SIMON J. BLANCHARD, DANIEL ALOISE, and WAYNE S. DESARBO*

In a sorting task, consumers receive a set of representational items (e.g., products, brands) and sort them into piles such that the items in each pile "go together." The sorting task is flexible in accommodating different instructions and has been used for decades in exploratory marketing research in brand positioning and categorization. However, no general analytic procedures yet exist for analyzing sorting task data without performing arbitrary transformations to the data that influence the results and insights obtained. This manuscript introduces a flexible framework for analyzing sorting task data, as well as a new optimization approach to identify summary piles, which provide an easy way to explore associations consumers make among a set of items. Using two Monte Carlo simulations and an empirical application of single-serving snacks from a local retailer, the authors demonstrate that the resulting procedure is scalable, can provide additional insights beyond those offered by existing procedures, and requires mere minutes of computational time.

Keywords: sorting task, categorization, positioning, optimization

Online Supplement: http://dx.doi.org/10.1509.jmr.15.0388

Extracting Summary Piles from Sorting Task Data

When the objective of a research study is to explore consumers' perceptions of a set of items (e.g., to determine natural associations between products or brands), a standard approach is to engage participants in a card sorting task (Bijmolt and Wedel 1995; Coxon 1999). In a sorting task, researchers assign participants a set of cards, each of which refers to a particular item (e.g., product, brand, object, or person). The participants are then asked to sort the cards into piles, such that the items in each pile are related to each other in some manner. The manner (i.e., criteria) in which cards are to be sorted can be specified either by participants (open sort) or by the marketing researcher (closed sort). The sorting task has been a popular data-gathering methodology in marketing and other social science areas. Popular survey platforms, such as Qualtrics, have implemented variants of the task, and specialized software and online interfaces provide comprehensive tools for obtaining sorting task data (e.g., cardsorting.net, OptimalSort, Syncaps, UXSort, Websort.net, XSort).

When used in exploratory marketing research to understand brand perceptions, sorting task data are particularly useful in developing concise perceptual structures for the positioning of a competitive set of brands. For example, consider a retailer interested in understanding how consumers choose among a set of brands in the single-serving snacks category, such as the items presented in Table 1. Although studying preferences might be best accomplished by way of a preference-based task such as conjoint analysis, such a task first requires the researcher to determine what consumers envision as the key salient set of product features and attributes. However, defining a comprehensive set of attributes might not be trivial because the associations and comparisons made by consumers are not always apparent to marketing managers—especially when such perceptions are heterogeneous. For instance, in the set of snack items listed

*Simon J. Blanchard is Assistant Professor of Marketing, Georgetown University (e-mail: sjb247@georgetown.edu). Daniel Aloise is Assistant Professor of Computer and Software Engineering, École Polytechnique de Montréal (e-mail: daniel.aloise@polymtl.ca). Wayne S. DeSarbo is Mary Jean and Frank P. Smeal Distinguished Professor of Marketing, Pennsylvania State University (e-mail: wsd6@psu.edu). The authors would like to thank the review team for their thoughtful feedback on the manuscript. They would also like to express special thanks to Daniel Nobre Pinhero for his programming expertise and to Brazil's National Council for Scientific and Technological Development for the partial funding through a grant (CNPq-308887/2014-0) awarded to the second author. All source codes and software are available at www.cardsorting.net. Coeditor: Robert Meyer; Associate Editor: Martijn de Jong.

ISSN: 0022-2437 (print)
1547-7193 (electronic)

Table 1
THE 29 FRITO-LAY PRODUCTS OFFERED BY THE CORP

Brand	*Products*
Lay's	Classic, Sour Cream & Onion, Salt & Vinegar, Garden Tomato & Basil, Sweet Southern Heat Barbecue, Dill Pickle, Barbecue, Kettle Cooked Mesquite Barbecue, Baked Original, Kettle Cooked Original
Chester's Fries	Flamin' Hot
Ruffles	Cheddar & Sour Cream, Sour Cream & Onion
Doritos	Nacho Cheese, Cool Ranch, Jacked Smoky Chipotle BBQ, Spicy Sweet Chili, Spicy Nacho
Cheetos	Puffs, Crunchy, Crunchy Flamin' Hot
Sun Chips	French Onion, Harvest Cheddar
Fritos	Original, Flavor Twists Honey BBQ, Bar-B-Q, Chili Cheese
Smartfood	White Cheddar
Funyuns	Original

in Table 1, we show in our empirical application that consumers consider the salient features to include not only brand families and flavor types, but also less obvious associations such as physical shape (e.g., fry-like products) and holistic judgments (e.g., unhealthy snacks, "classics"). Determining how consumers perceive brands according to their perceptions without a predetermined set of features may thus prove to be a powerful approach for improving brand managers' understanding of their brands and the nature of brand competition in the product class under study.

Sorting task results can also be used to adjust brand positions in the marketplace. For instance, consider Eurostar, a European railway service that has conducted research to understand how consumers navigated its website (Spencer 2007). After first identifying 80 topics that were most relevant to visitors, researchers collected sorting data and applied cluster analyses. They found that by redesigning the menu structure of the website to group content that consumers perceived as naturally going together, they could significantly improve navigation for both expert and novice customers. Not only did the redesign of the website lead to a substantial increase in online sales, but it also led to a decrease in customer service calls by about a third (Spencer 2007). Owing to such successes, research firms such as Websort.net (one of the sorting task platforms previously mentioned) have collected sorting task data for companies such as Allstate, Cabela's, Cisco, Intel, Staples, and Verizon.

Sorting tasks are an attractive methodology for gathering such perceptual data for several reasons. First, the sorting task is flexible and can adapt to different instructions as established by the researcher. Sorting tasks have been used to help understand preferences (e.g., Irwin and Walker Naylor 2009), perceptions of (dis)similarity (e.g., Alba and Chattopadhyay 1986; Sujan and Bettman 1989), group-level categorization decisions (e.g., Hamilton, Puntoni, and Tavassoli 2010), and consistency in categorization judgments (e.g., Peracchio and Tybout 1996; Scott and Canter 1997). Sorting has even become an integral part of understanding customer needs; see Griffin and Hauser's (1993) seminal work on the "voice of the customer."

Second, sorting task data are helpful for understanding consumer perceptions of large sets of items (Rao and Katz 1971). Compared with other approaches, such as paired comparison tasks, conditional rankings, or triadic combinations (Bijmolt and Wedel 1995), sorting tasks produce less fatigue and boredom in participants, provide equally good results, and offer faster completion times. Prior research has found that sorting tasks are intuitive and easy for participants (e.g., Coxon 1999). In a recent investigation about how sorting task structures influence the participant experience, Blanchard and Banerji (2016) found that even when sorting 40–60 items, participants did not find the task difficult and could complete it in just a few minutes.

A third factor underlying the popularity of sorting tasks is the availability of free online interfaces for gathering sorting data. These online interfaces, such as cardsorting.net, facilitate access and administration to large online panels such as Amazon's Mechanical Turk. For all these reasons and more, sorting tasks have become increasingly popular among marketing researchers and managers interested in gathering exploratory data on brand perceptions.

ANALYZING SORTING TASK DATA

Despite being used primarily in exploratory research, the insights extracted from sorting task data should be quantifiable. The biggest challenge involved in using sorting task data lies in its analysis. Without any transformation, the raw data involves i ($i = 1, ..., I$) consumers who sort J items into c_i piles such that $y_{il_i j} = 1$ if consumer i puts item j in his or her l_i^{th} pile ($l_i = 1, ..., c_i$), 0 otherwise. As an illustration, we present sample generated output data in Table 2 related to a sorting task of single-serving snacks (e.g., chips) in a business context that we shall return to later in the manuscript as our application.

Two difficulties stand out. The first is that, whereas each consumer sorts the same set of (J) items, a row in a sorting task data set is at the consumer-pile level. Making comparisons between consumers becomes challenging because the number of piles c_i created by individuals can differ substantially (e.g., in the example in Table 2, consumer 1 made four piles, but consumer 3 made six piles). The second difficulty is that the l_i indices have no correspondence between consumers because there is no logical ordering of the l_i piles consumer i made. That is, two consumers might have made the same pile (e.g., a "barbecue" pile), but these indices in the data do not allow direct comparison (e.g., consumer 3's first pile is the same as consumer 4's third pile). This lack of correspondence between indices creates difficulties in identifying similarities between piles from different consumers and in summarizing commonalities in how different consumers generated their piles.

Owing to such complexities, researchers and practitioners have often resorted to various data manipulations to make the data compatible with the analytical tools they are comfortable with. For instance, a typical approach involves converting sorting task data into pairwise similarities between the sorted items. That is, instead of recording the piles by setting $y_{il_i j}$ to 1 if consumer i sorted item j into the l_i^{th} pile (0 otherwise), the researcher converts the original data wherein $y'_{ijk} = 1$ if consumer i sorted items j and k ($j, k = 1, ..., J$) into the same pile. Doing so allows each consumer's sort to be represented by a J items × J items symmetric pairwise similarity matrix—often then summarized through a count of the number of consumers who put each pair into the same pile.

Table 2
ILLUSTRATION OF PERCEPTUAL SORTING TASK DATA (MULTIPLE-CARDS SORTING TASK)

Consumer's Pile (1_i)	*Label*	*Brand 1*	*Brand 2*	*Brand 3*	*Brand 4*	*Brand 5*	...	*Brand J*
Consumer 1								
1	Flavored chips	1	0	1	0	0	...	0
2	Nonchips	1	0	0	1	1	...	0
3	Ruffles	0	0	1	0	0	...	0
4	Potato chips	0	1	1	0	0	...	0
Consumer 2								
1	Cheesy	1	0	0	1	1	...	0
2	Classic-style chips	0	1	1	0	0	...	1
3	Alternatively cooked	0	0	0	0	0	...	1
4	Sunchips	0	0	0	1	1	...	0
Consumer 3								
1	Healthy	0	0	0	1	1	...	1
2	I like	0	1	0	1	0	...	0
3	I don't like	1	0	0	0	1	...	0
4	Yuk	0	0	1	0	0	...	0
5	Lay's	0	1	0	0	0	...	1
6	Sunchips	0	0	0	1	1	...	0
Consumer 4								
1	Doritos	1	0	0	0	0	...	0
2	Not chips	0	0	0	1	1	...	0
3	Not too bad for you	0	0	0	1	1	...	1
4	Spicy	1	0	1	0	0	...	0
...								
l_n		1	1	0	1	1	...	0

It is common to use such pairwise count matrices to identify groups of items that consumers tend to organize together, typically via cluster or multidimensional scaling analyses (e.g., Johnson 1967; Takane 1980; see also websort.net's implementation of hierarchical clustering). However, given that clustering and multidimensional scaling models often require distances as input (and not similarities), one must first convert the similarities into distances—a conversion that has a significant impact on the resulting solutions obtained (see Green and Rao 1969; Punj and Stewart 1983). Afterward, marketing researchers must arbitrarily choose from the various models available to analyze the pairwise distance data. These models range from continuous representations (e.g., multidimensional scaling) to more discrete representations (e.g., cluster analysis, such as Ward's [1963] method), which invariably also leads to different results and interpretations (Punj and Stewart 1983).

More recently, researchers have also modeled the sorting data *directly* by making assumptions about the underlying latent categorization process regarding the items being sorted into the same piles (i.e., the piles are the realized behavior of the underlying category structures). DeSarbo, Jedidi, and Johnson (1991) and Blanchard et al. (2012) propose latent structure models that identify the unobserved categories that consumers perceive in a set of items, with the assumption that y_{ijk} follows a Bernoulli distribution based on pairwise latent similarity judgments. Note that their procedures cannot accommodate the wide variety of sorting tasks that are not based on a latent pairwise similarity process (e.g., choice context, preferences). Blanchard and DeSarbo (2013) model a variant of the sorting task wherein participants are given multiple cards per item. Their model assumes that these pairwise counts are generated from a pairwise Poisson process with a gamma-distributed rate that is also a function of a latent pairwise similarity judgment between pairs of items. As is true in the research of DeSarbo, Jedidi, and Johnson (1991) and Blanchard et al. (2012), the procedures involve conditional maximum likelihood estimates (MLEs) of the parameters through numerical optimization; here, the model's optimization (MLE) computation requirements are quite extensive, even for moderately sized marketing research applications (e.g., 100–150 participants, 20–50 items). Most importantly, these parametric models proposed are not appropriate unless one knows that the piles are formed by a cognitive process relying on pairwise similarity judgments. Namely, unless the task instructions provided to participants are to sort by pairwise similarities, the resulting analyses can be misleading.

SUMMARIZING SORTING TASK DATA: THE IDENTIFICATION OF SUMMARY PILES

In this article, we propose a new methodology for the identification of summary piles (i.e., piles that best describe a collection of heterogeneous sorts) to explore perceptions among a set of items. The proposed methodology offers several advantages. First, it can accommodate a variety of sorting tasks in which respondents use one or multiple cards per item and for different instructions (e.g., sort by similarity, dissimilarity, preference, or any other criteria). The procedure is agnostic about the cognitive process consumers employ. That is, it can be used with a wide range of sorting tasks, including (but not limited to) those based on similarities. Second, unlike existing clustering or decomposition methodologies, the proposed methodology does not require arbitrary conversions of the created piles into pairwise similarities and distances that can influence the results; it models the piles directly. Third, the model can recover sorting data generated by participants who have latent heterogeneous perceptions. Fourth, unlike existing parametric models for identifying latent category structures, the proposed methodology does not require parametric assumptions about the data-generating process (that may or may not be satisfied). Fifth, our proposed optimization approach is fast and is scalable to large sorting task problem sizes. MATLAB code files and a Windows-based graphical user interface are available.

In the next section, we introduce the mathematical model at the core of our proposed methodology, along with the optimization strategies. After we illustrate the benefits of the proposed procedure for the identification of summary piles, we present the results of two Monte Carlo simulations and an application involving consumer perceptions of snack brands maintained by a local retailer. We conclude with a summary of our contributions and avenues for further research.

EXTRACTING SUMMARY PILES FROM GENERALIZED SORTING TASK DATA

Sorting task data involves I consumers $(i = 1, \ldots, I)$ who make c_i piles using J $(j = 1, \ldots, J)$ items; at the disaggregate level, the resulting data can be represented such that $y_{il_ij} = 1$ if consumer i puts item j into pile l_i $(l_i = 1, \ldots, c_i)$, and 0 otherwise. In this section, we propose a mathematical model whose objective is to identify a small set of piles that best summarize all the piles made by consumers in the data.

In our proposed model, the summary piles are selected so that each pile in the consumers' data is approximated by one of the K summary piles with minimal mispredictions of y_{il_ij}. Our model thus represents a new data reduction technique whereby two key features are simultaneously identified: (1) the K summary piles that best approximate the piles made by consumers across the data, and (2) the association of each of the N piles in the data (where $N = \sum_{i=1}^{I} c_i$) to one of these summary piles.

Model Specification

Formally, let p_{mj} $(m = 1, \ldots, K)$ be 1 if summary pile m includes item j, 0 otherwise; and let x_{il_im} be 1 if summary pile m is determined to be the best summary pile to predict consumer i's l_i^{th} pile, 0 otherwise. Further, let z_{il_ij} represent the misprediction error (if any) that occurs when p_{mj} is used to predict the presence of item j in consumer i's l_i^{th} pile (y_{il_ij}). Then, the proposed model can be expressed via the following binary integer programming model with quadratic binary constraints:

$$(1) \qquad \min\ Z = \sum_{i=1}^{I} \sum_{l_i=1}^{c_i} \sum_{j=1}^{J} z_{il_ij},$$

subject to

$$(2) \qquad \sum_{m=1}^{K} x_{il_im} p_{mj} \le z_{il_ij} \quad \forall i = 1, \ldots, I; \quad \forall l_i = 1, \ldots, c_i; \quad \forall j = 1, \ldots, J \text{ whenever } y_{il_ij} = 0;$$

$$(3) \qquad \sum_{m=1}^{K} x_{il_im}\left(1 - p_{mj}\right) \le z_{il_ij} \quad \forall i = 1, \ldots, I; \quad \forall l_i = 1, \ldots, c_i; \quad \forall j = 1, \ldots, J \text{ whenever } y_{il_ij} = 1;$$

$$(4) \qquad \sum_{m=1}^{K} x_{il_im} = 1 \quad \forall i = 1, \ldots, I; \quad \forall l_i = 1, \ldots, c_i;$$

$$(5) \qquad \sum_{j=1}^{J} p_{mj} \ge 1 \quad \forall m = 1, \ldots, K;$$

$$(6) \qquad x_{il_im} \in \{0, 1\} \quad \forall i = 1, \ldots, I; \quad \forall l_i = 1, \ldots, c_i; \quad \forall m = 1, \ldots, K;$$

$$(7) \qquad z_{il_ij} \in \{0, 1\} \quad \forall i = 1, \ldots, I; \quad \forall l_i = 1, \ldots, c_i; \quad \forall j = 1, \ldots, J;$$

and

$$(8) \qquad p_{mj} \in \{0, 1\} \quad \forall m = 1, \ldots, K; \quad \forall j = 1, \ldots, J.$$

In Expression 1 (the objective function Z), we aim to minimize the number of mispredictions among the $(J \times N)$ assignments of J items by the sample of consumers into their piles. It is, in simple terms, a measure of fit across all piles. For each of the N piles in the data and each of the J items, one constraint is generated from either Expression 2 or Expression 3. We note that the only data used by the proposed model is y_{il_ij}.

Obtaining mispredictions. If item j is observed to be assigned to consumer i's l_i^{th} pile (i.e., $y_{il_ij} = 1$), a constraint following Expression 3 is generated, and variable z_{il_ij} is 0 if the data point is predicted correctly, 1 if not. More specifically, the data will be predicted correctly if (1) consumer i's l_i^{th} pile is determined to be best approximated by summary pile m (i.e., $x_{il_im} = 1$), and (2) summary pile m does indeed include item j (i.e., $p_{mj} = 1$) as $x_{il_im}(1 - p_{mj}) = 0$. Given that the objective of the problem is to minimize the total errors and that z_{il_ij} is restricted to be either 0 or 1, the inequality in Expression 3 will assign $z_{il_ij} = 1$ (an error) if instead $p_{mj} = 0$ because $x_{il_im}(1 - p_{mj}) = 1(1 - 0) = 1$. That is, z_{il_ij} will have captured an error of *omission* (i.e., item j should have been predicted to be assigned to the pile, and it was not). The expression is summed over the K summary piles, as each pile is to be associated with only one summary pile.

If item j is *not* observed to have been assigned to consumer i's l_i^{th} pile (i.e., $y_{il_ij} = 0$), then the data will be predicted correctly if (1) consumer i's l_i^{th} pile is determined to be best approximated by summary pile m (i.e., $x_{il_im} = 1$), and (2) summary pile m does *not* include item j (i.e., $p_{mj} = 0$) as $x_{il_im}p_{mj} = 1 \times 0 = 0$. Given that the objective of the problem is to minimize errors and that z_{il_ij} is restricted to be either 0 or 1, the inequality in Expression 2 will assign $z_{il_ij} = 1$ (an error) if instead $p_{mj} = 1$ as $x_{il_im}p_{mj} = 1 \times 1 = 1$. That is, z_{il_ij} will have captured an error of *commission* (i.e., item j should not have been predicted to be assigned to the pile, and it was). This is summed over the K summary piles.

Restrictions. The model described above generates one constraint for each of the $N \times J$ piles made, and it aims at identifying simultaneously the composition of the summary piles and their assignment to predict each observed pile. To do so, several constraints need to be imposed. Specifically, the constraints in Expressions 4–8 concern the identification of the model's key features. The constraints in Expression 4 ensure that each pile is assigned one and only one summary pile. The constraints in Expression 5 ensure that there is no empty pile among the summary piles. The constraints in Expressions 6–8 ensure that x_{il_im}, p_{mj}, and z_{il_ij} are either 0 or 1.

Finally, we note that the proposed model in Expressions 1–8 does not require any assumption about whether any item is allowed to belong to one and only one pile or to belong to multiple piles (i.e., $\sum_{l_i=1}^{c_i} y_{il_ij} \geq 1 \;\; \forall i, j$). It can thus be used with both the multiple-cards and single-card sorting tasks (see Blanchard and Banerji 2016) as required by the researcher. The proposed model also does *not* require that every item belong to at least one pile because researchers sometimes instruct participants who do not recognize a particular item to leave it unsorted (Blanchard and Banerji 2016; Coxon 1999).

Incorporation of heterogeneity. Our model allows for heterogeneous solutions vis-à-vis the total set of K summary piles, **p**, and assignments to summary piles, **x**. To illustrate this, consider the following synthetic example. Given a set of identified summary piles $(\mathbf{p})$, recall that x_{il_im} is 1 if summary pile m predicts consumer i's l_i^{th} pile, 0 otherwise. Assume two consumers who made two and three piles respectively and that our model suggested the presence of four summary piles. Table 3 presents an illustrative example of how the model can recover and capture individual differences in their sorts.

In Panel A of Table 3, each row is a pile made by a consumer, and in the last column, we indicate the assignment of that observed pile to one of the summary piles. In Panel B, we show the composition of the summary piles. For consumer 1, the piles are approximated by summary piles 1 and 3, whereas for consumer 2, the piles are approximated by summary piles 2, 3, and 4. These heterogeneous assignments are crucial to the proposed model because the difference between what is predicted and what is observed results in the misprediction error that is captured in the objective function. We also note that the assignment rate for summary pile m $(\sum_{i=1}^{I}\sum_{l_i=1}^{c_i} x_{il_im}/N)$ represents the sample-level probability that pile m is made. It can be used as a measure of consensus across the sample, but at the individual level (via subscript i), the proposed model allows for the recovery of heterogeneous sorts. More specifically, through the assignment of piles to summary piles that varies across individuals x_{il_im}, the model can predict which of the identified summary piles relate to each of consumer i's observed piles.

Table 3

SYNTHETIC DATA (y_{il_ij}): ILLUSTRATION OF RESPONDENT HETEROGENEITY THROUGH ASSIGNMENTS (VIA x_{il_im} AND p_{mj}) TO THREE SUMMARY PILES

A: Data (y_{il_ij}) and Assignments to Summary Piles (x_{il_im})					
l_i	*Brand 1*	*Brand 2*	*Brand 3*	*Brand 4*	*Assigned Summary Pile*
Consumer 1					
1	0	1	1	0	Pile 3 ($x_{1,1,3} = 1$)
2	1	0	0	1	Pile 1 ($x_{1,2,1} = 1$)
Consumer 2					
1	0	1	1	0	Pile 3 ($x_{2,1,3} = 1$)
2	0	1	0	1	Pile 2 ($x_{2,2,2} = 1$)
3	1	0	0	1	Pile 1 ($x_{2,3,1} = 1$)
Assignment rate	.20	.20	.40	.20	

B: Summary Piles (p_{mj})				
Summary Pile Index (m)	*Brand 1*	*Brand 2*	*Brand 3*	*Brand 4*
1	1	0	0	1
2	0	1	0	1
3	0	1	1	0

Optimization Strategy

The model described in Expressions 1–8 is a binary mathematical program with quadratic 0–1 constraints. Even using commercial solvers for small data sets, one cannot solve such models to obtain the global optimum within reasonable computational times. In Appendix A, we describe our strategy to tackle the mathematical problem in Expressions 1–8. First, we take advantage of the nature of the sorting task data and identify redundancies in the constraints that can be eliminated to reduce the number of operational or binding constraints and accelerate optimization. Second, we propose a modified variable neighborhood search (VNS) heuristic (Hansen and Mladenović 2001; Mladenović and Hansen 1997) tailored to our optimization problem. VNS is a metaheuristic, that is, a framework for constructing heuristics, aimed at solving combinatorial and global optimization problems. Since its inception, VNS has undergone many developments and has been applied in numerous fields (for a recent survey, see Hansen, Mladenović, and Pérez 2010). In Appendix B, we present a proof that our problem is, in fact, NP-hard. Given the NP-hard nature of the problem, we must resort to heuristics to find approximate solutions to the problem in reasonable computational times.

Selection of K, the number of summary piles. The only parameter that a researcher must set a priori is the value of K—that is, the number of summary piles to extract. In parametric models that deal with MLE-based latent structure models (e.g., finite mixture models), researchers can employ information criteria (Akaike information criterion [AIC], Bayesian information criterion [BIC], consistent AIC [CAIC], modified AIC [MAIC], etc.) to select an appropriate number of components. As is well known, different criteria usually result in different decisions about which solution is most parsimonious (Yang and Yang 2007). In nonparametric models, such as the one we propose, a general recommendation is to examine solutions for increasing values of K and use a scree plot (or various deterministic goodness-of-fit measures) to determine the value of K, where increasing K by one additional unit does not produce a sufficient improvement in the objective function.

One key challenge in using the scree plot with our proposed methodology, however, is that unlike traditional data decomposition techniques (e.g., clustering, multidimensional scaling, principal component analysis) for which a small number of dimensions/clusters is expected, our proposed model incorporates heterogeneity in a way that results in the identification of more than just a few summary piles. Specifically, setting K at the mean number of actual consumer piles (i.e., average c_i) may provide poor fits and obscure the presence of heterogeneity. As we show in our second Monte Carlo simulation with synthetic data whose structure is known, using the scree plot (on the percentage improvement between increasing values of K) can still be successful at recovering the true value. In our experiments, we were able to perfectly recover K in 79.8% of trials. That said, we also believe it is thus critical to use the interpretation of the results as a supplemental guide to model selection (Wedel and Kamakura 2000), and we do so in our application.

Moreover, because we assess the performance of the model using the same criteria that is directly optimized by the model (i.e., mispredictions), it is possible that the model can overfit, such that the results poorly predict out-of-sample data, particularly at high values of K. Such concerns prompt a common practice of splitting the sample into a training and a validation, or holdout, data set, which is often suboptimal (e.g., Steckel and Vanhonacker 1993). As such, researchers advocate the use of more complicated procedures, such as T-fold cross-validation, in which the sample is randomly split into T equal-sized partitions; then each of the T samples serves once as a validation data set for testing the model, with the remaining T – 1 samples used to train the model. Thus, to avoid overfitting, we recommend that users compare the in-sample fit, for the selected value of K, with the fit that can be obtained from a tenfold cross-validation. A significant discrepancy between the in-sample and cross-validation fit likely indicates a problem with the selected value of K.

Computational times and random starts. Our proposed approach and optimization strategy is a nonparametric methodology that requires few decisions on the part of the user. Unlike local optimization models that terminate after a convergence criterion has been met, the use of a metaheuristic (i.e., VNS) changes how termination occurs. Specifically, VNS can be set to terminate after a maximum number of local searches has been performed or when the allowed computational time has elapsed. In our experiments, we find that 300 seconds seems to provide adequate confidence that the best possible solution has been found. Thus, in our code, we set optimization to terminate after 300 seconds. Of course, users might differ in their degree of tolerance for risk of being stuck in a poor local optimum and, as such, might decide to allocate more time. However, because our optimization already incorporates the exploration of neighborhoods to minimize risks of local optima, multiple "random starts" are often not necessary.

Robustness and Model Comparisons

To test the robustness of our proposed methodology and its ability to recover different types of sorting task data, we generated a Monte Carlo simulation that is presented in detail in online Technical Appendix 1. In this appendix, we show that the proposed model is able to predict sorting data that vary in terms of the number of consumers, the number of items to be sorted, and the types of sorting task (single-card sorting task vs. multiple-cards sorting task). We also show that the proposed VNS heuristic is relatively insensitive to parameters such as the amount of computational time and VNS tuning parameters. The model thus requires little calibration, although such options are made available in the code and software.

In online Technical Appendix 2, we present a second series of simulated data sets and compare three competing approaches: our proposed method; Ward's (1963) clustering, applied to a pairwise count matrix; and latent Dirichlet allocation (LDA; Blei, Ng, and Jordan 2003), a machine learning technique typically used in the context of document analysis. We now describe LDA in more detail.

A typical use of LDA is to analyze the vocabulary used across many documents and to identify "representative topics." In LDA terminology, documents are composed of various words (the vocabulary), and the aim is to identify a set of topics that best summarize the content (words) of the entire set of documents. In the context of summarizing sorting task data, if we assume that (1) the set of consumer-generated piles represents the set of documents, (2) the brands/items represent the available words and consist of the entire vocabulary, and (3) the summary piles are the representation of the topics, then a sorting data set can be analyzed using LDA algorithms.[1]

Applied to our sorting data and the identification of summary piles, the basic LDA assumes that the data-generating process for an observed pile is obtained from a mixture of K unobserved summary piles, each with its own associations to each of the J brands. First, let $P(\theta)$ be the probability distribution over the set of K summary piles θ, where θ is a $1 \times K$ vector. Second, let $P(\omega \mid \theta)$ be the probability distribution of the J brands over the set of K summary piles (i.e., brand-summary piles distribution). Then, the model determines the probability distribution of assigning brand j to observed piles through the following equation:

$$P(\omega_j) = \sum_{k=1}^{K} P(\omega_j \mid \theta_j = k) P(\theta_j = k), \tag{9}$$

where K is the number of summary piles, $P(\theta_j = k)$ is the probability that the kth summary pile is sampled for brand j, and $P(\omega_j \mid \theta_j = k)$ is the probability that brand j is sampled from summary pile k for the observed pile. When both the brand-summary piles and the summary piles are drawn from multinomial distributions with Dirichlet priors, one obtains the LDA model introduced by Blei et al. (2003).

Whereas several heuristics have been proposed to fit the statistical problem in LDA (see Blei 2012), for our comparative analyses we used the MATLAB Topic Modeling Toolbox (Steyvers and Griffiths 2007), which employs Gibbs sampling and takes advantage of the conjugate nature of the Dirichlet prior with the multinomial distribution. The LDA model has been shown to perform well in finding associations between concepts and in the context that relies on individual categorization structures (Griffiths, Steyvers, and Tenenbaum 2007). It has also been shown to be helpful for understanding brand associations (Tirunillai and Tellis 2014).

We produced 72 data sets as part of a factorial design that generated synthetic sorting task data for the single-card sorting task. For this second simulation, we used the single-card sorting task because most clustering algorithms do not allow items to belong to multiple clusters; using multiple-card sorting data could have provided an undue advantage to both LDA and our proposed methodology. We also generated the data through a different form of heterogeneity whereby the solutions vary across groups of consumers (i.e., latent segments). The results show that our proposed model performed at least as well as the competing approaches in all 72 trials and that when no error was added, our proposed model recovered the data perfectly in all but one trial. The total reduction in mispredictions of the sorting data by our proposed model was 75.15% better than

[1]We are indebted to an anonymous reviewer for recognizing the parallels between the two problems.

clustering, and 91.95% better than LDA. More details are provided in online Technical Appendix 2.

Simulations Summary

We have proposed a new methodology to identify the piles that best summarize a set of data from heterogeneous consumers performing a sorting task. We have shown through two simulations and a synthetic example (in Table 3) that not only can the proposed methodology recover heterogeneous structures for different types of sorting tasks, but also our instantiation of a VNS heuristic can produce better fits than competing models.

APPLICATION: PERCEPTIONS OF SINGLE-SERVING SNACKS

In the present application, we study brand perceptions for a set of single-serving snacks (chips) offered by a local retailer. We do so to help the retailer explore the salient features of single-serving snacks as perceived by consumers. First, we administer a multiple-cards sorting task (i.e., an item may belong to more than one pile) in which participants sort the brands according to their perceptions. Second, we identify summary piles and interpret the piles formed. Third, we compare the results in terms of fit and insights against LDA and cluster analysis.

The Corp's Single-Serving Snacks

The Corp is a store created by the students of Georgetown University that has been in business for more than 40 years. This full-service grocery store sells frozen foods, produce, sandwiches, chips, and toiletries, mainly to the university's resident students. It is a "one-stop shop" for students on campus, and it faces little to no competition from nearby establishments. The CEO, Board of Directors, and cashiers at the Corp are all undergraduate students attending the institution. In 2011, the store grossed $2,142,065 in sales, with a 23% gross profit. Among its key products, the Corp sells a substantial amount of single-serving snacks; the current single-serving section of the store features only Frito-Lay products, mainly to facilitate inventory management and scheduling. At the time of our study, the store offered the 29 Frito-Lay products listed in Table 1.

This sample offered by the Corp enables us to illustrate the capabilities of the proposed model. Although studying brand perceptions is only one potential use of the sorting task, it is a practical marketing application in a realistic business setting. Our brand set includes the full, actual product category assortment, which provides a suitable scenario for understanding competitive structures. Moreover, the assortment composition is fixed and limited, to be manageable for the retailer. Finally, there is potential for substantial heterogeneity in the way consumers sort products, including ways that the distributor and retailer may not have considered (e.g., by brand, flavor type, main ingredient, healthiness, spiciness).

Participants and Procedures

We collected data from sophomore undergraduates at Georgetown University who participated in our study in exchange for extra course credit. We obtained data for 101 such participants. After entering the research lab, all participants learned that the study would concern perceptions of snacks.

We first introduced participants to the idea of a sorting task. We followed Lickel, Hamilton, and Sherman (2001) and Blanchard and Banerji (2016), who showed a set of pretask instructions to explain the general purpose of a sorting task. This general set of instructions illustrates how the brands were to be dragged into piles. Specifically, a first computer screen displayed the following text, along with sample piles:

> In a sorting task, your job is simply to put things together as "piles" as you see fit. For example, consider the following list of musical instruments: "Clarinet, Flute, Guitar, Harp, Piano, Saxophone, Trumpet, and Violin" as sorted in the following way: **Strings**—Violin, Guitar; **Large Instruments**—Harp, Piano; and **Wind**—Clarinet, Flute, Trumpet, Saxophone.
>
> As you can see, the instruments have been put together in the same box. You may have seen a different way of grouping these instruments together. Your way of grouping them would have been just as valid as the one we provided here.

Participants then were directed to a second screen that displayed the following message: "Imagine that you are running between classes, and are looking to grab a snack. How would you choose between the snacks below? Take a minute to consider the process that would go through your mind in such a purchase situation." The rest of the screen showed the 29 boxes and pictures from the assortment, and participants had to stay on this page for at least 15 seconds.

After examining the screen, participants were told to sort on the basis of their perceptions using the multiple-cards sorting task (Blanchard and DeSarbo 2013). A window displayed the following text:

> Imagine that you went grocery shopping and you are selecting snacks. Standing in front of the snack shelf, you have 29 snacks in front of you.
>
> Drag all the snacks into piles such that snacks in the same piles "go together" according to your own perception.
>
> You can use any criteria you want when dragging these snacks into different piles, and you may put as many or as few of the snacks together as in a pile, and you may put snacks in multiple piles if you see fit.
>
> There are no right or wrong ways of putting the snacks together. We are only interested in how you see the snacks as going together. Please use all the snacks listed on the left side panel.

Participants then minimized the instructions window, revealing the online interface for the sorting task from cardsorting.net. Participants saw the assortment and a "surface" on which to sort the snacks into piles. The left side included an individually randomized list of the 29 snacks, along with a 75 × 75–pixel picture of the bags, just as consumers would see on shelves. Participants then proceeded to drag and drop the snacks into piles. Through the drag-and-drop functionality, participants could easily add new snacks to piles, move snacks from one pile to another, remove a snack from a pile, and/or remove a pile entirely. Participants were required to put all the snacks in at least one pile, a common requirement in sorting tasks (Blanchard and Banerji 2016; Coxon 1999). A counter indicated how many snacks had yet to be used. If participants tried to submit their sort (by clicking the "submit" button) without having used all the snacks, they received the following message and were prevented from continuing:

> Whoops! You did not put all snacks into piles. This study requires you to use all the snacks at least once. Please put all of the snacks once on the board before proceeding! Note that you can put a snack in a pile of its own if you don't feel it's similar to anything based on your criteria. Here are the snacks that you have not put into piles.

The list of unsorted snacks then appeared.

Results: Model Selection

Model fit. On average, the 101 participants made 6 piles each, with 245 unique piles across the total sample of 573 piles. Of note, we found that 28 out of the 29 brands were used more than once by the same consumer for at least one participant. About 16% of the participants sorted at least one brand into multiple piles, and about 9% of the piles contained a brand that had been used in another pile by the same consumer.

Consistent with the findings from our Monte Carlo simulation, we ran the algorithm five times for each sequential value of the number of summary piles for $K = 1, \ldots, 20$, limiting the results to five minutes per execution. The results are presented in Table 4. As discussed earlier, model selection in deterministic models is somewhat arbitrary (Blanchard, Aloise, and DeSarbo 2012); we used a combination of the sequential decrease in the objective function ("elbow in the curve") and interpretation to select a solution with $K = 10$ summary piles. Whereas the $K = 10$ solution had improved over $K = 9$ by 6.1%, the solution of $K = 11$ improved over $K = 10$ by only 4.3%, which appears to be a significant drop in improvement. We note that our online Technical Appendix 2 provides additional guidance on model selection.

The misprediction of 1,536 represents an accuracy of 92.08% over the $587 \times 29 = 16,617$ cells in the data. This result suggests an excellent explanatory power of the 10 summary piles (reduced from 587). Using the selected sample as training data, we found that 82.05% of all y_{il_ij} are 0, so a naïve model that predicts that all y_{il_ij} will be 0 would produce only 2,983 failed predictions, for a 17.95% misprediction rate. The proposed model reduces the error by a multiple of approximately two, with a misprediction rate of 7.92% and an improvement of 51.49% over the null model. We note that the model was stable. On average, at each level of $K = 1, \ldots, 20$, the minimum optimum found was obtained in 3.85 out of 5 executions.

To also ensure that our improvement in fit did not merely reflect overfitting due to having more predictive power, we calculated tenfold cross-validation for $K = 10$ for our model. Whereas the in-sample error reported above is 7.92%, the tenfold cross found a median cross-validation error of 8.53% ($M = 8.08\%$, $SD = 1.53$; min = 5.76%, max = 9.87%).

Summary Piles Solution Interpretation

The summary piles identified by the proposed model are displayed in Table 5, along with our proposed labels and average assignment information (a number that indicates the proportion of all 587 piles in the data that are predicted to be most similar to this summary pile).

Of all the piles made, the "popcorn" pile with a single item (Smartfood popcorn) is the most common, with a probability of 21%. The second–most common summary pile included fry-shaped snacks (12% assignment rate; brands included Cheetos Crunchy Flamin' Hot, Cheetos Puffs, and Chester's Flamin' Hot Fries). For our participants, brand names also

Table 4
EMPIRICAL APPLICATION: MODEL SELECTION AND MODEL FIT (MISPREDICTIONS)

Number of Summary Piles	*Proposed Methodology* Execution 1	2	3	4	5	Min	Decrease	Percentage Decrease	*LDA* Execution 1	2	3	4	5	Min	Decrease	Percentage Decrease
1	4,810	4,810	4,810	4,810	4,810	4,810	—	—	—	—	—	—	—	—	—	—
2	2,924	2,930	2,932	2,924	2,932	2,924	1886	39.2%	6,800	6,105	6,276	6,362	6,253	6,105	—	—
3	2,440	2,440	2,452	2,440	2,521	2,440	484	16.6%	4,113	4,291	3,986	4,116	4,149	3,986	2,119	34.7%
4	2,187	2,316	2,238	2,238	2,162	2,162	278	11.4%	3,327	3,242	3,288	3,456	3,469	3,242	744	18.7%
5	1,974	1,980	1,915	1,915	1,985	1,915	247	11.4%	2,979	3,009	2,544	2,727	2,863	2,544	698	21.5%
6	1,754	1,754	1,754	1,754	1,754	1,754	161	8.4%	2,316	2,270	2,297	2,496	2,584	2,270	274	10.8%
7	1,646	1,646	1,646	1,646	1,646	1,646	108	6.2%	2,204	2,258	2,156	2,054	2,144	2,054	216	9.5%
8	1,536	1,536	1,536	1,536	1,536	1,536	110	6.7%	2,119	1,987	2,064	2,023	2,072	1,987	67	3.3%
9	1,444	1,444	1,444	1,444	1,444	1,444	92	6.0%	1,995	1,969	2,083	2,074	2,185	1,969	18	.9%
10	**1,356**	**1,356**	**1,356**	**1,356**	**1,356**	**1,356**	**88**	**6.1%**	1,959	1,995	1,933	2,010	2,004	1,933	36	1.8%
11	1,298	1,298	1,298	1,298	1,298	1,298	58	4.3%	1,966	1,994	2,013	2,040	2,067	1,966	−33	−1.7%
12	1,243	1,242	1,242	1,242	1,242	1,242	56	4.3%	2,010	1,867	1,863	1,906	1,984	1,863	103	5.2%
13	1,186	1,186	1,186	1,191	1,186	1,186	56	4.5%	1,827	1,979	1,936	2,006	1,834	1,827	36	1.9%
14	1,134	1,144	1,134	1,134	1,134	1,134	52	4.4%	1,987	1,863	2,037	1,834	1,838	1,834	−7	−.4%
15	1,089	1,089	1,089	1,089	1,089	1,089	45	4.0%	1,848	1,836	1,905	1,804	1,887	1,804	30	1.6%
16	1,047	1,047	1,055	1,047	1,051	1,047	42	3.9%	1,757	1,864	1,955	1,814	1,785	1,757	47	2.6%
17	1,013	1,013	1,013	1,013	1,008	1,008	39	3.7%	1,801	1,780	1,792	1,812	1,772	1,772	−15	−.9%
18	974	975	974	976	974	974	34	3.4%	1,846	1,873	1,792	1,777	1,841	1,777	−5	−.3%
19	940	949	940	945	940	940	34	3.5%	1,671	1,918	1,855	1,905	1,806	1,671	106	6.0%
20	911	911	914	916	911	911	29	3.1%	1,676	1,813	1,724	1,811	1,733	1,676	−5	−.3%

Notes: Boldface indicates the selected solution.

TABLE 5
APPLICATION: IDENTIFIED SUMMARY PILES

Assignment rate	*Popcorn 21%*	*Fry-Like 12%*	*Ruffles 12%*	*Doritos 12%*	*Sun Chips 11%*	*Fritos 10%*	*All Lay's 10%*	*Classic 6%*	*BBQ 4%*	*Unhealthy 2%*
Cheetos Crunchy	0	1	0	0	0	0	0	0	0	1
Cheetos Crunchy Flamin' Hot	0	1	0	0	0	0	0	0	0	1
Cheetos Puffs	0	1	0	0	0	0	0	0	0	1
Chester's Flamin' Hot Fries	0	1	0	0	0	0	0	0	0	1
Doritos Cool Ranch	0	0	0	1	0	0	0	0	0	1
Doritos Jacked Smoky Chipotle BBQ	0	0	0	1	0	0	0	0	1	1
Doritos Nacho Cheese	0	0	0	1	0	0	0	0	0	1
Doritos Spicy Nacho	0	0	0	1	0	0	0	0	0	1
Doritos Spicy Sweet Chili	0	0	0	1	0	0	0	0	0	1
Fritos Bar-B-Q	0	0	0	0	0	1	0	0	1	1
Fritos Chili Cheese	0	0	0	0	0	1	0	0	1	1
Fritos Flavor Twists BBQ	0	0	0	0	0	1	0	0	1	1
Fritos Original	0	0	0	0	0	1	0	1	0	1
Funyuns Original	0	0	0	0	0	0	0	0	0	1
Lay's Baked Original	0	0	0	0	0	0	1	1	0	0
Lay's Barbecue	0	0	0	0	0	0	1	0	1	1
Lay's Classic	0	0	0	0	0	0	1	1	0	1
Lay's Dill Pickle	0	0	0	0	0	0	1	0	0	1
Lay's Garden Tomato & Basil	0	0	0	0	0	0	1	0	0	1
Lay's Kettle Cooked Barbecue	0	0	0	0	0	0	1	0	1	1
Lay's Kettle Cooked Original	0	0	0	0	0	0	1	1	0	0
Lay's Salt & Vinegar	0	0	0	0	0	0	1	0	0	1
Lay's Sour Cream & Onion	0	0	0	0	0	0	1	0	0	1
Lay's Sweet Southern Heat Barbecue	0	0	0	0	0	0	1	0	1	1
Ruffles Cheddar & Sour Cream	0	0	1	0	0	0	0	0	0	1
Ruffles Sour Cream & Onion	0	0	1	0	0	0	0	0	0	1
Smartfood Popcorn	1	0	0	0	0	0	0	0	0	0
Sun Chips French Onion	0	0	0	0	1	0	0	0	0	0
Sun Chips Harvest Cheddar	0	0	0	0	1	0	0	0	0	1

seemed highly salient. Common piles included groupings of the two Sun Chips snacks (11.5% assignment rate), the Doritos snacks (12.0% assignment rate), the Fritos snacks (10% assignment rate), and all the Lay's chips (10% assignment rate). However, participants also made piles that varied with respect to flavors: a "classic/original flavors" pile (including Fritos Original, Lay's Baked Original, Lay's Kettle Cooked Original; 6% assignment rate) and a "barbecue" pile (including Doritos Jacked Chipotle BBQ, Fritos Bar-B-Q, Fritos Chili Cheese, Fritos Flavor Twists BBQ, Lay's Kettle Cooked BBQ, and Lay's Sweet Southern Heat Barbecue; 4% assignment rate). Finally, some participants made a pile that included all the "unhealthy" snacks (i.e., all the snacks except Lay's Baked Original and Kettle Cooked Original, Smartfood popcorn, and Sun Chips).

Exploring heterogeneity. The assignment rate probabilities presented in Table 5 represent the probability that a pile is made across the entire sample, irrespective of the consumer. Higher probabilities do not, however, necessarily represent higher probabilities that a single consumer makes such a pile. Such probabilities can be derived from the assignments per consumer; the probability that a consumer makes a pile is presented in Table 6.

In the present section, we discuss ways that researchers can explore the heterogeneity in sorting task data using summary piles. First, we discuss how (asymmetric) pairwise associations can be used to illustrate how the making of one pile influences the making of another. Second, we illustrate how the assignments of piles to summary piles can be used as input to clustering algorithms to identify segments of consumers who make similar piles.

Asymmetric pairwise associations. Prior research has shown that to understand the nature of competition among brands, it is often necessary to consider how associations may differ even between pairs of brands (e.g., DeSarbo and Grewal 2007). That is, in understanding market perceptions, it is important to understand not only the global asymmetry in preference across all products but also how the relationships between pairs of brands might differ (i.e., local asymmetry; see Ringel and Skiera 2016). We now turn to the asymmetry in covariation across assignment of summary piles, and how this information can be utilized to derive implications about the nature of the competition across types of single-serving snacks.

In Table 6, the first column of each row represents the unconditional probability that a consumer makes a pile similar to the given summary pile. For instance, the model predicts that 67.3% of consumers make a pile that included only popcorn and that 24.8% of consumers make a pile of barbecue-flavored brands. Each row in the matrix then represents the conditional probability of a consumer *also* making the summary pile labeled in the column, given that he/she makes the one labeled in the row. To illustrate, consider the "barbecue" summary pile further. The probability that, in addition to a barbecue pile, the same consumer also makes an "original flavors" pile is 92.0%, which is much higher than the base probability of 35.6% of a consumer making an original flavors pile (about three times more

Table 6
EMPIRICAL APPLICATION: CONDITIONAL PROBABILITY OF COASSIGNMENTS

	Proportion of Consumers	*Ruffles*	*Doritos*	*Fritos*	*Original*	*Barbecue*	*All Lay's*	*Sun Chips*	*Fry-Like*	*Unhealthy*	*Popcorn*
Ruffles	61.4%	1	64.50%	54.80%	45.2%*	32.3%*	50.00%	59.70%	75.8%*	3.2%*	74.20%
Doritos	65.3%	60.60%	1	77.3%**	18.2%**	6.1%**	78.8%**	74.2%**	72.70%	3.0%*	75.8%*
Fritos	56.4%	59.60%	89.5%**	1	14.0%**	3.5%**	87.7**	77.2%**	78.9%**	.0%**	80.7%**
Original	35.6%	77.8%*	33.3%**	22.2%**	1	63.9%**	11.1%**	36.1%**	66.70%	8.30%	61.10%
Barbecue	24.8%	80.0%*	16.0%**	8.0%**	92.0%**	1	.0%**	24.0%**	72.00%	.0%**	60.00%
All Lay's	56.4%	54.40%	91.2%**	87.7%**	7.0%**	.0%**	1	80.7%**	78.9%*	.0%**	77.2%*
Sun Chips	59.4%	61.70%	81.7%**	73.3%**	21.7%**	10.0%**	76.7%**	1	76.7%*	8.30%	75.50%
Fry-like	68.3%	68.1%*	69.60%	65.2%*	34.80%	26.10%	65.2%*	66.7%*	1	1.40%	73.90%
Unhealthy	8.9%	22.2%*	22.2%**	.0%**	33.30%	.0%**	.0%**	55.60%	11.1%**	1	33.30%
Popcorn	67.3%	67.60%	73.5%*	67.6%**	32.40%	22.10%	64.7%**	66.20%	75.0%**	4.40%	1

*Proportion is different from the base rate at the .05 level (two-tailed).
**Proportion is different from the base rate at the .01 level (two-tailed).
Notes: Each row represents the conditional probability of the pile in the column being sorted, conditional on the pile in the row being sorted. For example, the probability that a consumer makes a pile of all Ruffles chips is 61.4%. The probability that the same consumer also makes a pile of barbecue chips is 32.3%, higher than the base rate probability of 24.8% of making a barbecue pile. Given that a consumer makes a barbecue pile, however, the probability that the same consumer also makes a Ruffles pile is 80.0%.

likely). However, given that a consumer makes an original summary pile, the probability the same consumer also makes a barbecue pile is 63.9%. Thus, whereas there is evidence that consumers who discriminate brands that have barbecue flavors will also single out brands that have original flavors, discriminating brands that have original flavors does not necessarily suggest strongly that the consumer will single out brands that have barbecue flavors.

In general, the results suggest that consumers who focused on brands were more likely to single out all brands, and yet most consumers recognized the singular nature of the popcorn snack. Moreover, those who made a summary pile grouping all the brands that are unhealthy tended not to discriminate by brand but were likely to single out the original flavors. Although such associations are only bivariate, such that multivariate associations might better represent the complexity of the heterogeneity patterns in the data, they represent an easy way for managers to explore brand associations.

Post hoc segmentation analyses. For practitioners interested in understanding the patterns of heterogeneity, one post hoc analysis of managerial relevance might be to identify market segments (Wedel and Kamakura 2002). To illustrate how we can utilize the output of the proposed model to identify such segments, we clustered the 101 participants' according to the summary pile assignments, using Ward's hierarchical clustering. We find that a two-segment solution describes the data well ($N_1 = 39$, $N_2 = 62$); we present the proportion of segments' members being assigned summary piles in Table 7.

The two derived segments (S1 and S2) differ significantly regarding which summary piles best describe their data. More striking are the differences in the number of summary piles ($M_1 = 3.97, M_2 = 5.68$) and consumers' tendencies to differentiate snacks according to snack type (S1 consumers) versus brand (S2 consumers). Participants who focused on snack types (i.e., S1) made up 39% of the sample and were more likely than S2 consumers to group snacks by flavor type (e.g., classic, barbecue) and healthiness. Participants who focused on brands (i.e., S2) made up the remainder of the sample and were highly likely to group snacks according to the brand, including making Lay's, Doritos, and popcorn summary piles. These participants were unlikely to make a barbecue pile, a classic pile, or an unhealthy pile. Further t-tests from follow-up survey measures find that those who grouped by snack type (S1) were less likely to report choosing single-serving snacks by brand (4.95 vs. 4.03) than those who sorted by brands (S2).

At the end of the sorting study, we also collected measures relating to the snack items, participants' snacking habits, psychographic variables, and demographic (concomitant) variables. Although such personality and demographic variables are often difficult to successfully relate to segments (Wedel and DeSarbo 2002), it remains possible to attempt to use them in post hoc analyses to identify drivers of assignments of certain summary piles. To do so, we performed a series of two-sample t-tests comparing those who were assigned a summary pile with those who were not, as a function of the set of demographic and psychographic variables. Relating the covariates to the segments, we find no differences in gender, age, primary language, or the number of snacks in the study that participants were previously unaware of. We did, however, find differences in patterns of consumption. Participants who sorted by type (S1) were less likely to report purchasing in the category than those who sorted by brand (S2), and they were also less likely to report choosing snacks by brands more generally.

Model Comparisons

Comparison with LDA. For our application, we also utilized LDA to summarize the sorting data. Whereas there seems to be a significant improvement up to the addition of seven topics (i.e., summary piles), the effect of additional topics diminishes quickly after that. On average (with the number of topics/summary piles held constant), our proposed model predicted the data with 35.21% less error than LDA. The difference was particularly noticeable when the number of topics (i.e., summary piles) increased above seven. When the number of summary piles tested was greater than seven, the multiple executions did not produce the same fit/solution.

Table 7

EMPIRICAL APPLICATION: SAMPLE POST HOC CUSTOMER SEGMENTATION AND ASSIGNMENT OF SUMMARY PILES

	Popcorn	*Fry-Like*	*Ruffles*	*Doritos*	*Sun Chips*	*Fritos*	*All Lay's*	*Classic*	*Barbecue*	*Unhealthy*	*Mean Number*
Segment 1: snack type focus (N = 39)	51.3%	53.8%	71.8%	23.1%	30.8%	5.1%	.0%	76.9%	61.5%	23.1%	3.97
Segment 2: brand focus (N = 62)	77.4%	77.4%	54.8%	91.9%	77.4%	88.7%	91.9%	9.7%	1.6%	.0%	5.68
Significance (*p*)	<.1	<.01	<.1	<.01	<.01	<.01	<.01	<.01	<.01	<.01	<.01

It is possible that our inability to get stable results from LDA stems from our choice of hyperparameters that, at least in the case of our data, were quite difficult to determine a priori. In contrast to our model that only requires heuristics or knowledge to determine the number of summary piles, in LDA, there are three decisions that a user must make: α, β, and the "cutoff," which helps to determine a hard assignment of brands to summary piles. Whereas α represents a smoothing parameter and β a prior on the count of the number of times brands are sampled from summary piles, the choice of the "cutoff" is more arbitrary and is necessary for model comparisons between our nonparametric model and LDA. We first used the recommendations provided ($\alpha = 50/K$, $\beta = 200/J$, and a cutoff at 50%), but the results led to a degenerate solution. We then experimented with alternative tuning parameter values and settled with $\alpha = 5/K$, $\beta = 20/J$, and cutoff at 5%. This took considerable experimentation because α, β, and the cutoff essentially represent a three-dimensional search space. Yet, in our many attempts, our "best" LDA solution had 57% more error than our model. It was also subject to convergence issues.

For ease of comparison regarding the interpretation of the solution, we selected LDA's ten-topic (summary pile) solution, which we present in Table 8. We find some overlap with the solution proposed by our methodology; both solutions provide summary piles that group Cheetos and Fritos each in their own group. The solution presented in Table 8 also provides a grouping of the sour cream–flavored snacks (i.e., Lay's Sour Cream & Onion, as well as the two Sun Chips snacks), which our proposed methodology did not identify. However, some of the LDA's piles are difficult to interpret. For instance, the solution separates Lay's chips into three groups: one includes Baked Lay's and Kettle Cooked (presumably healthier varieties), and two others separate Lay's BBQ, Salt & Vinegar, and Sweet Southern Heat from Lay's Classic, Dill Pickle, Garden Tomato & Basil, and Sour Cream & Onion. These second and third summary piles are difficult to interpret. This solution also separates Doritos between cheese and "strong flavors," with Spicy Nacho Doritos being in both. Finally, a pile consisting of Chester's Flamin' Hot Fries, Fritos Original, and Funyuns is difficult to interpret.

Comparison with cluster analysis. We applied a clustering algorithm (e.g., Ward's [1963] method) on a pairwise count matrix converted into distances. Our results suggest that clustering alone produced, on average (within each level of K), 86.91% more mispredictions than our proposed methodology, and 31.01% more error than LDA. The resulting interpretation of the scree plot (and the average number of piles) would suggest the presence of six clusters. Unfortunately, the six-cluster solution provides clusters that provide little insight: (1) Cheetos and Chester's, (2) Doritos, (3) Fritos, (4) Lay's, (5) Ruffles, and (6) a mixed group (Funyuns, Smartfood, and both Sun Chips snacks). The Ward solution with ten clusters provides similar brand groupings: (1) Doritos, (2) Cheetos and Chester's, (3) Fritos, (4) Ruffles, and (5) Sun Chips. However, it does single out (6) Smartfood popcorn and (7) Funyuns. Finally, it also breaks down the Lay's snacks into three groups: (8) barbecue (Lay's Kettle Cooked BBQ, Lay's BBQ, and Lay's Sweet Southern Heat Barbecue), (9) healthy classics (Baked Original and Kettle Cooked Original), and (10) other Lay's snacks.

Although this ten-cluster Ward solution might seem reasonable at first glance, a thorough examination suggests that considerable heterogeneity is masked. To illustrate, consider the implications for Lay's Southern Heat Barbecue, a snack that in the clustering solution is placed in a barbecue pile consisting of only Lay's products. One could infer that the brand is not associated with other snacks like Doritos Jacked BBQ, Fritos Flavor Twists BBQ, or Fritos Bar-B-Q but is only associated with other Lay's products. Our results suggest that this is not the case. Lay's Sweet Southern Heat Barbecue is associated with three summary piles: Lay's (56.4% of consumers), barbecue (24.8% of consumers), and unhealthy (8.9% of consumers). Whereas the Doritos Jacked BBQ, Fritos Flavor Twists BBQ, and Fritos Bar-B-Q are not associated with Lay's Sweet Southern Heat Barbecue through being Lay's products, they are associated through being barbecue-flavored snacks. Using Ward's cluster analyses, one could have concluded that the association between Lay's Sweet Southern Heat Barbecue and the other non-Lay's snacks was nonexistent. Moreover, because no heterogeneity information would be available, one would have no indication of which associations are more salient to consumers. That is, unlike in our analyses above, one would interpret the clustering solution and conclude that all participants sort by brands, when, in fact, it seems that the percentage of consumers who do so is much lower (approximately 62%, according to our post hoc analyses).

Summary of Implications for the Corp

The Corp provides a considerable amount of shelf space to single-serving snacks, and through its use of Frito-Lay products, it can offer its consumers a large number of snacks that vary in brands (Doritos, Cheetos, Fritos, Lay's, Smartfood, Chester's, Ruffles), flavors (barbecue, spicy, cheese, regular), healthiness (fried, baked, popped, kettle cooked), and texture (potato chips, waffles, popcorn, fry-shaped snacks). In the past, the retailer has primarily grouped the items by brands.

Table 8
EMPIRICAL APPLICATION: SOLUTION FROM LDA

	Lay's Baked, Kettle Cooked, and Classic	*Lay's Pile 2*	*Lay's Pile 3*	*Doritos Cheese*	*Doritos Strong Flavors*	*Fritos*	*Pile 7*	*Cheetos*	*Sun Chips and Popcorn*	*Sour Cream*
Cheetos Crunchy	0	0	0	0	0	0	0	1	0	0
Cheetos Crunchy Flamin' Hot	0	0	0	0	0	0	0	1	0	0
Cheetos Puffs	0	0	0	0	0	0	0	1	0	0
Chester's Flamin' Hot Fries	0	0	0	0	0	0	1	0	0	0
Doritos Cool Ranch	0	0	0	1	0	0	0	0	0	0
Doritos Jacked Smoky Chipotle BBQ	0	0	0	0	1	0	0	0	0	0
Doritos Nacho Cheese	0	0	0	1	0	0	0	0	0	0
Doritos Spicy Nacho	0	0	0	1	1	0	0	0	0	0
Doritos Spicy Sweet Chili	0	0	0	0	1	0	0	0	0	0
Fritos Bar-B-Q	0	0	0	0	0	1	0	0	0	0
Fritos Chili Cheese	0	0	0	0	0	1	0	0	0	0
Fritos Flavor Twists BBQ	0	0	0	0	0	1	0	0	0	0
Fritos Original	0	0	0	0	0	1	1	0	0	0
Funyuns Original	0	0	0	0	0	0	1	0	0	0
Lay's Baked Original	1	0	0	0	0	0	0	0	0	0
Lay's Barbecue	0	1	0	0	0	0	0	0	0	0
Lay's Classic	1	0	1	0	0	0	0	0	0	0
Lay's Dill Pickle	0	0	1	0	0	0	0	0	0	0
Lay's Garden Tomato & Basil	0	0	1	0	0	0	0	0	0	0
Lay's Kettle Cooked Barbecue	1	0	0	0	0	0	0	0	0	0
Lay's Kettle Cooked Original	1	0	0	0	0	0	0	0	0	0
Lay's Salt & Vinegar	0	1	0	0	0	0	0	0	0	0
Lay's Sour Cream & Onion	0	0	1	0	0	0	0	0	0	1
Lay's Sweet Southern Heat Barbecue	0	1	0	0	0	0	0	0	0	0
Ruffles Cheddar & Sour Cream	0	0	0	0	0	0	0	0	0	1
Ruffles Sour Cream & Onion	0	0	0	0	0	0	0	0	0	1
Smartfood Popcorn	0	0	0	0	0	0	0	0	1	0
Sun Chips French Onion	0	0	0	0	0	0	0	0	1	0
Sun Chips Harvest Cheddar	0	0	0	0	0	0	0	0	1	0

For the assortment of 29 snacks provided, our results suggest that tremendous heterogeneity exists regarding which features are salient to consumers. Whereas a majority seems to differentiate snacks by brands, a nontrivial proportion of participants (approximately 38%) seems to be focused on other attributes that relate to texture, healthiness, and flavor. Our results suggest the type of snack is an important differentiator; despite having a similar flavor, cheese snacks (e.g., Doritos Spicy Nacho, Ruffles Cheddar, Cheetos) were rarely grouped together. Rather, the dominant flavor categorizations seem to have been barbecue and the more traditional, or classic, flavors. Moreover, we find that many consumers end up seeing all unhealthy snacks as similar to one another, regardless of type. In addition, the fry-shaped nature of some snacks stood out more than flavors (spicy, cheese, etc.), which could have formed the basis of piles.

Because considerable consumer heterogeneity exists, and because consumers may not easily find what they want when an assortment is organized by brand, several recommendations can be drawn from in-store experiments or subsequent analyses. For example, it may be desirable to continue organizing the main assortment of snacks by brand type. However, it may also be desirable to arrange the proximity of the snacks such that barbecue-flavored snacks are located near one another even if they are of different brands. It may also be desirable to organize a separate (smaller) shelf with the brands that are not associated with the unhealthy perception of most single-serving snacks (e.g., Lay's Kettle-Cooked, Lay's Baked, Ruffles). One could perhaps even single out Smartfood popcorn and place it not only in the single-serving snack assortment but also by the popcorn that is not sold as single-serving snack. Moreover, if the retailer expected a greater consumer focus on different flavor types (e.g., spicy, cheese, barbecue, salt and vinegar), it might consider providing additional cues in the environment (e.g., stickers) to help consumers find the flavor types they enjoy.

GENERAL DISCUSSION

We have introduced a flexible methodology for the extraction of summary piles (i.e., piles that best describe heterogeneous consumer sorts) from generalized sorting task data. The proposed methodology is flexible enough to accommodate consumers sorting different types of items (e.g., products, brands, objects, items, people) by a wide variety of criteria (e.g., similarities, dissimilarities) using either the single-card or the multiple-cards sorting task. The method does this without arbitrary conversions of piles into pairwise proximities or of similarities into pairwise distances, which are typically performed before choosing one of many available ad hoc clustering algorithms that hinder the detection of latent heterogeneity.

To illustrate the capabilities of both the methodology and the proposed estimation algorithm, we first report a Monte Carlo simulation based on a partial factorial experimental design. The results indicate that the method is robust and can identify known structures that include both the single-card and multiple-cards sorting tasks. Second, we report the results from a Monte Carlo simulation that provides additional evidence for the methodology's ability to recover heterogeneous sorting task data. We show not only that the model is able to do so but also that it consistently outperforms some comparable benchmark models, especially in the presence of heterogeneity. Third, we collect sorting task data about an assortment of 29 single-serving snacks from a local retailer. Our methodology obtains excellent in-sample and out-of-sample predictive ability, and the resulting summary piles are much easier to interpret than the solutions obtained by competing methods.

In our proposed methodology, we purposely avoid presuming a particular cognitive process on the part of the consumer. Not directly incorporating theory could be construed as a weakness of the proposed method. Although we agree that there is merit to theory-rich methodologies, we also believe there is merit to empirical methodologies that aim to explore data—and which can do so efficiently. We thus hope that the proposed model can be used to develop and refine theory in other models, in line with other data decomposition approaches that continue to play a vital role in the marketing literature and marketing practice (for recent examples, see Chung and Rao 2012; Kim et al. 2013; Nam and Kannan 2014; Ringel and Skiera 2016).

Our formulation entails a complex binary programming problem that involves many quadratic constraints. In fact, we show in Appendix B that the formulation results in a problem that is NP-hard. Given this complexity, we take special advantage of the mathematical model's structure to accelerate the identification of summary piles. As shown in Appendix A, we were able to design an effective optimization VNS heuristic that, in the context of our empirical application, can accelerate the optimization without affecting the optimal solution. We hope that VNS can be adapted to even more challenging combinatorial problems. Moreover, we believe there is additional merit in providing a methodological approach that can reveal insights in a matter of minutes and without deep technical knowledge. To this end, we have provided both the full MATLAB codes and compiled software on the first author's faculty web page.

LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

Conceptualization and Characterization of Latent Heterogeneity

Consumer heterogeneity exists when consumers differ in the benefits they seek or in their response to marketing strategies. Summarizing the extent to which such heterogeneity exists can be done in a variety of ways, including finite-mixture models, clustering methods (Wedel and Kamakura 2000), and continuous compound mixtures (Allenby, Arora, and Ginter 1998). In statistical models that accommodate heterogeneity, the researcher must make assumptions about the nature of the distribution of heterogeneity. To the extent that these assumptions are reasonable and that the model fits well, statistical inferences can be made regarding the nature of the latent heterogeneous groups that lead to the data generation.

Because we do not assume a particular data-generating mechanism for the sorts and only aimed at data reduction to aid interpretation, our model does not enforce a particular parametric framework and is more similar to nonparametric approaches, which can be particularly useful in understanding consumer behavior when stochastic assumptions are not met (Kamakura et al. 2005). We have thus purposely also avoided specifying a definitive form of heterogeneity or relationships among consumer piles (e.g., enforcing that for each consumer, no two observed piles can be assigned to the same summary pile). We did so for two reasons. First, a pile may be the subset of another because a consumer's set of piles reflects different levels of abstraction in the consumer's latent category structures (Blanchard 2011). Second, computationally, such a constraint severely restricts the range of the values of K (the number of summary piles) at which the model has feasible solutions. That is, with such a constrained added, K must be evaluated at $K \geq \max(c_i)$, which can put a large dependence on consumers who generate very large numbers of piles.

Although we do not explicitly *model* parametric heterogeneity, our model is able to *recover* meaningful heterogeneity in the data. In our Monte Carlo data sets, our data-generating mechanism involves a finite number of latent category structures (i.e., segments), and our model can recover the data quite well. Moreover, in our empirical application, we have shown through post hoc analyses (segmentation and the use of covariates) that patterns exist in combinations of summary piles being assigned to different unobserved groups of consumers. We also characterize heterogeneity as being related to broad types of categorizers, namely, those who focus on brands and those who do not. This contrasts with aggregate models, such as hierarchical clustering, which obscure such differences and would lead to an inference that consumers strongly sort by brands. We believe it would be interesting for future research to develop models that simultaneously identify segments of consumers that vary in terms of the similarity of the summary piles they employ.

Adapting LDA to Better Analyze Sorting Task Data

We have proposed a model and an optimization strategy that is uniquely suited to analyze sorting task data. Although analyses on such data have tended to be performed through more aggregate-oriented methodologies (i.e., cluster analysis or multidimensional scaling), recent developments in text analysis have also introduced methods that could be applied to sorting data. In the present work, we have used a formulation of LDA as an alternative model for the analysis of sorting task data through a very popular (and fast) implementation of Gibbs sampling. Although the results suggest a mixed predictive ability of LDA and limited interpretability of the solution provided, it is possible that such results are influenced by the choices we made in adapting LDA to the analysis of sorting task data.

LDA is a mixed-membership probability model that can help summarize the piles that are used to generate the piles observed in the data. However, given that typical LDA applications involve larger amounts of data, and much larger numbers of words (i.e., brands), a model in which hyperparameters are estimated using a hierarchical model would likely provide a better fit. Furthermore, LDA models typically have single words as the unit of analysis, such that brands (e.g., Cheetos Crunchy Flamin' Hot) could be tokenized (i.e., broken into multiple words) as input data. Whereas doing so would perhaps help LDA perform better due to a greater number of words being used as input, tokenizing would prevent us from being able to associate brands to piles (e.g., there may be multiple brands with both "Lay's" and "BBQ" in their names). We also note that tokenization cannot explain why our model performed better in the Monte Carlo simulation on data that were generated with binary elements (not expressions). However, we believe that recent developments of LDA in marketing (e.g., Jacobs, Donkers, and Fok 2016; Puranam, Narayan, and Kadiyali 2016; Tirunillai and Tellis 2014) might prove beneficial, particularly with the incorporation of individual-level heterogeneity.

APPENDIX A: REFORMULATIONS AND GLOBAL OPTIMIZATION STRATEGIES

Previous research has noted that sorting task data can be inherently "noisy" (Day, Shocker, and Srivastava 1979) in that consumers can differ greatly in the piles that they form. However, consumers often have similar perceptions and thus sometimes create identical piles. Identical piles represent redundant data for our optimization model because we do not enforce any relationship among the piles. We thus reduce the data to only unique piles and weight the piles as a function of the number of times they are made. In this appendix, we show why doing so does not affect the optimal solution.

To begin, assume that there exist two consumers (1 and 2) whose piles $l_1 = \omega_1\ (1 \le \omega_1 \le c_1)$ and $l_2 = \omega_2\ (1 \le \omega_2 \le c_2)$ are identical such that $\sum_{j=1}^{J} |y_{1,\omega_1,j} - y_{2,\omega_2,j}| = 0$. Then, it follows that the product of $x_{1,\omega_{1,m}}p_{mj}$ will be equal to $x_{2,\omega_{2,m}}p_{mj}$ for all values of m. If the products of $x_{1,\omega_{1,m}}p_{mj}$ and $x_{2,\omega_{2,m}}p_{mj}$ are equal, then it also follows that the misprediction obtained through $z_{1,\omega_1,j}$ and $z_{2,\omega_2,j}$ will be equal for all j. That is, if two piles are identical, they will be identified by the same summary pile and produce the same number of mispredictions.

By doing this, we can reduce the number of piles necessary to achieve the global optimum to that of unique piles N_u ($N_u \le N$) only if we account for the number of times each unique pile n (n = 1, ..., N_u) is used across all participants. Specifically, by introducing κ_n as the number of times unique pile n is made in the data, and by further replacing all i and l_i indices with n, we can solve the optimization problem of minimizing Z_R in Expressions A1–A7 to exactly identify the same solution with the same objective function as that of the larger problem Z in Expressions 1–8 in the main text:

$$\text{(A1)} \quad \min Z_R = \sum_{n=1}^{N_u} \sum_{j=1}^{J} \kappa_n z_{nj},$$

subject to

$$\text{(A2)} \quad \sum_{m=1}^{K} x_{nm}p_{mj} \le z_{nj} \quad \forall n = 1, ..., N_u; \quad \forall j = 1, ..., J \text{ such that } y_{nj} = 0;$$

$$\text{(A3)} \quad \sum_{m=1}^{K} x_{nm}\left(1 - p_{mj}\right) \le z_{nj} \quad \forall n = 1, ..., N_u; \quad \forall j = 1, ..., J \text{ such that } y_{nj} = 1;$$

$$\text{(A4)} \quad \sum_{m=1}^{K} x_{nm} = 1 \quad \forall n = 1, ..., N_u;$$

$$\text{(A5)} \quad \sum_{j=1}^{J} p_{mj} \ge 1 \quad \forall m = 1, ..., K;$$

$$\text{(A6)} \quad x_{nm},\ z_{nm} \in \{0, 1\} \quad \forall n = 1, ..., N_u; \quad \forall m = 1, ..., K;$$

and

$$\text{(A7)} \quad p_{mj} \in \{0, 1\} \quad \forall m = 1, ..., K; \quad \forall j = 1, ..., J,$$

where κ_n is now the number of consumers who make unique pile n in the original data. The reduced problem requires $(J + 1) \times (N - N_u)$ fewer binary quadratic constraints, and $(K + J) \times (N - N_u)$ fewer decision variables, yet it shares the same optimal solution of problem in Expressions 1–8.

Although minimizing Z_R through Equations A1–A7 substantially reduces the complexity of the problem, exact methods are still not practical because of the large number of nonconvex binary quadratic constraints. Metaheuristics address the problem of escaping from local optima. A local minimum x_L of an optimization problem is such that

$$\text{(A8)} \quad f(x_L) \le f(x), \quad \forall x \in N(x_L),$$

where N(x) denotes the feasible neighborhood of x, which can be defined in many ways, each one yielding a different neighborhood structure. In discrete optimization problems, a neighborhood structure consists of all vectors obtained from x by some simple modification. For instance, if x is binary, one neighborhood structure can be defined by the set of all vectors obtained from x by complementing one of its components. Another possible neighborhood structure can be defined as the set of all vectors obtained from x by complementing two complementary components of x (i.e., one component is set from 0 to 1 and the other changes from 1 to 0). A local search or improving heuristic consists of choosing an initial solution x and then moving to the best neighbor $x' \in N(x)$ in the case of $f(x') < f(x)$. If no such neighbor exists, the heuristic stops; otherwise, it is iterated. If many local optima exist for a problem, the range of values they span may be large. Moreover, the global optimum $f(x^*)$ may differ substantially from the average value of a local minimum or even from the best such value among many, obtained by simple randomly restarting a procedure (a phenomenon called the Tchebycheff catastrophe; Baum 1986). To escape local optima and the valleys which contain them, VNS exploits the idea of neighborhood change through the following observations:

Fact 1: A local minimum for one neighborhood is not necessarily a local minimum for another.

Fact 2: A global minimum is a local minimum with respect to all possible neighborhoods.

Fact 3. For many problems, local minima with respect to one or several neighborhoods are relatively close to one another.

This last observation, which is empirical, implies that a local optimum often provides some information about the global one.

Let N_t ($t = 1, ..., t_{max}$) denote a finite set of preselected neighborhood structures, and let $N_t(x)$ represent the set of feasible solutions in the t-th neighborhood of x. Given a real-valued objective function f, a local optimum x_L of a minimization problem with respect to N_t is

$$f(x_L) \leq f(x),\ \forall x \in N_t(x_L).$$

In the VNS framework, the neighborhoods used correspond to various types of moves, or perturbations, of the current solution; they are thus problem specific. The current best solution x found is the center of the search. When one is looking for a better solution x, a solution x' is drawn at random in an increasingly far neighborhood, and a local search is performed from x', leading to another local optimum x''. If $f(x'') \geq f(x)$, x'' is ignored and one chooses a new neighbor solution x' in a further neighborhood of x. If otherwise, $f(x'') < f(x)$, the search is recentered around x'' restarting with the closest neighborhood. If all neighborhoods of x have been explored without success, one begins again with the closest neighborhood to x, until a stopping condition (e.g., maximum CPU time, number of major iterations, tolerance) is satisfied.

The steps of the basic VNS are given in Algorithm 1, shown next. They combine deterministic (local search) and stochastic (shaking) changes of neighborhood.

Algorithm 1: VNS

Initialization: Select the set of neighborhoods structures N_t, for $t = 1, ..., t_{max}$, that will be used in the search; find an initial solution *x*.

repeat

$t \leftarrow t_{min}$;

repeat

Shaking: Generate a point x' at random from the t-th neighborhood of x (i.e., $x' \in N_t(x)$);

Local Search: Apply a local search method (see below) with x' as initial solution; denote with x'', the so obtained local optimum;

Move or Not: If the local optimum x'' is better than the incumbent x, move there ($x \leftarrow x''$), and continue the search with N_{tmin} ($t \leftarrow t_{min}$); otherwise, set $t \leftarrow t + t_{step}$;

until $t = t_{max}$

until a stopping criterion is met.

As the size of neighborhoods tends to increase with their distance from the current best solution x, close-by neighborhoods are explored more carefully than far-away ones. This strategy takes advantage of the three facts listed earlier. Indeed, it is often observed that most local optima of combinatorial problems are concentrated in few parts of the solution space.

In the context of our model, the local search procedure involves an iterative procedure conditionally minimizing Expression A1. We have two sets of variables to estimate: **p** (composition of the summary piles) and **x** (assignments to summary piles). In the first step, we minimize Expression A1 with respect to ***x***, holding **p** fixed to current values. To do so, we note that if the summary piles **p** are known in Expressions A1–A7, the optimal assignment of piles to summary piles (**x** variables) is trivially obtained by assigning $x_{nm} = 1$ for

$$m = \underset{\beta=1...K}{\text{argmin}} \left\{ \sum_{j=1}^{J} \left| p_{\beta j} - y_{nj} \right| \right\},$$

and $x_{nm'} = 0$ for all $m \neq m'$. That is, if the summary piles are known (variables **p** are fixed), then the assignment of consumer piles to summary piles can be made such that each pile is assigned to the one of K summary piles it is the most similar to. In the second step, we minimize Expression A1 with respect to **p**, holding ***x*** fixed to the last obtained values. Specifically, the optimal summary piles **p** (conditional on **x**) can be obtained by making $p_{mj} = 1$ if $\sum_{n:x_{nm}=1} \kappa_n y_{nj} > \sum_{n:x_{nm}=1} \kappa_n / 2$, and 0 otherwise. For the specific case in which $p_{mj} = 0$ for all $j = 1, ..., J$, p_{mj} is set to 1 for the index j with the largest amount of 1s. Once **p** and **x** are fixed, variables **z** are obtained by the closed formulas A2–A3, and the current solution value for Z_R can be obtained. These provide local optimality conditions, and the local search settles to a local optimum when applying both Steps 1 and 2 does not improve beyond a tolerance level.

Our *Shaking* procedure consists of moves that stochastically reassign individual piles to summary piles to perturb the existing solution. Thus, if the parameter t = 2 for *Shaking*, then two reassignments are performed; if t = 3, then three reassignments are performed; and so on. The selection of the pile n to be reassigned is done randomly, while its new summary pile m* is also randomly selected from {1, ..., K} with probability

$$\frac{\sum_{m=1}^{K} \kappa_n \mu_{nm} - \kappa_n \mu_{nm^*}}{\sum_{m=1}^{K} \left(\sum_{m'=1}^{K} \kappa_n \mu_{nm'} - \kappa_n \mu_{nm} \right)},$$

where μ_{nm} is the number of mispredictions made by using pile m to predict n. Finally, we note that unlike other metaheuristics (e.g., simulated annealing, Tabu search), VNS requires just three parameters to control neighborhood change, namely, t_{min}, t_{step} and t_{max}, which makes it easy to understand and, most importantly, easy to use. A common setting (which we use) is $t_{min} = t_{step} = 1$.

APPENDIX B: PROOF OF NP-HARDNESS

The problem in Expressions A1–A7 can be reformulated as follows:

(B1) $$\min_{x,p} \sum_{n=1}^{N_u} \kappa_n \min_{m=1,...,K} \left\{ \sum_{j=1}^{J} z_{nj}^m \right\},$$

subject to

(B2) $$p_{mj} - y_{nj} \leq z_{nj}^m \quad \forall n = 1, ..., N; \quad \forall m = 1, ..., K; \quad \forall j = 1, ..., J;$$

(B3) $$p_{mj} - y_{nj} \geq -z_{nj}^m \quad \forall n = 1, ..., N; \quad \forall m = 1, ..., K; \quad \forall j = 1, ..., J;$$

(B4) $$\sum_{j=1}^{J} p_{mj} \geq 1 \quad \forall m = 1, ..., K;$$

(B5) $$z_{nj}^m \in [0, 1] \quad \forall n = 1, ..., N; \quad \forall j = 1, ..., J;$$

and

(B6) $$p_{mj} \in \{0, 1\} \quad \forall n = 1, ..., N; \quad \forall m = 1, ..., K,$$

where N_u is the number of unique piles in the data and κ_n is the number of consumers who had pile n in their sorts. Constraints B2 and B3 in this reformulation are equivalent to constraints in Expressions 2–3 in the main text. As in the previous formulation, a variable $z^m_{nj} = 1$ if p_{mj} does *not* correctly predict unique pile n's jth item. Note that in the formulation, we have (temporarily) eliminated x from the problem entirely; yet its assignment can be recovered through the following closed expression:

$$\text{(B7)} \qquad x_{m^*n} = \begin{cases} 1 & \text{for } m^* = \underset{m=1,\ldots,K}{\operatorname{argmin}} \left\{ \sum_{j=1}^{J} z^m_{nj} \right\} \\ 0 & \text{otherwise} \end{cases}.$$

That is, once we assume z to be known, we can determine which summary pile produces the fewest errors for each pile in the original data.

To prove that our problem is NP-hard, we first note that objective function B1 is the sum of the minima of K linear functions, which is a piecewise-linear concave function. Second, we note that the minimization of piecewise-linear concave functions in a polyhedral set is NP-hard (Mangasarian 1978). As such, to prove that our problem is NP-hard, given it also involves the minimization of a piecewise-linear concave function in a polyhedral set, we have to express our constraints such that they define half spaces, and thus together provide a polyhedral set. This requires being able to replace constraint B6 by a constraint that states $p_{mj} \in [0, 1]$ without affecting the solution.

To do so, let us consider the following:

$$N_{m^*} = \left\{ n \in \{1, \ldots, N_u\} \,|\, m^* = \underset{m=1,\ldots,K}{\operatorname{argmin}} \left\{ \sum_{j=1}^{J} z^m_{nj} \right\} \right\}.$$

Thus, for a particular m = m*, the solution of $\sum_{n \in N_{m^*}} \kappa_n \min \sum_{j=1}^{J} z^{m^*}_{nj}$ subject to constraints B2–B6 can be obtained equivalently by solving $\sum_{n \in N_{m^*}} \kappa_n \min \sum_{j=1}^{J} |p_{m^*j} - y_{nj}|$ subject to B4–B6. The latter problem is trivially solved by choosing p_{m^*j} equal to the median value of $(y_{n_1 j}, y_{n_2 j}, \ldots, y_{n_{|N_{m^*}|} j})$, where $n_1, n_2, \ldots, n_{|N_{m^*}|}$ are the elements of the set N_{m^*}. Given that y is binary, these median values are also binary, and thus p is also binary regardless of constraint B6, which can be replaced by:

$$\text{(B8)} \qquad p_{mj} \in [0, 1] \quad \forall n = 1, \ldots, N; \quad \forall m = 1, \ldots, K.$$

For the specific case in which $p_{m^*j} = 0$ for all j = 1, ..., J, $p_{m^*j'}$ is made equal to 1 for the index j′ with the largest number of 1s to satisfy B4. In summary, given that all our constraints (B2, B3, B4, B5, B8) provide a polyhedral set and that our objective (B1) is the minimization of a piecewise-linear concave function and that such problems have been shown to be NP-hard (Mangasarian 1978), our problem is thus NP-hard as well.

REFERENCES

Alba, Joseph W., and Amitava Chattopadhyay (1986), "Salience Effects in Brand Recall," *Journal of Marketing Research*, 23 (November), 363–69.

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1998), "On the Heterogeneity of Demand," *Journal of Marketing Research*, 35 (August), 384–89.

Baum, E.B. (1986), "Toward Practical 'Neural' Computation for Combinatorial Optimization Problems," in *Neural Networks for Computing*, N.J. Denker, ed. New York: American Institute of Physics, 53–64.

Bijmolt, Tammo H.A., and Michel Wedel (1995), "The Effects of Alternative Methods of Collecting Similarity Data for Multidimensional Scaling," *International Journal of Research in Marketing*, 12 (4), 363–71.

Blanchard, Simon J. (2011), "A Methodology for Identifying Unobserved Categories When Consumers Assign Brands to Multiple Categories," doctoral dissertation, Pennsylvania State University.

Blanchard, Simon J., Daniel Aloise, and Wayne S. DeSarbo (2012), "The Heterogeneous P-Median Problem for Categorization Based Clustering," *Psychometrika*, 77 (4), 741–62.

Blanchard, Simon J., and Ishani Banerji (2016), "Evidence-Based Recommendations for Designing Free-Sorting Experiments," *Behavior Research Methods*, 48 (4), 1318–36.

Blanchard, Simon J., and Wayne S. DeSarbo (2013), "A New Zero-Inflated Negative Binomial Methodology for Latent Category Identification," *Psychometrika*, 78 (2), 1–19.

Blanchard, Simon J., Wayne S. DeSarbo, A. Selin Atalay, and Nukhet Harmancioglu (2012), "Identifying Consumer Heterogeneity in Unobserved Categories," *Marketing Letters*, 23 (1), 177–94.

Blei, David M. (2012), "Probabilistic Topic Models," *Communications of the ACM*, 55 (4), 77–84.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Chung, Jaihak, and Vithala R. Rao (2012), "A General Consumer Preference Model for Experience Products: Application to Internet Recommendation Services," *Journal of Marketing Research*, 49 (June), 289–305.

Coxon, Anthony Peter Macmillan (1999), *Sorting Data: Collection and Analysis*. Thousand Oaks, CA: Sage Publications.

Day, George S., Allan D. Shocker, and Rajendra K. Srivastava (1979), "Customer-Oriented Approaches to Identifying Product-Markets," *Journal of Marketing*, 43 (October), 8–19.

DeSarbo, Wayne S., and Rajdeep Grewal (2007), "An Alternative Efficient Representation of Demand-Based Competitive Asymmetry," *Strategic Management Journal*, 28 (7), 755–66.

DeSarbo, Wayne S., Kamel Jedidi, and Michael D. Johnson (1991), "A New Clustering Methodology for the Analysis of Sorted or Categorized Stimuli," *Marketing Letters*, 2 (3), 267–79.

Green, Paul E., and Vithala R. Rao (1969), "A Note on Proximity Measures and Cluster Analysis," *Journal of Marketing Research*, 6 (August), 359–64.

Griffin, Abbie, and John R. Hauser (1993), "The Voice of the Customer," *Marketing Science*, 12 (1), 1–27.

Griffiths, Thomas L., Mark Steyvers, and Joshua B. Tenenbaum (2007), "Topics in Semantic Representation," *Psychological Review*, 114 (2), 211–44.

Hamilton, Rebecca W., Stefano Puntoni, and Nader T. Tavassoli (2010), "Categorization by Groups and Individuals," *Organizational Behavior and Human Decision Processes*, 112 (1), 70–81.

Hansen, Pierre, and Nenad Mladenović (2001), "Variable Neighborhood Search: Principles and Applications," *European Journal of Operational Research*, 130 (3), 449–67.

Hansen, Pierre, Nenad Mladenović, and José A. Moreno Pérez (2010), "Variable Neighborhood Search: Methods and Applications," *Annals of Operations Research*, 175 (1), 367–407.

Irwin, Julie R., and Rebecca Walker Naylor (2009), "Ethical Decisions and Response Mode Compatibility: Weighting of Ethical Attributes in Consideration Sets Formed by Excluding Versus Including Product Alternatives," *Journal of Marketing Research*, 46 (April), 234–46.

Jacobs, Bruno, Bas Donkers, and Dennis Fok (2016), "Model-Based Purchase Predictions for Large Assortments," *Marketing Science*, 35 (3), 389–404.

Johnson, Stephen C. (1967), "Hierarchical Clustering Schemes," *Psychometrika*, 32 (3), 241–54.

Kamakura, Wagner, Carl F. Mela, Asim Ansari, Anand Bodapati, Pete Fader, Raghuram Iyengar, et al. (2005), "Choice Models and Customer Relationship Management.," *Marketing Letters*, 16 (3/4), 279–91.

Kim, Sunghoon, Simon J. Blanchard, Wayne S. DeSarbo, and Duncan K.H. Fong (2013), "Implementing Managerial Constraints in Model-Based Segmentation: Extensions of Kim, Fong, and DeSarbo (2012) with an Application to Heterogeneous Perceptions of Service Quality," *Journal of Marketing Research*, 50 (October), 664–73.

Lickel, Brian, David L. Hamilton, and Steven J. Sherman (2001), "Elements of a Lay Theory of Groups: Types of Groups, Relational Styles, and the Perception of Group Entitativity," *Personality and Social Psychology Review*, 5 (2), 129–40.

Mangasarian, Olvi L. (1978), *Characterization of Linear Complementarity Problems as Linear Programs*. Berlin: Springer.

Mladenović, Nenad, and Pierre Hansen (1997), "Variable Neighborhood Search," *Computers & Operations Research*, 24 (11), 1097–100.

Nam, Hyoryung, and P.K. Kannan (2014), "The Informational Value of Social Tagging Networks," *Journal of Marketing*, 78 (July), 21–40.

Peracchio, Laura A., and Alice M. Tybout (1996), "The Moderating Role of Prior Knowledge in Schema-Based Product Evaluation," *Journal of Consumer Research*, 23 (3), 177–92.

Punj, Girish, and David W. Stewart (1983), "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research*, 20 (May), 134–48.

Puranam, Dinesh, Vishal Narayan, and Vrinda Kadiyali (2016), "The Effect of Calorie Posting Regulation on Consumer Opinion: A Flexible Latent Dirichlet Allocation Model with Informative Priors," working paper, National University of Singapore.

Rao, Vithala R., and Ralph Katz (1971), "Alternative Multidimensional Scaling Methods for Large Stimulus Sets," *Journal of Marketing Research*, 8 (November), 488–94.

Ringel, Daniel M., and Bernd Skiera (2016), "Visualizing Asymmetric Competition Among More Than 1,000 Products Using Big Search Data," *Marketing Science*, 35 (3), 511–34.

Scott, M.J., and David V. Canter (1997), "Picture or Place? A Multiple Sorting Study of Landscape," *Journal of Environmental Psychology*, 17 (4), 263–81.

Spencer, Donna (2007), "Eurostar Card Sorting Case Study," Rosenfeld Media, http://rosenfeldmedia.com/card-sorting/eurostar-card-sorting-case-stu/.

Steckel, Joel H., and Wilfried R. Vanhonacker (1993), "Cross-Validating Regression Models in Marketing Research," *Marketing Science*, 12 (4), 415–27.

Steyvers, Mark, and Tom Griffiths (2007), "Probabilistic Topic Models," in *Latent Semantic Analysis: A Road to Meaning*, T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, eds. Mahwah, NJ: Laurence Erlbaum Associates.

Sujan, Mita, and James R. Bettman (1989), "The Effects of Brand Positioning Strategies on Consumers' Brand and Category Perceptions: Some Insights from Schema Research," *Journal of Marketing Research*, 26 (4), 454–67.

Takane, Yoshio (1980), "Analysis of Categorizing Behavior by a Quantification Method," *Behaviormetrika*, 7 (8), 75–86.

Tirunillai, Seshadri, and Gerard J. Tellis (2014), "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation," *Journal of Marketing Research*, 51 (August), 463–79.

Ward, Joe H., Jr. (1963), "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, 58 (301), 236–44.

Wedel, Michel, and Wayne S. DeSarbo (2002), "Market Segment Derivation and Profiling via a Finite Mixture Model Framework," *Marketing Letters*, 13 (1), 17–25.

Wedel, Michel, and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*. Norwell, MA: Kluwer.

Wedel, Michel, and Wagner A. Kamakura (2002), "Introduction to the Special Issue on Market Segmentation," *International Journal of Research in Marketing*, 19 (3), 181–83.

Yang, Chih-Chien, and Chih-Chiang Yang (2007), "Separating Latent Classes by Information Criteria," *Journal of Classification*, 24 (2), 183–203.