Mark Lett (2012) 23:177–194
DOI 10.1007/s11002-011-9145-2

Identifying consumer heterogeneity in unobserved categories

Simon J. Blanchard · Wayne S. DeSarbo · A. Selin Atalay · Nukhet Harmancioglu

Published online: 10 August 2011

Abstract Categorization has been extensively studied in both the psychology and marketing literatures. However, very little methodological research has demonstrated the heterogeneity in consumers' unobserved category structures and activations. We propose a new latent structure procedure that simultaneously identifies the unobserved categories that consumers use and represents consumer heterogeneity via different groups of consumers who have activated different unobserved latent categories. The results of an empirical study in Sports Marketing about sports fans' perceptions of various sports illustrates how the proposed methodology can capture heterogeneity at the group level and account for a variety of different category structures.

Keywords Categorization · Latent structure analysis · Heterogeneity · Sports marketing

1 Introduction

Consumers naturally categorize stimuli (i.e., objects, products, brands, etc.) to simplify their decision making processes (Rosch and Mervis 1975; Mervis and Rosch 1981).

Electronic supplementary material The online version of this article (doi:10.1007/s11002-011-9145-2) contains supplementary material, which is available to authorized users.

S. J. Blanchard
McDonough School of Business, Georgetown University, Washington, DC 20057, USA
e-mail: sjb247@georgetown.edu

W. S. DeSarbo (✉)
Marketing Department, Smeal College of Business, The Pennsylvania State University,
433 Business Building, University Park, PA 16802, USA
e-mail: wsd6@psu.edu

A. S. Atalay
HEC Paris, Paris, France
e-mail: atalay@hec.fr

N. Harmancioglu
Koc University, Koc, Turkey
e-mail: nharmancioglu@ku.edu.tr

Categories in Marketing are sets of products, services, brands, entities, or states that are perceived by consumers as being related in some way (Loken et al. 2008). How these sets are perceived to be related to each other varies significantly across individuals. Consider a consumer who is trying to make a choice among various snack options. How would this consumer perceive the options, categorize them, and make a choice that satisfies her/his needs? A craving for salty food, a nut allergy, as well as a newly activated fitness goal are likely to have an impact. The options can be further grouped based on package size, color of package, price, how filling the option is, etc. In addition, would another consumer facing the same options with a different set of needs and preferences use the same criteria to group them? Clearly, many ways of grouping are possible, yet it is difficult to know a priori the bases for such categorizations. Nonetheless, from a product development, positioning, and targeting perspective, predicting how various options will be perceived and categorized by different consumers is critical for marketers.

In general, perceptions of similarity (Rosch and Mervis 1975; Tversky 1977) on the basis of feature identification (Bettman and Sujan 1987; Felcher et al. 2000) are instrumental in the consumer categorization process. The categorization process involves the identification and comparison of products and allows consumers to maximize the information that they gather while minimizing the cognitive effort they spend in the information acquisition process (Cohen and Basu 1987). In this respect, categorization is central to product evaluations (Sujan 1985). The evaluation of a product is often based on category expectations, and categories also influence the inferences that consumers make about a product's features and performance (Gregan-Paxton et al. 2005). From this perspective, brand extension judgments (Boush and Loken 1991; Keller and Aaker 1992), product category assessments (Loken and Ward 1990), information search (Ozanne et al. 1992), ad evaluations and the motivation to process an ad (Goodstein 1993), and product-impression formation (Cohen and Basu 1987) are influenced by such categorization processes.

Consumers' reliance on category structures to evaluate products and product information makes the study of consumers' natural categorization process managerially relevant. Many marketing decisions rely heavily on consumers' knowledge structures (Gregan-Paxton and Roedder John 1997), which are functions of the categorization process. For instance, store lay out (Loken and Ward 1990; Stayman et al. 1992), product presentation and positioning (Gregan-Paxton et al. 2005; Sujan and Bettman 1989), packaging and labeling (Moreau et al. 2001), pricing (Chen et al. 2010), and advertising (Snyder 1992; Sujan and Dekleva 1987) decisions for both existing and new products, as well as brand extensions, must take into account how consumers perceive and categorize the options available to them to match their needs. When a new brand is introduced, the perceived similarity of this brand with existing known brands (i.e., the categorization of the new brand with other similar product offerings) impacts the inferences that the consumer will draw about the new brand (Snyder 1992). As such, understanding how consumers naturally categorize is a key to the successful implementation of marketing strategies.

Nonetheless, understanding the consumer categorization process is not at all straightforward. There is substantial heterogeneity in the ways consumers categorize, as each consumer may focus on different features when faced with the same items. Furthermore, consumers may use categories based on how the items relate to their individual needs and goals that are salient to them. Such goal-driven categories that are not based solely on the typical perceptual features of items (Barsalou 1983) are the main drivers of the heterogeneity in categorization. Within these goal-driven categories, the attributes consumers elicit to define a product may substantially differ from the ones that typically characterize the products (Ratneshwar and Shocker 1991). Personal and situational goals that are salient may contribute to perceived similarity in category representations (Ratneshwar et al. 2001), and the perceived similarity between products may increase as the perceived relations transcend intrinsic features (Jones and Love 2007). Consequently, goal-driven categories often span *across* typical product category boundaries (Loken et al. 2008). In brief, the observed heterogeneity in the categories activated can be based on salient goals or perceptual features and attributes. In addition to the heterogeneity in the composition of the internal category structures, there is also heterogeneity with respect to the factors that influence the activation of the categories. Individual differences and situational factors play an important role in consumers' category activation. For instance, chronically accessible (Fazio, and Dunton 1997) and situationally activated attitudes (Smith et al. 1996), positive affect (Isen 1984), age (John and Sujan 1990), cultural background (Jain et al. 2007), and experience (Sujan and Dekleva 1987) are found to influence the categories consumers activate. Thus, there exists significant evidence that consumers differ on the categories that they activate depending on their attitudes, goals, affective states, and individual differences.

Given the heterogeneity in the composition and activation of categories and the importance of the categorization process for both consumers and marketers, how consumers naturally form and store categories has been studied using different procedures (c.f. Coxon 1999). For instance, categorization has been studied by prompting participants to explicitly think of attributes that differentiate products (Viswanathan and Childers 1999). Similar studies have asked participants to label the categories that they form, and/or asked participants to judge the appropriateness of various products for different usage situations (Ratneshwar and Shocker 1991). Despite the useful insights they provide, these approaches are not distinct as they are suggestive of the categories that consumers will use a priori. More specifically, such methodologies that require participants to group stimuli upon a predetermined set of categories assume that consumers have a specific predetermined category structure. However, many categories are constructed as needed, extemporaneously, and may not reflect the categories that exist in one's memory (Barsalou 1985). A shortcoming of such methods is that requiring individuals to label categories prompt the use of simple rule-based categories which can potentially hinder the accessibility of affective or holistic categories, while focusing on usage contexts may lead to neglecting the perceptual aspects involved. Hence, studies that use a predetermined set of existing categories are confirmatory as opposed to exploratory in nature, making them appropriate only if the categories activated are known with great confidence.

Sorting tasks (also called free-sorting) have been used extensively to study categorization. These tasks impose less structure upon participants and allow natural categories to emerge (e.g., Evans and Arnoult 1967; Coxon 1999). The traditional sorting task requires participants to allocate a set of stimuli into categories that they construe based on their *own* perceptions of the stimuli. This procedure provides the researcher with a pairwise binary pairing or co-occurrence of the stimuli being studied (e.g., one if consumer i places product j and k in the same pile, 0 otherwise) and has several advantages: It is easy to perform (young children can do it), and it is reflective of natural mental activity (Coxon 1999). The sorting task method for collecting data on the perceptions of categories is particularly helpful when participants face a large number of alternatives (Rao and Katz 1971) and is well-suited to explore category perceptions (DeSarbo et al. 1991; Takane 1980; Coxon 1999). It is thus a conventional tool to study the inherent use of categories.

Despite the advantages involved in using the sorting task, analyzing sorting data can be difficult to properly assess as the vast majority of existing analytical procedures typically require aggregation which obscures the heterogeneity in the ways that consumers categorize. Daws (1996) suggested that there are two types of sorting task (aggregate) analyses: (1) those that aggregate over individuals and (2) those that aggregate over objects. In the former case, the number of times two objects are grouped together is used as a proxy for similarity such that the more consumers paired two objects/products together (aggregating over consumers) the more similar the objects/products are assumed to be. Johnson (1967) employed a product by product count matrix as input for hierarchical clustering analyses and interpreted the clusters as categories. With Daws (1996), *individuals* are compared based on the similarity of their categorizations. This requires obtaining an aggregate similarity measure between pairing matrices such that the resulting partitions can be compared. The resulting similarities are often used in multidimensional scaling techniques (Kruskal 1964; Takane 1980). However, in all these cases, heterogeneity is lost with this aggregation.

In this paper, we propose a new methodology that can be used to identify heterogeneous latent category structures using sorting task data without using arbitrary aggregations or ad hoc conversions into similarities or dissimilarities. We extend the literature on two-way segmentation models that simultaneously group consumers and variables either by permutations of rows and columns (e.g., Hartigan 1972; Green et al. 1973; Eckes and Orlik 1993), by ultra-metric structures (e.g., DeSarbo and De Soete 1984; De Soete et al. 1984; Ramaswamy and DeSarbo 1990) or by additive clustering (e.g., Shepard and Arabie 1979; Carroll and Arabie 1983). In this paper, we generalize the DeSarbo et al. (1991) individual level categorization model based on additive clustering to a latent structure context, given the statistical difficulties associated with their approach and the restrictions on the model parameters that may obscure the underlying structure of the categories consumers use. We follow the previous literature suggesting that individuals sort large sets of stimuli based on an underlying similarity judgment (Goldstone 1994; Tversky 1977), which are unobserved and inferred from the sorting task data (DeSarbo et al. 1991; Hampton 1998). We propose two brands to be members of the same unobserved category if they are judged to be similar beyond some unobserved threshold (sufficient level). We further argue that this unobserved similarity is a function of the objects' memberships in unobserved categories, but only if those unobserved categories are actually activated. We model heterogeneity at the group level, allowing researchers to go beyond aggregate sample level generalizations and avoid aggregation issues, while still providing insight into which unobserved categories are more intuitive than others. In the next section, we detail the new procedure and the estimation algorithm. As an illustration, we conduct a study that entails an analysis of sorting data of various sports and examine how groups of consumers naturally differ in their perceptions. Finally, we present some directions for future research.

2 The proposed latent structure category extraction procedure

Let:

i = 1, …, I consumers
j,k = 1, …, N stimuli (e.g., brands)
r = 1, …, R unknown latent categories
g = 1, …, G unknown consumer groups (e.g., market segments).

We observe the following:

$$\delta_{\mathrm{ijk}} = \begin{cases} 1 \text{ if consumer } i \text{ places stimuli } j \text{ and } k \text{ in the same pile,} \\ 0 \text{ otherwise,} \end{cases}$$

and use these observed categorization judgments obtained from a sorting task to provide us with insight into the latent category structure, as well as the extent to which groups of consumers differ in activating those structures. To accommodate heterogeneity in the category structures, we propose that the observed categorization behavior when forming the piles can be expressed as a function of unobserved similarity judgments of consumers who use similar latent categories. As such, if we assume the presence of G unobserved groups of consumers who have similar categorization structures, we can define the latent similarity between stimulus j and k as perceived by consumer i, conditional on membership in group g, as:

$$S_{\mathrm{i,j,k|g}} = s_{\mathrm{i,j,k|g}} + e_{\mathrm{i,j,k|g}}, \tag{1}$$

where:

$$e_{\mathrm{i,j,k|g}} \overset{\mathrm{iid}}{\sim} N\left(0, \sigma_g^2\right). \tag{2}$$

We now define:

$$\begin{aligned} P\left(\delta_{\mathrm{i,j,k|g}} = 1\right) &= P\left(S_{\mathrm{i,j,k|g}} \geq t_g\right) \\ &= P\left(e_{\mathrm{i,j,k|g}} \geq t_g - s_{\mathrm{i,j,k|g}}\right) \\ &= 1 - \Phi\left(\frac{\left(t_g - s_{\mathrm{i,j,k|g}}\right)}{\sigma_g}\right), \end{aligned} \tag{3}$$

where Φ(.) is the standard normal cumulative distribution function. This indicates that the more similar two stimuli are perceived, the more likely they are to be grouped together in the same pile.

As in DeSarbo et al. (1991), we propose that the perceived similarity is a function of the degree to which the stimuli are prototypical of similar latent categories and to the extent that those categories are activated by consumers when engaging in the categorization task. We now introduce p_{jr} as the degree to which stimulus j is a member of latent category r (each p_{jr} is bounded between 0 and 1), and w_{gr} as the degree to which the gth group of consumers activates the latent category r when engaging in the categorization task (w_{gr} is restricted to positive values). Overall, our proposed similarity model is as follows:

$$\mathrm{s}_{\mathrm{i,j,k|g}} = \sum\nolimits_{r=1}^{R} w_{gr} p_{jr} p_{kr}. \tag{4}$$

This formulation suggests that stimuli j and k are perceived as similar to consumer i belonging to group g if s/he judges *both* j and k to be typical of similar categories, considering that the categories have been activated by the consumer. This follows existing categorization literature that suggests that consumers resort to the selective weighting (attention) of category dimensions when engaging in a categorization task (Medin and Schaffer 1978; Nosofsky 1986). Furthermore, this decomposition of the similarity into a tri-product of activation (weight/salience) and the typicality of stimuli resemblance have been useful in modeling similarity and the categorization process (see Carroll and Arabie 1983; Shepard and Arabie 1979; DeSarbo et al. 1991; Carroll and Winsberg 1995, and Chaturvedi and Carroll 2006 for justification of this trinary product in Eq. 4 for modeling similarities).

Substituting $s_{\mathrm{t|g,j,k}}$ in Eq. 4, we obtain a reformulation of Eq. 3 in terms of p_{jr} and w_{gr}:

$$P\left(\delta_{\mathrm{i,j,k|g}} = 1\right) = 1 - \Phi\left(\frac{\left(t_g - \sum_{r=1}^{R} w_{gr} p_{jr} p_{kr}\right)}{\sigma_g}\right)$$

and without loss of generalizability:

$$P\left(\delta_{\mathrm{i,j,k|g}} = 1\right) = 1 - \Phi\left(t_g - \sum\nolimits_{r=1}^{R} w_{gr} p_{jr} p_{kr}\right) \tag{5}$$

given that σ_g can be absorbed in the group threshold and weights parameters (i.e., σ_g is not identifiable[1]). Similarly:

$$P\left(\delta_{\mathrm{i,j,k|g}} = 0\right) = \Phi\left(t_g - \sum\nolimits_{r=1}^{R} w_{gr} p_{jr} p_{kr}\right) \tag{6}$$

Thus, for consumer i belonging to group g, the conditional likelihood is given by:

$$L_{\mathrm{i|g}} = \prod\nolimits_{i>k}^{N} \prod \left(1 - \Phi(\cdot)\right)^{\delta_{\mathrm{i,j,k|g}}} \Phi(\cdot)^{1-\delta_{\mathrm{i,j,k|g}}} \tag{7}$$

where (.) denotes $\left(t_g - \sum_{r=1}^{R} w_{gr} p_{jr} p_{kr}\right)$ for simplicity. This is merely an accounting of the instances two brands are placed in the same pile (or not) by a given consumer in a sorting task. Given an independent sample of consumers, the complete likelihood is given by:

$$L = \prod_{i=1}^{I} \sum_{g=1}^{G} \lambda_g L_{\mathrm{i|g}} \tag{8}$$

$$\prod_{i=1}^{I} \sum_{g=1}^{G} \lambda_g \prod\nolimits_{j>k}^{N} \prod \left(1 - \Phi(\cdot)\right)^{\delta_{\mathrm{i,j,k|g}}} \Phi(\cdot)^{1-\delta_{\mathrm{i,j,k|g}}} \tag{9}$$

[1] There is also an indeterminacy between w_{gr}, p_{jr}, and p_{kr} as all three terms are indexed by R. To resolve this issue, we normalize latent category memberships such that $\max\left(p_{jr}\right) = 1, \forall r = 1 \ldots R$. This is equivalent to requiring that every latent category has at least one item that is most prototypical.

where λ_g are the unknown mixing parameters such that $0<\lambda_g<1$ and $\sum \lambda_g = 1$. The complete log likelihood is:

$$Ln\,L = \sum_{i=1}^{I} \ln\left[\sum_{g=1}^{G} \lambda_g \prod\nolimits_{j>k}^{N} \prod \left(1-\Phi(\cdot)\right)^{\delta_{\mathrm{i,j,k|g}}} \Phi(\cdot)^{1-\delta_{\mathrm{i,j,k|g}}}\right]. \tag{10}$$

Bishop (2006; pp. 444–446) proves how such Bernoulli mixtures permit non-zero covariances such that pure independence is not required. Note that we impose the following constraints in the estimation: (1) $w_{gr}\geq 0$, (2) $t_g\geq 0$, (3) $0\leq p_{jr}\leq 1$, (4) $0<\lambda_g<1$, and (5) $\sum \lambda_g=1$. The non-negativity constraints for t_g and w_{gr} are implemented by substituting squared quantities for these parameters (e.g., substituting t_g^2 and w_{gr}^2 for their respective unsquared terms in the likelihood expression (see Gill et al. 1981, pp. 268–269). To ensure that $0\leq p_{jr}\leq 1$, we reparameterize each p_{jr} by $\exp\left(q_{jr}\right) / \left(1+\exp\left(q_{jr}\right)\right)$ and estimate q_{jr}.[2]

This specification allows for different categorization structures for each group (equivalent to R_g), and for different perceptions of stimuli as belonging to different categories. For example, given that $\underline{\boldsymbol{P}}$ represents a total set of latent categories used across the sample, the procedure can estimate w_{gr} much smaller than t_g when category r (potentially zero) is not used or activated by group g, whereas it will be larger than t_g if there is evidence that the group uses or activates it. This can also accommodate groups perceiving category memberships differently. For instance, assume that members of group 1 do not see stimulus j as a member of category r, but members of group 2 do. Then, the procedure would identify two categories: one without stimulus j as a member (that group 1 activates) and another with stimulus j as a member (that group two activates). Furthermore, because no restriction is made regarding the membership of stimuli to one and only one category, the latent categories can also be used to represent different levels of abstraction where some categories are subsets of others.

As with other finite mixture-based latent structure models, within any iteration, one can estimate the posterior probabilities of consumer i belonging to group g via:

$$v_{ig} = \frac{\lambda_g L_{i|g}}{\sum_{g=1}^{G} \lambda_g L_{i|g}} \tag{11}$$

conditioned on the current estimates of the model parameters. Thus, given only the sorting data (δ_{ijk}) and trial values for R and G, our proposed model *simultaneously* estimates: (1) the number of latent categories utilized by the complete sample of consumers (R^*), (2) the number of latent groups of consumers that activate different patterns of latent categories (G^*), (3) the size of these latent consumer groups (λ_g), (3) the probabilistic membership of each consumer in each derived group ($\underline{\boldsymbol{V}}=((v_{ig}))$), (4) the membership probabilities of each object in each latent category ($\underline{\boldsymbol{P}}=((p_{jr}))$), (5) the extent to which each derived group use or activate each derived latent category in the evocation of its categorization judgments ($\underline{\boldsymbol{W}}=((w_{gr}))$), and (6) the group-level activation thresholds ($\underline{\boldsymbol{t}}=(t_g)$). This analysis is repeated for $g=1\ldots$G

[2] See Dayton and Macready (1988) for the use of this reparameterization in constrained latent class models.

groups and r=1…R latent categories where a variety of model selection heuristics are examined to determine the numbers of latent categories and groups. The technical details of the estimation procedure and model selection heuristics utilized are provided in the Web Appendix.

Note, given that there is no requirement that the p_{jr}'s sum to one (or any other constant value), one can inspect $\underline{\boldsymbol{P}}$ to examine stimuli that are categorized in multiple categories across groups. Extant literature in psychology and consumer behavior has shown that consumers may perceive brands as belonging to different categories depending on contextual factors and individual states (Gregan-Paxton et al. 2005; Moreau et al. 2001; Ross and Murphy 1999). Heterogeneity in the categorization process is represented by G^* and $\underline{\boldsymbol{W}}$. In particular, the cardinality of G^* renders insight into the magnitude of response heterogeneity (i.e., the larger the number of groups required, the more heterogeneous the categorizations are *across* the sample). $\underline{\boldsymbol{W}}$ reflects the heterogeneity in the patterns of activations for the derived groups using the latent categories activated across the sample. In particular, $\underline{\boldsymbol{W}}$ indicates the different categorizations activated by each of the estimated groups.

As mentioned earlier, our proposed methodology generalizes the DeSarbo et al. (1991) model which estimates individual-level activations. While the DeSarbo et al. (1991) individual-level model accommodates heterogeneity with respect to such categorization processes, their maximum likelihood estimation procedure suffers from the presence of incidental parameters whose order varies by sample size as there are a huge number of individual level parameters to estimate. As a result, the derived maximum likelihood parameter estimates are inconsistent. Our latent structure generalization does not suffer from this problem. Additionally, the formulation by DeSarbo et al. (1991) enforces the following constraint: $\sum_{r=1}^{R} p_{jr} = 1, \forall j$, which overly restricts the nature of membership structures permitted. Finally we note that just as with DeSarbo et al. (1991), the likelihood of the proposed model also includes structural dependencies such that if Y_{ijk}=1 and Y_{ijm}=1, then Y_{imk}=1 due to the structure of the traditional sorting task in which items belong to one and only one pile. A synthetic data example that illustrates how the methodology can capture various category structures is presented in the Web Appendix.

3 A sports marketing illustration: categories of sports

To demonstrate the usefulness and relevance of our proposed methodology, we investigate the ways in which consumers categorize sports. According to the 2008 annual survey of the *Sports Business Journal*, the Sports business industry is a $213 billion industry which is more than twice the size of the US auto industry (Sports Business Journal 2009) and over seven times as large as the movie industry. Approximately 30% of the expenditures are related to advertising and sporting good purchases (each approximately 15% of the industry spending). Despite the magnitude of the sports industry, there is a paucity of academic research attention directed to it in Marketing.

Furthermore, little is known about the ways that consumers perceive and categorize sports. One notable exception is Koivula (2001) who asked participants to evaluate some 41 sports with respect to 117 different characteristics. Using a principal component analysis, Koivula (2001) found 12 factors that represent different types of sports. This particular study brings up the question: What would have happened if participants were not prompted with any attributes or characteristics of sports prior to the categorization task? Asking participants to evaluate a list of sports on a set of characteristics can lead to a category structure such that some of the characteristics or situations presented would not have been considered if the participants were not prompted with them. Additionally, their task is highly taxing (a simple calculation based on the description of the procedure suggests that participants each made 4,797 Likert-type scale evaluations of sports in the study); thus, the categorization process may have been ill affected by the cumbersomeness of the task.

3.1 Our study

We employed the Sports Manufacturing Goods Association's (SMGA) yearly US sport participation reports to identify a preliminary list of sports that we could request participants to categorize. The objective was to create a list of sports that participants were familiar with and were representative of the different forms of such recreational and athletic activities. In addition, the list of sports we created needed to be categorized in a relatively short amount of time and to allow participants to exhibit the variation in their category structures. Consequently, we used the list from the SMGA (113 sports) and asked 12 undergraduate students with different backgrounds to review the list of sports and indicate how much they knew about the listed sports. Based on the feedback from these informants, we selected 50 sports that would be relevant and well understood by our participants. The 50 sports selected for use in our study are presented in Table 1.

The names of the 50 sports were placed on flash cards, one sport per card. Upon entering the research lab, participants were asked to sit in front of a computer and were given an envelope containing the randomly shuffled flash cards. The cards were shuffled and hence randomly ordered such that the order of the cards would not bias the categorization process. First, participants were instructed to sort the cards into as many or few piles as they desired based on their own perceptions of similarities of the sports. Participants were allowed to form their own piles as the extant literature has demonstrated that objects can belong to different categories for

Table 1 List of selected sports

Aerobics	Boxing	Hunting	Rugby	Swimming
Archery	Canoeing	Ice hockey	Running	Table tennis
Badminton	Cardio kick boxing	Ice skating	Sailing	Tennis
Baseball	Climbing	Kayaking	Scuba diving	Track and field
Basketball	Field hockey	Lacrosse	Skateboarding	Triathlon
Bicycling	Fishing	Mountain biking	Snorkeling	Volleyball
Billiards	Football	Pilates	Soccer	Wakeboarding
BMX	Gymnastics	Racquetball	Softball	Weight lifting
Boardsailing	Hiking	Roller skating	Squash	Wresting
Bowling	Horseback riding	Rowing	Surfing	Yoga/Tai chi

different consumers (Murphy and Ross 1994; Moreau et al. 2001; Gregan-Paxton et al. 2005; Lajos et al. 2009). Upon completion of this sorting task, participants responded to a questionnaire which asked them to indicate the composition of each pile they made. After reporting the composition of the piles, participants were asked to describe how each of their piles were formed, one category at a time. Next, they responded to questions about their knowledge related to these 50 sports. This part of the questionnaire was adopted from Brucks (1985) and asked participants to compare their knowledge of each sport to the knowledge of an average person on a 1–7 scale, where 1 indicated "I know a lot less" and 7 indicated "I know a lot more. Participants also provided reports of interest and participation in each of the sports, as well as the extent to which they follow sports in the media (all within the last year). Finally, they responded to various psychographic, demographic, and mood questions. One hundred seventy-two undergraduate students (52% male and 48% female) completed the study for course credit. The participants were undergraduate students at a large northeastern university enrolled in an introductory class, and 9% of the participants were non-native English speakers. The categorization task on average took 33 min to complete. All participants completed the entire study in less than 45 min.

3.2 The results

We performed the analysis for $G=1\ldots8$ groups and $R=1\ldots8$ latent categories. Using the variety of model selection heuristics discussed in the Web Appendix, we selected the $G=6$ groups, $R=6$ latent categories solution with a common threshold across groups as the most parsimonious representation of the structure in the data. The estimated latent category structure ($\underline{\boldsymbol{P}}$) and activations ($\underline{\boldsymbol{W}}$) for $G=6$ and $R=6$ are shown in Tables 2 and 3.

In examining the estimated $\underline{\boldsymbol{P}}$ matrix in Table 2, we can obtain valuable insight into the derived latent categories. We interpret the first category as *fitness sports*. The variability in the probabilities for this category suggests that the participants varied in how they viewed the sports in relation to this category. For instance, sports typically performed at a gym (e.g., aerobics, Pilates, yoga/tai chi, cardio kick-boxing) were of the highest membership to the category. For sports such as running, swimming, boxing, and triathlon that also involve a high degree of fitness, participants varied in whether they viewed them as typical members as their probabilities of membership in this particular category were lower in magnitude. They were still considered members of the category, but not as highly typical.

We label the second category as *water sports*. Members of the category are sports that are associated with water such as boardsailing, rowing, snorkeling, wakeboarding, kayaking, surfing, etc. This category also includes sports performed in the water (swimming, snorkeling, scuba diving), sports that require using boards on the water (surfing, wakeboarding, boardsailing), and sports that involve something to sit on while in water (kayaking, canoeing).

The third category involves *sports that are practiced outdoors*. It includes sports that are considered nature activities such as climbing, hunting, hiking, and also sports that are traditionally conducted outside (biking, archery, and various water

Table 2 Derived latent sport categories (estimated $\underline{P}$ matrix)

	Category 1, fitness sports	Category 2, water sports	Category 3, outdoors	Category 4, team sports	Category 5, recreational activities	Category 6, motion sports
Aerobics	1.00	0.00	0.00	0.00	0.00	0.00
Archery	0.00	0.00	0.91	0.00	0.50	0.00
Badminton	0.00	0.00	0.00	0.31	1.00	0.01
Baseball	0.00	0.00	0.00	1.00	0.00	0.00
Basketball	0.00	0.00	0.00	1.00	0.00	0.00
Bicycling	0.49	0.00	0.52	0.00	0.00	0.63
Billiards	0.00	0.00	0.37	0.00	0.88	0.01
BMX	0.11	0.00	0.49	0.00	0.00	0.98
Boardsailing	0.00	0.99	0.47	0.00	0.00	0.25
Bowling	0.00	0.00	0.37	0.23	0.78	0.02
Boxing	0.54	0.00	0.00	0.49	0.01	0.00
Canoeing	0.00	0.85	0.79	0.00	0.00	0.00
Cardio kick boxing	1.00	0.00	0.00	0.00	0.00	0.00
Climbing	0.39	0.00	1.00	0.00	0.00	0.30
Field hockey	0.00	0.00	0.00	0.93	0.01	0.00
Fishing	0.00	0.36	0.96	0.00	0.30	0.00
Football	0.00	0.00	0.00	1.00	0.00	0.00
Gymnastics	0.59	0.00	0.01	0.47	0.16	0.24
Hiking	0.44	0.00	1.00	0.00	0.01	0.20
Horseback riding	0.27	0.00	0.90	0.00	0.41	0.20
Hunting	0.00	0.00	0.99	0.00	0.36	0.01
Ice hockey	0.00	0.00	0.00	0.85	0.00	0.01
Ice skating	0.30	0.00	0.29	0.29	0.38	0.44
Kayaking	0.00	0.85	0.78	0.00	0.00	0.03
Lacrosse	0.00	0.00	0.00	0.98	0.00	0.00
Mountain biking	0.32	0.00	0.83	0.00	0.00	0.56
Pilates	1.00	0.00	0.00	0.00	0.00	0.00
Racquetball	0.00	0.00	0.00	0.39	0.88	0.01
Roller skating	0.21	0.00	0.38	0.00	0.45	0.72
Rowing	0.21	0.81	0.53	0.24	0.00	0.10
Rugby	0.00	0.00	0.00	0.86	0.13	0.00
Running	0.78	0.00	0.14	0.00	0.00	0.47
Sailing	0.00	0.97	0.65	0.00	0.00	0.01
Scuba diving	0.00	0.99	0.61	0.00	0.00	0.02
Skateboarding	0.00	0.00	0.44	0.00	0.24	1.00
Snorkeling	0.00	0.97	0.62	0.00	0.00	0.01
Soccer	0.00	0.00	0.00	1.00	0.00	0.00
Softball	0.00	0.00	0.00	0.97	0.01	0.00
Squash	0.00	0.00	0.00	0.39	0.86	0.00
Surfing	0.00	1.00	0.39	0.00	0.00	0.46
Swimming	0.56	0.59	0.00	0.38	0.00	0.27
Table tennis	0.00	0.00	0.01	0.19	1.00	0.01
Tennis	0.00	0.00	0.00	0.72	0.49	0.00
Track and field	0.50	0.00	0.01	0.56	0.00	0.48

Table 2 (continued)

	Category 1, fitness sports	Category 2, water sports	Category 3, outdoors	Category 4, team sports	Category 5, recreational activities	Category 6, motion sports
Triathlon	0.65	0.00	0.20	0.20	0.00	0.47
Volleyball	0.00	0.00	0.00	0.94	0.15	0.00
Wakeboarding	0.00	0.98	0.38	0.00	0.00	0.49
Weight lifting	0.83	0.00	0.00	0.19	0.00	0.03
Wresting	0.42	0.00	0.00	0.64	0.00	0.00
Yoga/Tai chi	0.99	0.00	0.00	0.00	0.01	0.00

sports). Some sports that were regarded as water sports were less typical of outdoor sports such as swimming which can be practiced indoors or outside. Furthermore, one notices that groups typically have either the water sports category or the outdoor sports category salient (negative correlation of −0.67). This can be explained as many outdoor sports are also water sports and that participants were forced to make tradeoffs between the categories because of the structure of the sorting task.

The fourth category is *team sports*. Highly prototypical members were mostly popular team sports visible in the media such as basketball, baseball, softball, volleyball, football, ice hockey, and rugby. Some sports in this category evoke the concept of a "team" as a group, such as the tennis team, the wrestling team, the swim team, etc. Given their lower probabilities of membership, however, individuals varied in their conceptualization of a team from conceptual to feature based.

The fifth category involves sports that are perceived as *recreational activities*. This category includes sports like table tennis, bowling, billiards, and badminton that are often considered activities rather than sports. Some of the racquet sports like tennis and squash are seen as less representative as participants have categorized tennis as a team sport.

Finally, the sixth category is *motion sports*. Sports in this category involve sports related to bicycles (BMX, bicycling, mountain biking), sports that involve the use of a board or skates (boardsailing, ice skating, roller skating, skateboarding, wakeboarding),

Table 3 Derived group level category activations (estimated $\underline{W}^2$ matrix)

Group	Category 1, fitness sports	Category 2, water sports	Category 3, outdoor sports	Category 4, team sports	Category 5, recreational activities	Category 6, motion sports
1	29.12	16.33	20.19	44.00	24.59	22.65
2	27.84	14.53	31.64	28.43	23.57	24.74
3	27.30	48.39	10.60	25.49	23.66	22.84
4	59.38	23.46	18.27	30.79	30.97	26.82
5	28.76	33.11	20.36	47.69	27.20	22.07
6	23.19	17.28	15.15	23.30	24.03	20.81

Note: The estimated squared threshold t^2 =21.67. As squared activations increase past 21.67, the more likely it is that a group activates the latent category.

or to a lesser extent, fitness activities that involve a large component of leg strength and endurance (gymnastic, hiking, horseback riding, running, track and field, triathlon).

The mixing parameters λ_g provide estimates of group sizes in this dataset (group 1, 9.93%; group 2, 26.16%; group 3, 19.00%; group 4, 6.41%; group 5. 23.37%; group 6, 15.13%). How did participants in different groups vary in their category activations? In the following section, we illustrate the heterogeneity in categorization structure activations of the six groups using significant demographic, experience, attitudinal, and psychographic variables. We compare the estimated matrix $\underline{\boldsymbol{W}}$ of category activations per group to the estimated common threshold parameter (presented in Table 3) as an indication of how likely members of the groups are to activate a particular category. We use significance tests (not shown) to assess the mean differences between groups (for all results reported, $p<.10$) with respect to these background variables collected in our study.

Of the participants who are members of the first group, 65% are females. The highest activated category by Group 1 is *team sports*. In comparison to members of other groups, they are *less* interested in activities that are *not* physical. This was evidenced by their reported lack of subjective knowledge about bowling, fishing, hunting, sailing, and table tennis, and with their reported lack of experience with such activities. In addition to this, they are more likely on average than other group members to agree with the statements that “My friends would consider me athletic,” “Exercise is very important to me,” “I enjoy playing sports,” “Sports had been an important aspect in my family when I grew up,” “I played many varsity sports in high school,” and “I work out every week”. Accordingly, we label this group as “*Athletic Females*”.

When asked to indicate whether they believed they know more or less than other students about each of the sports, members of Group 2 reported that they feel that they are less knowledgeable than others about aerobics, cardio kickboxing, gymnastics, basketball, field hockey, rowing, softball, swimming, and volleyball. They also reported having less experience in many of the sports. This self-reported lack of knowledge about the sports is reflected in their category structures by the activation the outdoor sports category in contrast to the water sports category (the latter being at a lower level of abstraction). Hence, we label members of this group as “*Sports Novices*”.

For members of Group 3, one distinctive aspect is that they had the highest activation for the water sports category. Yet, members of this group are much less interested in water sports and practice them less (e.g., canoeing, fishing, and swimming). It is thus possible that these individuals’ permanent dislike for water-based sports makes the category highly salient to them (Fazio and Dunton 1997). We therefore label members of this group as “*Water Averse*”.

Members of Group 4 have much higher salience for the fitness sports category. Members of this group differ from others in that they report being more likely to “ignore professional sports on TV”, although they report participating in sports just as much as others. For individuals who may not be interested in the entertainment aspects of sports, the fitness aspect is likely to be salient. Furthermore, considering that they reported participating in recreational activities just as much as others, their focus on fitness is likely to make it apparent that these recreational activities do not involve much fitness (hence recreational). It is probably salient to members of this group that recreational activities lack a fitness component, and they consequently wanted to differentiate those sports from others. We label members of this group as the “*Non-sports Fan*”.

In comparison to members of other groups, the members of Group 5 reported being more careful during the card sorting task, and they feel that they are more familiar with gymnastics, weightlifting, Pilates, yoga, and cardio kickboxing compared with others. Additionally, they have less experience with sports exercised with others such as football, rugby, wrestling, or rowing. They also agree less with the following statements: "I attended parties every weekend," "school is stressful to me," "I enjoy playing sports," "I am a competitive person," "I have a favorite team in every major sport". Thus, we can explain their high activation in the team sports category by their lack of interest in sports for socializing, making the team sports particularly salient. Hence, we label the members of this group as "*Exercisers*".

Finally, members of the Group 6 agreed less with the statements "I rarely get sick" and more with the statements "In comparison to my classmates, I tend to be more stressed about my grades and studies" and "School is stressful to me" than members of other groups. They also reported feeling more knowledgeable compared with others about activities such as bowling, roller-skating, sailing, and tennis and attended more events related to "activities" type sports. Also, they report having less experience than others with some of the rougher sports like hockey and rugby. They also have consistently lower activations for many of the categories, suggesting their familiarity with sports. Therefore, we label this group as "*Recreationals*".

Obviously, our student convenience sample is not representative of the US sports market place. The major contribution of this manuscript is methodological in terms of introducing a new procedure for representing heterogeneity in consumer categorization judgments. Were this study to be replicated with a national representative sample of consumers from the entire US sports market place, several managerial benefits could be derived from such analyses. One, the derived groups of consumers can represent market segments which can be profiled and examined for targeting. The estimated mixing proportions can provide approximations as to the substantiality of such market segments, and the estimated posterior probability of memberships can be utilized to profile these derived segments when crossed with interest and activity by sport, demographics, psychographics, amount spent on sports and sports-related merchandise, etc. In this manner, target segments can be derived depending upon the nature of the client business (e.g., Nike). Two, our analyses derives the basis on which consumers naturally organize sports in their mind via the $\underline{\boldsymbol{P}}$ matrix. And this matrix renders important relationships between the various sporting activities that can also provide substantial managerial insight. Sporting activities that are prototypical of a latent category can be jointly marketed to in terms of respective target segments that readily activate such latent categories. Symbiotic marketing programs can be jointly devised across different sporting events most representative of the same latent category. For example, broadcasting commercials for roller-skating equipment during a television cycling tour may seem natural to viewers and result in a transfer of the positive associations of cycling to the advertised roller-skating equipment. This is particularly relevant for manufacturers of multiple sports equipment and/or clothing (e.g., Champion). Finally, such analyses may point to areas for brand extensions into new areas of sport where such interrelationships exist and may offer a mechanism to expand current markets into new customer bases and sporting activities.

4 General discussion

In this paper, we have modeled consumer categorizations on the basis of sorting task judgments collected using a free-sorting dataset of sports and proposed a new methodology for the identification of the unobserved categories and the differences in the activation of those categories across consumer groups. In our empirical study, we demonstrated the presence of six unobserved categories which exhibited a graded structure and differed based on whether they represented taxonomic features with distinct levels of abstraction (e.g., *water sports* and *outdoors sports*), were related to goals (e.g., *fitness sports*), or were at a higher level of abstraction (e.g., *motion sports, recreational activities*). We found six groups of individuals that showed diverse patterns of activation or salience over the unobserved categories, providing evidence for a significant amount of heterogeneity. The salience of the categories was related to participants' preferences, goals, experiences, and subjective knowledge of the categories.

We provided a new empirical methodology to explore consumers' categorization structures in an unsupervised manner. By providing evidence that individuals differ in the categories they activate, we contribute to the growing research on the consumers' categorizations of brands and products. This is particularly important given that earlier categorization researchers have specified the categories a priori for their experiments, which may not have resulted in the categories that consumers would actually activate. Our methodology can be employed to explore these latent category structures and to ensure that the unique ways in which consumers may categorize are taken into account.

Given that the methodology presented reveals the latent bases for categorization (i.e., we do not ask for rules, labels or examples), it may allow marketing managers to understand the 'true' categories their target consumers may activate when faced with products and to evaluate the salience of various categories for different consumer segments. Hence, the current model can be particularly useful when managers are facing uncertainty with regards to the differences among consumer segments in terms of the categories they retrieve in evaluating existing or new products. This can be strategically important throughout the marketing management process: Our methodology may facilitate more effective segmentation and targeting as it allows differentiating among consumer groups based on their categorization process, which reflects their salient goals, preferences, and experiences. In light of this knowledge, one can better succeed in defining the value proposition and designing the marketing mix strategies. Our methodology can be employed together with other exploratory techniques (e.g., focus groups) to identify consumer differences with respect to the categories they activate when faced with products.

Consistent with the categorization theory, the interpretation of the latent categories identified through this methodology may be idiosyncratic. To extend this approach and facilitate interpretation, one can reparametrize the $\underline{\boldsymbol{P}}$ matrix such that the typicality of an object to the latent category is a function of attributes, features, and/or marketing mix. Similarly, future research can also be directed towards reparameterizing $\underline{\boldsymbol{W}}$ as a function of relevant individual background variables, need states, goal orientations, etc. Although we argue that the value of our method lies in its exploratory nature, adapting the methodology to test the importance of specific attributes, relevant goals, or individual differences more explicitly and a priori (through a reparametrization of both the $\underline{\boldsymbol{W}}$ and $\underline{\boldsymbol{P}}$ matrices) may be useful to researchers interested in testing the importance of specific features in the natural categorization of individuals (e.g., size, color, brand). Such reparameterization features would also provide a mechanism for performing predictive validation. If one is further interested in comparing group activations and structures for a priori defined individual groups, an "external analysis" may be conducted by fixing $\underline{\boldsymbol{V}}$ (and thus the mixing proportions), and estimating the remaining parameters in the M-Step ($\underline{\boldsymbol{W}}$, $\underline{\boldsymbol{P}}$, and $\underline{\boldsymbol{t}}$).

References

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11*(3), 211–227.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 11*(4), 629–654.

Bettman, J. R., & Sujan, M. (1987). Effects of framing on evaluation of comparable and noncomparable alternatives by expert and novice consumers. *Journal of Consumer Research, 14*(2), 141–154.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Boush, D. M., & Loken, B. (1991). A process-tracing study of brand extension evaluation. *Journal of Marketing Research, 28*(1), 16–28.

Brucks, M. (1985). The effects of product class knowledge on information search behavior. *Journal of Consumer Research, 12*(June), 1–16.

Carroll, J. D., & Arabie, P. (1983). INDCLUS: An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika, 48*(2), 157–169.

Carroll, J. D., & Winsberg, S. (1995). Fitting an extended INSCAL model to three-way proximity data. *Journal of Classification, 12*, 57–71.

Chaturvedi, A., & Carroll, J. D. (2006). CLUSCALE: A three-way hybrid model incorporating overlapping clustering and multidimensional scaling structure. *Journal of Classification, 23*, 269–299.

Chen, Y., Iyer, G., & Pazgal, A. (2010). Limited memory, categorization, and competition. *Marketing Science, 29*(4), 650–673.

Cohen, J. B., & Basu, K. (1987). Alternative models of categorization: Toward a contingent processing framework. *Journal of Consumer Research, 13*(4), 455–472.

Coxon, A. P. M. (1999). *Sorting Data: Collection and Analysis*. Thousand Oaks, CA: Sage.

Daws, J. T. (1996). The analysis of free-sorting data: Beyond pairwise co-occurrence. *Journal of Classification, 13*(1), 57–80.

Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association, 83*, 173–178.

De Soete, G., DeSarbo, W. S., Furnas, G. W., & Carroll, D. J. (1984). The estimation of ultrametric and path length trees from rectangular proximity data. *Psychometrika, 49*(3), 289–310.

DeSarbo, W. S., & De Soete, G. (1984). On the use of hierarchical clustering for the analysis of nonsymmetric proximities. *Journal of Consumer Research, 11*(1), 601–610.

DeSarbo, W. S., Jedidi, K., & Johnson, M. D. (1991). A new clustering methodology for the analysis of sorted or categorized stimuli. *Marketing Letters, 2*(3), 267–279.

Eckes, T., & Orlik, P. (1993). An error variance approach to two-mode hierarchical clustering. *Journal of Classification, 10*(1), 51–74.

Evans, S. H., & Arnoult, M. (1967). Schematic concept formation: Demonstration in a free sorting task. *Psychonomic Science, 9*(4), 221–222.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology, 33*(5), 451–470.

Felcher, E. M., Malaviya, P., & McGill, A. L. (2000). Taxonomic versus goal-derived product categorization: Effects on within and across category choice processes. *Psychology and Marketing, 18*(8), 865–887.

Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical Optimization*. New York: Academic.

Goldstone, R. L. (1994). The role of similarity in categorization: Providing groundwork. *Cognition, 52*(2), 125–157.

Goodstein, R. C. (1993). Category-based applications and extensions in advertising: Motivating more extensive ad processing. *Journal of Consumer Research, 20*(1), 87–99.

Green, P. E., Wind, Y., & Jain, A. K. (1973). Analyzing free-response data in marketing research. *Journal of Marketing Research, 10*(February), 45–52.

Gregan-Paxton, J., & Roedder John, D. (1997). Consumer learning by analogy: A model of internal knowledge transfer. *Journal of Consumer Research, 24*(3), 266–284.

Gregan-Paxton, J., Hoeffler, S., & Zhao, M. (2005). When categorization is ambiguous: Factors that facilitate the use of a multiple category inference strategy. *Journal of Consumer Psychology, 15*(2), 127–140.

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition, 65* (2–3), 137–165.

Hartigan, J. A. (1972). Direct clustering of a data matrix. *Journal of the American Statistical Association, 67*(337), 123–129.

Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Handbook of Social Cognition*. Hillsdale, NJ: Lawrence Erlbaum.

Jain, S. P., Desai, K. K., & Mao, H. (2007). The influence of chronic and situational self-construal on categorization. *Journal of Consumer Research, 34*(June), 66–76.

John, D. R., & Sujan, M. (1990). Age differences in product categorization. *Journal of Consumer Research, 16*(March), 452–460.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32*(3), 241–254.

Jones, M., & Love, B. C. (2007). Beyond common features: The role of roles in determining similarity. *Cognitive Psychology, 55*(3), 196–231.

Keller, K. L., & Aaker, D. A. (1992). The effects of sequential introduction of brand extensions. *Journal of Marketing Research, 29*(1), 35–60.

Koivula, N. (2001). Perceived characteristics of sports categorized as gender-neutral, feminine and masculine. *Journal of Sport Behavior, 24*(4), 377–393.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*(1), 1–27.

Lajos, J., Katona, Z., Chattopadhyay, A., & Sarvary, M. (2009). Category activation model: A spreading activation network model of subcategory positioning when categorization uncertainty is high. *Journal of Consumer Research, 36*(June), 122–136.

Loken, B., & Ward, J. (1990). Alternative approaches to understanding the determinants of typicality. *Journal of Consumer Research, 17*(2), 111–126.

Loken, B., Barsalou, L. W., & Joiner, C. (2008). In C. P. Haugtvedt, P. M. Herr, & F. R. Kardes (Eds.), *Handbook of Consumer Psychology* (pp. 133–163). Mahwah, NJ: Lawrence Erlbaum Associates.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207–238.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32* (January), 89–115.

Moreau, C. P., Markman, A. B., & Lehmann, D. R. (2001). What is it? Categorization flexibility and consumers' responses to really new products. *Journal of Consumer Research, 27*(March), 489–498.

Murphy, G., & Ross, B. (1994). Predictions from uncertain categorization. *Cognitive Psychology, 27* (October), 148–193.

Nosofsky, R. M. (1986). Attention, similarity, and the identification categorization relationship. *Journal of Experimental Psychology. General, 115*(1), 39–57.

Ozanne, J. L., Brucks, M., & Grewal, D. (1992). A study of information search behavior during the categorization of new products. *Journal of Consumer Research, 18*(4), 452–463.

Ramaswamy, V., & DeSarbo, W. S. (1990). SCULPTRE: A new methodology for deriving and analyzing hierarchical product-market structures from panel data. *Journal of Marketing Research, 27*(4), 418–427.

Rao, V. R., & Katz, R. (1971). Alternative multidimensional scaling methods for large stimulus sets. *Journal of Marketing Research, 8*(4), 488–494.

Ratneshwar, S. R., & Shocker, A. D. (1991). Substitution in use and the role of usage context in product category structures. *Journal of Marketing Research, 28*(3), 281–295.

Ratneshwar, S. R., Barsalou, L. W., Pechmann, C., & Moore, M. (2001). Goal-derived categories: The role of personal and situational goals in category representations. *Journal of Consumer Psychology, 10*(3), 147–157.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psych, 38*(4), 495–554.

Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review, 86*(2), 87–123.

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology, 71*(5), 888–898.

Snyder, R. (1992). Comparative advertising and brand evaluation: Toward developing a categorization approach. *Journal of Consumer Psychology, 1*(1), 15–30.

Sports Business Journal. (2009). About Us—The Sports Industry, http://www.sportsbusinessjournal.com/index.cfm?fuseaction=page.feature&featureId=43.

Stayman, D. M., Alden, D. L., & Smith, K. H. (1992). Some effects of schematic processing on consumer expectations and disconfirmation judgments. *Journal of Consumer Research, 19*(2), 240–255.

Sujan, M. (1985). Consumer knowledge: Effects on evaluation strategies mediating consumer judgments. *Journal of Consumer Research, 12*(1), 31–46.

Sujan, M., & Bettman, J. R. (1989). The effects of brand positioning strategies on consumers' brand and category perceptions: Some insights from schema research. *Journal of Marketing Research, 26*(4), 454–467.

Sujan, M., & Dekleva, C. (1987). Product categorization and inference making: Some implications for comparative advertising. *Journal of Consumer Research, 14*(3), 372–378.

Takane, Y. (1980). Analysis of categorizing behavior using a quantification method. *Behaviormetrika, 7*(8), 75–86.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352.

Viswanathan, M., & Childers, T. L. (1999). Understanding how product attributes influence product categorization: Development and validation of fuzzy-set based measures of gradedness in product categories. *Journal of Marketing Research, 36*(1), 75–94.