
Milk ^f	38	22.6%
Ice Cream ^v	57,58,59,60	19.6%
Bread ^f	52,53,61,93	19.4%
Candy/Gum/Mints ^v	60,91,92	17.3%
Cereal ^f	49,50,94	17.1%
Eggs ^f	36	14.7%
Canned Vegetables ^f	47,61	12.7%
Baking Ingredients	18,24,25,26,27	12.2%
Frozen Prepared Dinners	55,56	11.9%
Drinks (others) ^f	52,53,94	11.9%
Yogurt ^f	81	11.5%
Pasta Sauce ^f	14,30	11.2%
Fruit Juice ^f	36	10.8%
Canned Dried Fruit ^f	20,95	10.8%
Pet Care	60,65,66,67	10.7%
Meat/Poultry/Seafood Manufactured Prepack	31,35	10.3%
Canned Soup ^f	44,61	9.7%
Frozen Pizza Snacks ^v	55,56	9.1%
Bath Tissue	37,77	9.0%
Frozen Vegetables ^f	54	8.6%
Peanut Butter/Jams	48,61	7.7%
Bottled Water ^f	23,40	7.6%
Prepared Food/Dried Dinners ^f	29,95	7.4%
Frozen Meat/Poultry/Seafood	54,56	7.0%
Pasta	30	6.9%
Frozen Drinks	57	6.1%
Pastry/Snack Cakes ^v	51	5.8%
Granola Bars ^f	19,94	5.3%
Bagels/Breadsticks ^f	52,53,73	5.2%
Spices/Seasonings	16,26,46,95	4.9%
Magazines	77,91,92	4.9%
Condiments/Sauces ^f	24,25,26	4.7%
Frozen Baked Goods	57,58	4.6%
Tobacco ^v	90,91	4.6%
Household Cleaners ^f	78,79	4.4%
Facial Tissue ^f	76,84	4.4%
Paper Towels ^f	37,75	4.4%
Coffee ^v	50	4.3%
Frozen Potatoes/Onions ^f	54	4.2%
Oral Care ^f	74,91,92	4.2%
Prepackaged Deli Meat	34	4.2%
Frozen Dessert/Fruit ^v	58,93	4.0%
Canned Seafood	40	3.7%
Non-Refrigerated Dressings	25	3.6%
Disposable Tableware ^v	69,94	3.6%
Olives/Peppers/Pickles ^f	24	3.5%
Dough Products	39	2.9%
OTC Medicines ^f	74,88,91,92	2.9%
Beer ^v	62,63,93	2.9%
Non-Carbonated Flavored Drinks ^v	51	2.8%
Skin/eye care ^f	84,85,86,87,88	2.6%

Category Name	Zones	%buy
Shampoo/Conditioner ^f	81,82	2.5%
Laundry Supplies ^f	78,79	2.5%
Natural/Organic Food ^f	7	2.5%
Pudding/Dry Dessert ^v	25	2.1%
Rice	42	2.1%
Shelf-Stable Milk	27	1.9%
Bakery Service	8,10	1.7%
Hot Beverage Add-Ins ^f	49	1.7%
Canned RTE Meat Entrées	40	1.7%
Baby Food ^f	71	1.6%
Stationery/School Supplies	69,70	1.6%
Wine ^v	28,29	1.5%
Refrigerated Snacks	81	1.5%
Ethnic (Oriental)	41	1.5%
Ethnic (TexMex)	43	1.5%
Toaster Pastries ^v	48	1.4%
Paper and Plastic Bags	68	1.4%
Special Diet Items ^f	9	1.4%
Cooking Oil	27	1.3%
Salad Add-Ins ^f	27	1.3%
Natural/Organic Snacks ^f	11	1.3%
Canned Meat	40	1.2%
Toiletries	87,90,91,92	1.2%
Meat/Poultry/Seafood Fresh Prepack	32	1.2%
Ethnic (Hispanic)	43	1.1%
Rolls/Buns/Pitas ^f	52,53	1.0%
Prepackaged Deli Prepared Lunch ^f	14	1.0%
Prepared Food/Potatoes ^f	45	1.0%
Tea ^f	49	0.9%
Frozen Dough/Bread/Bagel ^f	58	0.9%
Electronic Media	89	0.9%
Cosmetics/Deodorant ^v	86	0.9%
Pancake/Syrup ^v	26,48	0.9%
Deli Prepack	13,15	0.8%
Feminine Hygiene	72	0.7%
Dry Soup ^f	45	0.7%
Hard Liquor ^v	42,43	0.6%
Baby Medical Needs ^f	71,72	0.6%
Baking Supplies	61	0.6%
Hair Color Accessories ^v	83	0.6%
Batteries ^f	80,84	0.5%
Light Bulbs ^f	80	0.5%
Office Supplies ^f	75	0.5%
Plastic Wrap ^f	68	0.5%
Deli Service	12	0.4%
Dried Beans/Peas ^f	43,47	0.4%
Natural/Organic Drinks ^f	11	0.4%
Aluminum Foil ^f	68	0.4%
Napkins ^f	76	0.4%
Hot Chocolate Mix ^f	49	0.3%
Deli Amenities	15	0.3%
Automotive Supply	67	0.1%
Apparel	73	0.1%
Meat/Poultry/Seafood Fresh Service	17,31	0.1%
Meat/Poultry/Seafood Fully/Partially Cooked	33	0.1%
Floral	2,6	0.0%
Natural/Organic (Others) ^f	7	0.0%

Table 1. Locations of product categories. Superscripts ^v: vice category, ^f: virtue category.

	Posterior Mean	Posterior S.D.	95% Posterior Interval
μ_{κ}	-1.323	0.015	(-1.351, -1.291)
μ_{α_s}	-2.509	0.049	(-2.597, -2.411)
μ_{β_s}	0.665	0.019	(0.632, 0.708)
μ_{α_b}	-2.940	0.048	(-3.028, -2.837)
μ_{β_b}	1.529	0.029	(1.470, 1.578)
μ_{Δ_s}	-0.336	0.012	(-0.358, -0.314)
μ_{Δ_b}	-0.348	0.015	(-0.377, -0.320)
μ_{λ}	-0.817	0.023	(-0.860, -0.771)

Table 2. Estimation results for model hyperparameters.

Product Category	Attraction	Product Category	Attraction
Fruits	2.83	Salty Snacks	1.57
Vegetables	2.29	Meat/Poultry/Seafood Manufactured Prepack	1.44
Natural/Organic Food	2.26	Pastry/Snack Cakes	1.31
Special Diets	2.11	Rice	1.29
Butter/Cheese/Cream	1.92	Milk	1.19

Table 3. Posterior mean for category attractions for the 10 categories with the highest attraction, sorted in decreasing order.

Hypothesis	Parameter	Posterior Mean	95% Posterior Interval	Interpretation
H1a (<i>time pressure-visit</i>)	ω_v	0.008*	(0.007, 0.009)	Supported
H1b (<i>time pressure-shop</i>)	ω_s	0.0012*	(0.0011, 0.0014)	Supported
H1c (<i>time pressure-buy</i>)	ω_b	0.0005*	(0.0004, 0.0007)	Supported
<hr/>				
H2a (<i>licensing-visit</i>)	ψ_v	-0.021	(-0.070, 0.032)	Not supported
H2b (<i>licensing-shop</i>)	ψ_s	0.142^	(-0.021, 0.321)	(marginally) supported
H2c (<i>licensing-buy</i>)	ψ_b	-0.086	(-0.214, 0.028)	Not supported
<hr/>				
H3a (<i>social influence-visit</i>)	γ_v	0.012*	(0.005, 0.019)	Supported
H3b (<i>social influence-shop</i>)	γ_s	-0.034^	(-0.085, 0.012)	(marginally) supported
H3c (<i>social influence-buy</i>)	γ_b	0.017	(-0.036, 0.068)	Not supported

Table 4. Hypothesis testing results (* $p < .05$; ^ $p < .1$)

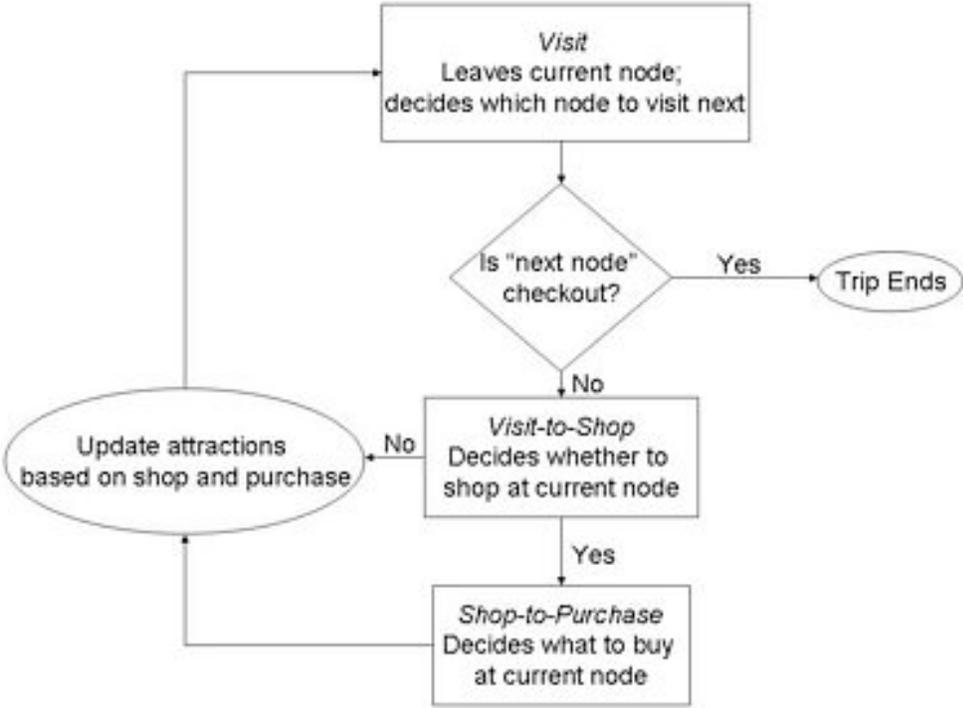


Figure 1. The shopper's in-store decision process.

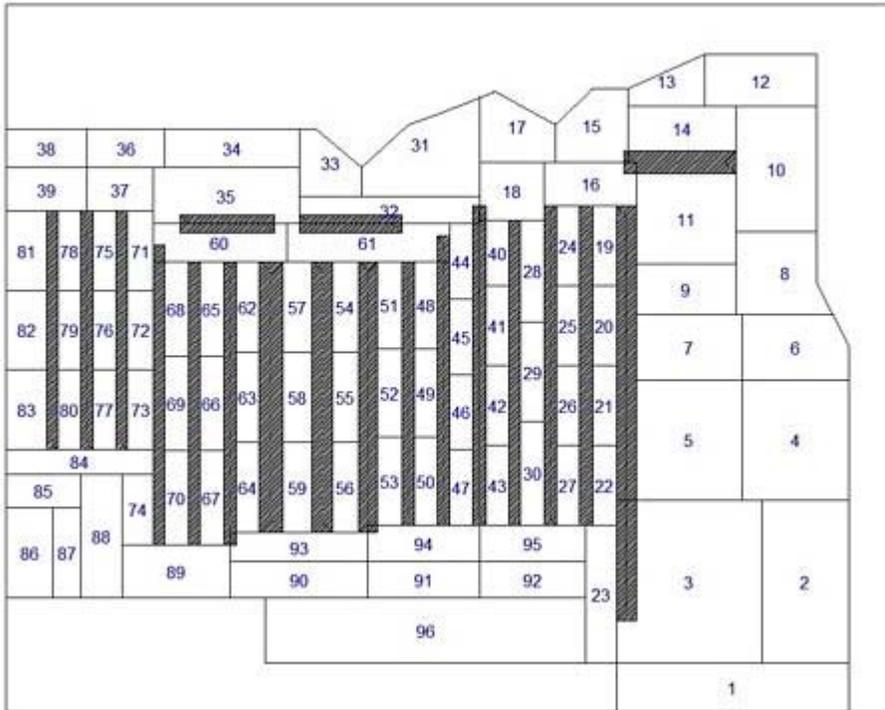


Figure 2. Grocery store discretized into 96 zones.

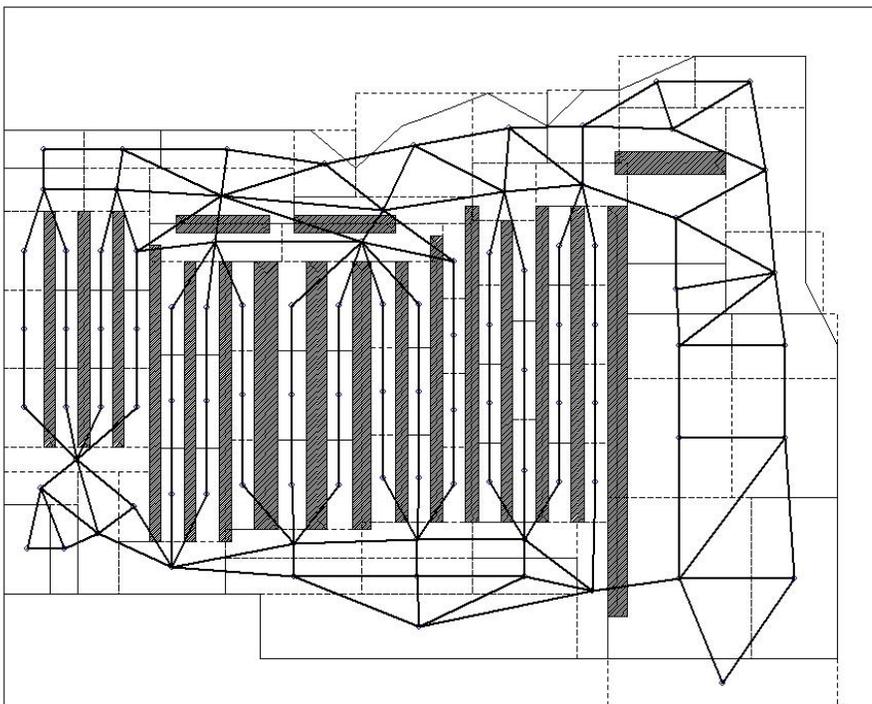


Figure 3. Grocery store represented by a graph of 96 nodes.

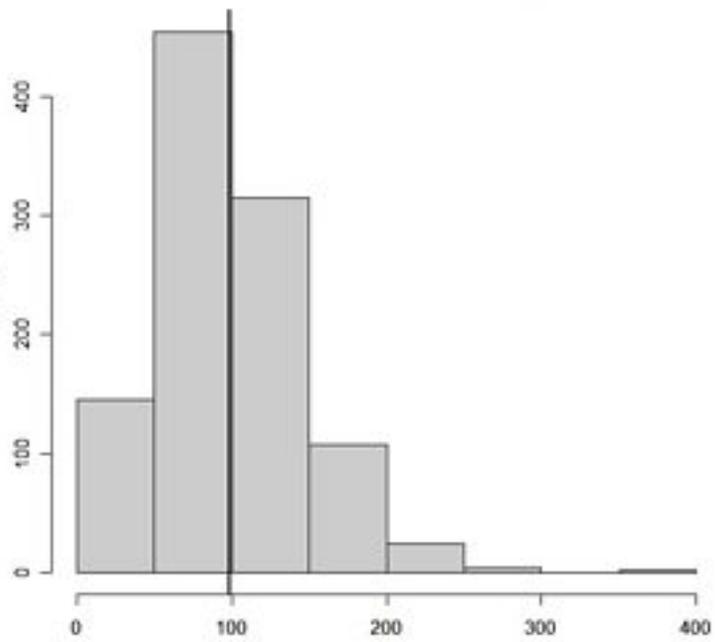


Figure 4. Histogram of number of steps (vertical line denotes the mean).

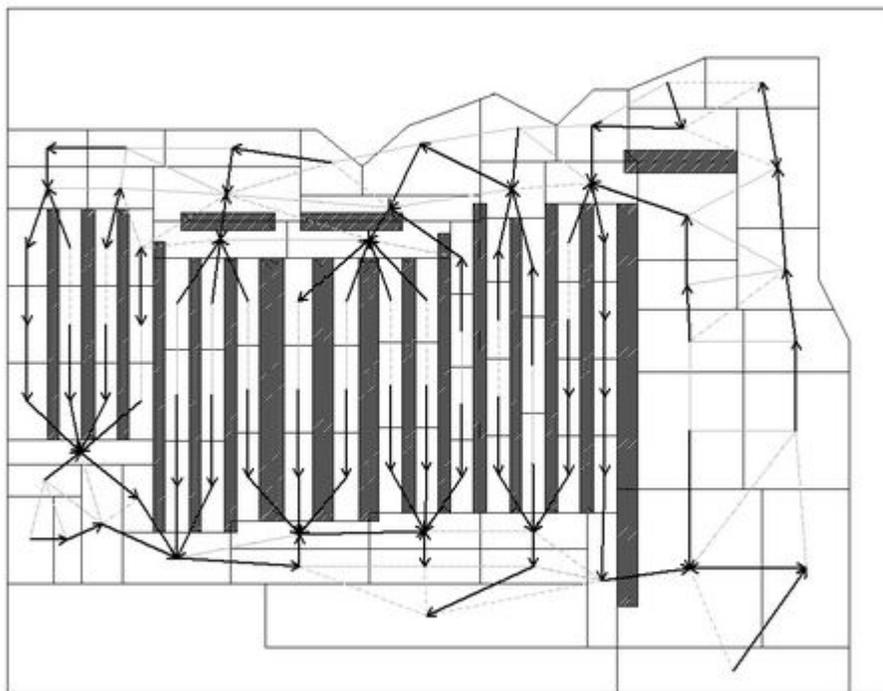


Figure 5. Most frequent transition out of each zone.

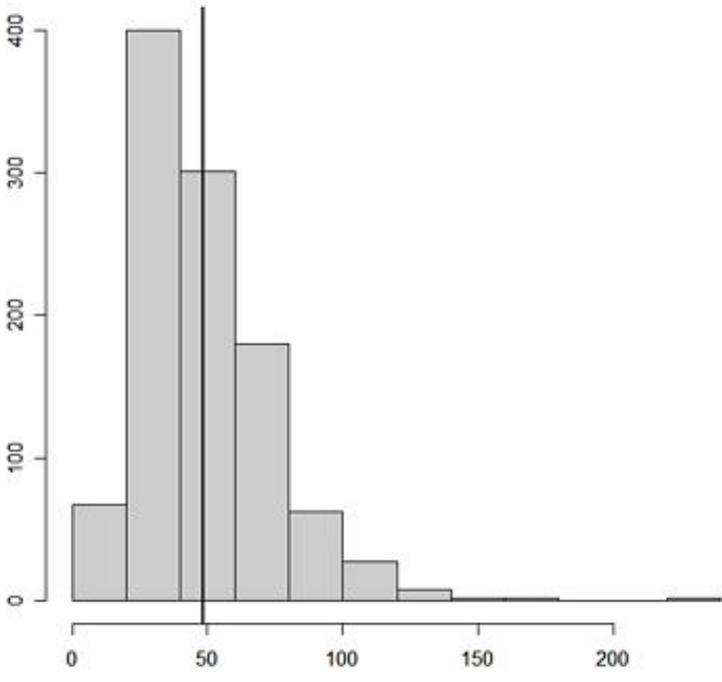


Figure 6. Histogram of total in-store time in minutes (vertical line denotes the mean).



Figure 7. Average time a shopper spent (in minutes) in each zone.

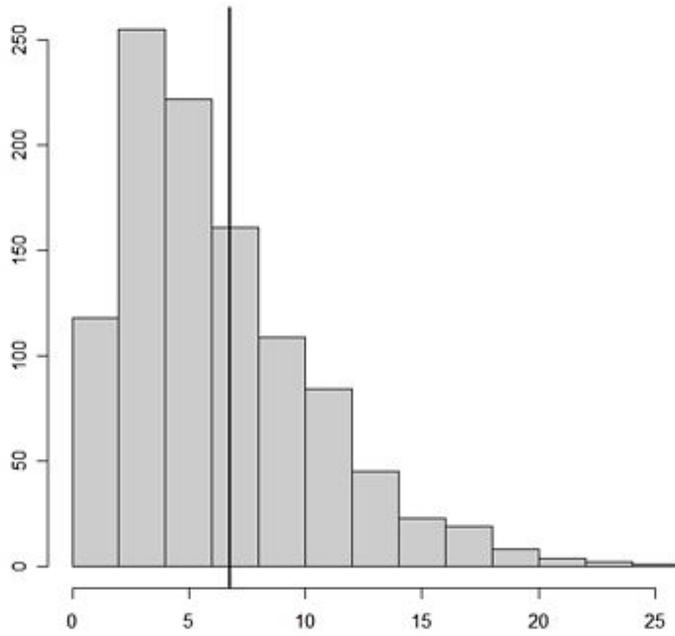


Figure 8. Histogram of the total number of product categories purchased (vertical line denotes the mean).



Figure 9. τ_j^{SHOP} for each zone; zones with longer shopping time are shaded in darker gray.



Figure 10. Z_j for each zone; zones with higher Z_j are shaded in darker gray.