Mental Representation of Budgeting Categories

Lin Fei (lfei@chicagobooth.edu)
University of Chicago, 5807 S Woodlawn Ave
Chicago, IL 60637 USA

Daniel Bartels (bartels@uchicago.edu)
University of Chicago, 5807 S Woodlawn Ave
Chicago, IL 60637 USA

Abstract

Understanding how people mentally represent expenditures is crucial to understanding how they manage their money. In this paper, we report three studies that investigate people's representation of budgeting categories by asking people to categorize common expenditures of money (e.g., rent, dining out, etc.). We then examine the implications of these taxonomic representations of expenditures for how people selectively restrict their uses of money (e.g., when overspending on one item, for which other items would people choose to spend less). We found that there is consensus in people's representations of expenditures, and that both the category membership and taxonomic distance between items predict how people restrict their spending.

Keywords: mental representation; categorization hierarchy; mental budgeting; mental accounting; successive pile sort

Introduction

An important and relatively novel domain in which to investigate people's representation of concepts and categories is mental budgeting. Mental budgeting is the practice of mentally categorizing expenditures and restricting the amount of money allocated to each category of spending (Heath & Soll, 1996). The categorization of expenditures is therefore an essential part of the budgeting process (Henderson & Peterson, 1992). This paper investigates the structure of budgeting categories that people have in mind and the restrictions that this structure imposes on spending behavior.

Budgeting is a common practice that leads to important behavioral changes. Many people use budgets to control their spending (for review, see Zhang & Sussman, 2017; Zhang, Sussman, Wang-Ly, & Lyu, 2020), and much work has explored this restrictive effect. For example, people would be less willing to spend on a category if they had already consumed a typical item in the category (Heath & Soll, 1996). But what are the budgeting categories that people naturally have in mind as they budget? The study of mental budgeting behavior often makes assumptions about the categories people have in mind, such as food and entertainment (Heath & Soll, 1996; Cheema & Soman, 2006). Yet it remains unclear whether these are the categories that naturally come to people's minds, how common they are, and how they relate to each other. The current research attempts to address these questions and better understand these mental budgeting processes by asking people to sort expenditures into categories.

The investigation of representation in the context of mental budgets further allows us to deepen our understanding of the representation of concepts in general. Most, but not all of our current understanding of how we represent concepts is informed by methods where people are asked to form groups of natural kinds like plants (Medin, Lynch, Coley, & Atran, 1997) and animals (Lopez, Atran, Coley, Medin, & Smith, 1997). This research has documented both consensus of taxonomic representations within a culture, though the taxonomy differs between cultures and across occupations. For example, landscape workers tend to categorize trees in terms of landscape roles, while botanists' categories better resemble their field's biological taxonomy (Medin et al., 1997). Less attention has been devoted to studying people's representation of ad-hoc product categories, with some notable exceptions (Ratneshwar & Shocker, 1991; Ross & Murphy, 1999), where concepts are likely to be influenced by functional and goal-related features (Barsalou, 1985; Ross & Murphy, 1999). However, it is less clear whether people have consensus in their representations of concepts such as expenditures and how such representations influence subsequent behaviors. In this research, we examine how these functional objects (i.e., expenditures) are represented across individuals and how the representation subsequently influences people's behaviors.

Hierarchical Elicitation of Product Categories

The current research aims to investigate people's mental representations of budgeting categories, and we take the approach of asking people to successively sort a range of items that people spend money on (Boster, 1994; Medin et al., 1997). A successive pile-sort task requires participant to initially sort items into however many categories they would like. Then, participants further split and merge the categories that they deem appropriate to generate a hierarchy of categories. Doing so allows us to learn the categories of items that are most natural to participants (Boster, 1994).

Even though such hierarchical structure has not been investigated as extensively in expenditure and budget categories, it is reasonable to assume that there is a stable hierarchical structure on budgeting categories for several reasons. First, hierarchical structure prevails in concepts of the natural world (Murphy, 2002). For example, the concept

of mammals includes dogs and cats, while the concept of dogs includes bulldogs (Murphy, 2002). Similar to natural categories, budgeting categories have graded membership and people often can identify the purchases that are more and less typical of the category (Heath & Soll, 1996; Reinholtz, Bartels & Parker 2015). It is therefore likely that people also represent budget categories taxonomically as they do natural categories. In addition, some early research has provided support by finding that people have a hierarchically clustered representation of food items, and the representation involves ad-hoc functional groupings such as snack food (Ratneshwar & Shocker, 1991). Therefore, we think that a hierarchical structure is natural to people's representation of spending items.

A hierarchy allows for an approximation of the taxonomic distance between items. Specifically, an item will have distance 0 with itself, while the items grouped in the lower-level group (i.e., the group after splitting) will have a distance of 1; items grouped together in the initial sort but not in the lower-level group a distance of 2; those shared only higher-level category a distance of 3 while those never grouped together a distance of 4. Such taxonomic distance characterizes the relationship between two items, and therefore we hypothesize that it can predict downstream behaviors like budgeting where the relationship plays an important role.

In addition, in the hierarchy of categories, levels contain different amounts information. Lower-level categories contain more detailed information while the higher-level categories contain only the most overarching features that describe all the categories they contain (Murphy, 2002; Rosch, Mervis, Grey, Johnson, & Boyes-Braem, 1976; cf. Sloman, 1998). It is consequently often easier to verify that an item belongs to a lower-level level category than a higher-level category (e.g., "a pine is an evergreen" is easier to evaluate than "a pine is a plant"; Collins & Quillian, 1969). In the context of budget items, spending on an item may be more readily associated with other spending on a lower-level category than a higher-level category. Therefore, we hypothesize that items in higher-level categories will be less influenced by the expenditure.

Study 1

Study 1 investigates people's representation of budgeting categories and items. The study asks participants to perform a successive pile-sort task (Boster, 1994) of 64 budget items (determined through a pilot). Through this study, we aim to explore the representational consensus among participants, the dimensions that people categorize on, the categories constructed, and the similarity relationship between the expenditures.

Participants

Twenty-seven participants were recruited on Amazon Mechanical Turk (MTurk) to perform a successive pile-sort task of 64 common expenditure items. The average age of the participants was 38.4 and 37.0% were female.

Procedure

Participants learned that we were interested in how they budget their spending and performed a successive pile-sort task on a web interface. The participants saw a set of 64 cards labeled with items on the left-side of the screen. The order in which the index cards were presented was random. Participants put together the items that went together with however many budgeting categories as were natural to them (the initial category). After the groupings were recorded, participants were asked to review the categories they had just put together. They were then asked to merge the categories that were most similar to each other, forming the higher-level categories. Then, the interface restored the initial categories and asked participants to split out the items that were the most different from each other into smaller, lower-level categories[1]. Prior to this task, participants were familiarized with the interface by merging and splitting groups of simple symbols.

Analysis and Results

With the sorting task, we obtained each of the 64 items' hierarchical relationship with respect to all other items. To quantify the taxonomy, each participant's sorting was translated into distances between 0 to 4 and characterized as a 64 by 64 symmetrical distance matrix.

Consensus Result A Cultural Consensus Model (CCM; Romney, Weller, & Batchelder, 1986) was fit to assess consensus in people's groupings (Medin et al. 1997). Each participant's distance matrix was correlated with every other participant, which yielded a 27 by 27 correlation matrix. Then, a principal component analysis was performed on the inter-subject correlation matrix. The data would suggest that there was consensus according to CCM if the loadings on the first factor were positive, the first latent root (eigenvalue) was relatively large compared to the rest. As hypothesized, the loadings for all participants were positive, and the first eigenvalue was 12.23, which was large compared to the second eigenvalue of 1.39[2].

This result suggested that people had consensus in their representation of items that they spend on, which implied that people grouped items with a set of dimensions that was shared across individuals. For instance, people shared the understanding of why shampoo and toilet paper were closely related to each other, while shampoo and movies were not.

[1] Unlike Medin et al. (1997) that elicited merging and splitting iterativly, we only asked participants to merge and split once to keep the task concise considering the attention limit of online workers.

[2] The rule of thumb to evaluate whether the first eigenvalue is sufficiently larger than the second is to see if the ratio exceeds 3:1 (Weller, 2007)

Aggregate Level Categorization Since there was relative consensus, we further identified categories that were constructed on the aggregate level. The aggregate distance matrix was obtained by adding up all participants' distance matrices. A multidimensional scaling (MDS; Cox & Cox, 2008) with two dimensions on the aggregate distance matrix yields the loadings of each item onto the two dimensions (the loading matrix). To identify the categories, we performed k-means clustering on the loading matrix. We chose five as the number of clusters as five clusters allows for a small within-clusters sum of squares as well as variation in the categories. Further, five was the modal number of groups that participants constructed on the initial sort.

Applying five clusters to the loading matrix, we obtained the grouping as shown in Figure 1. In MDS, any orthogonal directions can be interpreted as the dimensions (Cox & Cox, 2008). Dashed lines represented one possible interpretation, where the two dimensions roughly mapped to the amount needed to spend on the item and how hedonic (or fun) it is to acquire the item (Hirschman & Holcrook, 1982). A dendrogram of the item distances based on a hierarchical clustering of the aggregate distance matrix is presented in Figure 2.[3]

From the groupings in Figure 1 and Figure 2, we observed that the five major groups that people readily have in mind were centered around necessary expenses (e.g., rent), groceries (e.g., juice), household products (e.g., shampoo), clothing items (e.g., shirts), and entertainment spending (e.g., dining out, airplane tickets). Though this grouping was likely to be specific to the budgeting context, as participants were explicitly instructed to think about how they budget, it offers important insight to this context. This methodology provided a better understanding of expenditures' membership to any specific category. We observed that participants' representation of the "entertainment" group was much broader than that assumed in the literature, and it ranged from books to travel expenses. In addition, eliciting the representations allowed us to better understand the categories readily accessible in people's mind and the nature of such categories. For example, the representation showed that both the taxonomy and the function of the expenditures influenced people's representation. Airplane tickets and hotel booking were categorized together because they served the same goal, while foods were mostly categorized together even though wine might serve different functions than juice.

Discussion

Study 1 provided a picture of people' mental representation of the budgeting categories, and this structure should impose a pattern on people's budgeting behaviors. Namely, work in mental budgeting found that people who deviate from budgetary restrictions should make more adjustments to items within the category (Heath & Soll, 1996; Henderson & Peterson, 1992). Therefore, we can verify the represented structure by using it to predict how people make adjustments to their spending. Study 2 assessed downstream implications of our recovered representations.

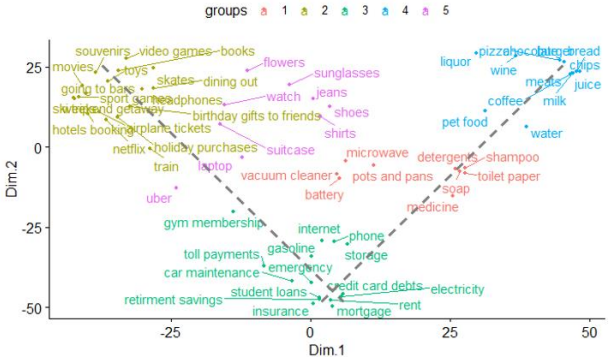

Figure 1: Spatial representation of items in two dimensions. Colors represent the five clusters.

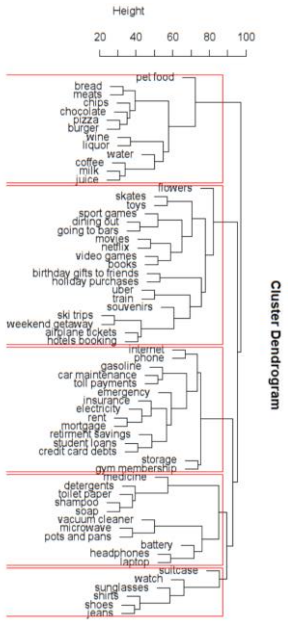

Figure 2: Average dendrogram for Study 1 showing the relationship between the expenditures.

Study 2

Participants

Three hundred and seventy-two MTurk workers participated in the survey and passed the attention check. The average age in the sample was 33.4 (range:18-77). In our sample, 172

[3] Note that the groupings in Figure 1 are slightly different from that in Figure 2, and this discrepancy is reasonable since the clustering in Figure 1 was performed based on the reduced representation of the items in two dimensions, while Figure 2 is based on the aggregated distances between items.

participants were male, 192 were female, four were other and three preferred not to answer.

Design and Procedures

We adopted our paradigm from the Study 2 of Heath and Soll (1996) by prompting people to consider a spending scenario on a focal item and eliciting their adjustment on several comparison items.

The study used a 2 (spending condition: overspend vs underspend on the focal item) by 5 (scenarios with different sets of items) within–subject design. Each set of scenarios consisted of four items, a focal item that they "spent" on (e.g., microwave) and three comparison items: (i) a "same-category" item—an item that was grouped together with the focal item in the initial sort by more than 90% of the participants in Study 1 (e.g., pots and pans), (ii) a "marginal category" item—one grouped together with the focal item by 50% of Study 1's participants (e.g., laptop), and (iii) a "different-category" item—one never grouped together with the focal item in Study 1 (e.g., weekend getaway).

Participants read that they budgeted for the focal item, and they were randomly assigned to have underspent [overspent] on the item with respect to the budget. They then rated their likelihood to adjust their spending on the three comparison items on a 7-point scale (1- decrease spending by a lot, 4 - no change, 7 - increase spending by a lot). Participants repeated this process for all five focal items. The budget and the respective deviation amount varied across all five scenarios for more generalizable results. Participants also finished the same successive pile sort task as Study 1 in addition to the rating task, and the order in which they completed the pile-sort task and the rating task was counterbalanced across participants. Then, they entered their age, gender, and income, and were compensated for the study.

Analysis and Results

Spending Adjustments with Respect to Aggregate Categorization In order to test for spending adjustment, we ran two regressions, one for overspending and one for underspending, regressing people's spending adjustment onto the comparision items' taxonomy (contrasted coded; 0 = different, 1 = marginal, 2 = same), order of the tasks (categorization-first or rating-first), controlling for the scenarios. Order of the tasks did not significantly affect the pattern of spending adjustment in either models and will not be discussed further (p's > 0.40).

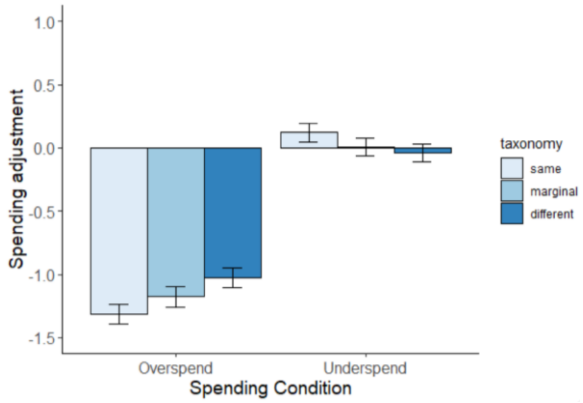

Figure 3: Spending adjustment in Study 2. Negative values indicate the spending will be adjusted downward while positive values indicate the spending will be adjusted upward. Error bars represent 95% confidence intervals.

The regression on overspending revealed that when people overspent on the focal item, they adjusted consumption downward for same-category items the most and the different-category items the least (slope = -0.14, $t(2783) = -5.06$, $p < 0.001$; Figure 3). Compared to the "different category" condition, we found a significant negative effect on adjustment of marginal-category items ($b = -0.15$, $t(2782) = -2.63$, $p < 0.01$) and a larger significant negative effect on adjustment of same-category items ($b = -0.29$, $t(2782) = -5.06$, $p < 0.001$).

In our underspending conditions, people were more likely to adjust consumption upward for items that were more likley to be grouped together with the focal item ($b = 0.08$, $t(2783) = 3.20$, $p < 0.005$), though the adjustment was smaller than that in overspending, which was consistent with previous findings (Zhang et al., 2020). Compared to the different category condition, there was a significant positive effect of same-category items ($b = 0.16$, $t(2782) = 3.20$, $p < 0.01$), and a similar, nonsignificant trend for the marginal-category items ($b = 0.04$, $t(2782) = 0.94$, $p = 0.34$; Figure 3). These results remained significant when we controlled for individual random effects or additional norming measures such as the average price of each item on offer.

Spending Adjustments with Respect to Individual Categorization We further examined these patterns by testing whether individuals' categorizations lead to a similar adjustment pattern. We dummy-coded the comparison items as whether they are in the same category as the focal item in each participants' first sort. Then, we performed the same two regressions, finding again that there was a negative adjustment of spending on same-category items when participants overspent on the focal item ($b = -0.17$, $t(2784) = -3.59$, $p < 0.001$) and a positive adjustment when they

underspent on the focal item(b = 0.12, t(2784) = 3.08, p < 0.002).

Discussion

Study 2 found that our recovered initial sort of budgeting categories conceptually replicates the past findings (Heath & Soll 1996) that a budgeting category restricts the fungibility of money. In previous work, participants were prompted to think about the membership of constituent items within a category (e.g., "entertainment spending"). Study 2 offered a more conservative test of this notion, because categorical labels of expenditures were not provided, and any priming of the category came from the expenditures themselves. Yet we still observed patterns that were consistent with the budgetary category constraint on both the aggregate and the individual level. This pattern also supported the claim that Study 1's categories aligned with how people represent expenditures, and therefore restricted the fungibility of money.

In Study 2, we only investigated expenditures in relation to the categories of initial sort. We did not study the difference between, for example, an item in the higher-level category and one that never shared the same category. Our recovered taxonomies can also further allow us to investigate some finer details of spending adjustment. Specifically, we can investigate how different levels of grouping impact one's spending adjustment. We predict that we should observe larger adjustments in spending for items with a smaller taxonomic distance (e.g., sharing a lower-level category) and smaller adjustments for items with greater taxonomic distance (e.g., not even sharing a higher-level). Study 3 investigated this hypothesis.

Study 3

Participants

One hundred and ninety-eight Prolific participants completed Wave 1. Of them, 171 (86%) people were eligible for Wave 2, and 168 participated. After screening by attention check, 161 participants were included for analysis. The average age in the sample was 31.4 (range:18-77), and 66 were male, 93 female, and two other.

Design and Procedure

Study 3 used two waves of data collection, separated by two days. In Wave 1, participants did the categorization task from Study 1. We then tailored the stimuli for each participant. In Wave 2, they took part in a 2 (spending condition: overspend vs underspend on the focal item) by 2 (scenarios with different sets of items) within-subject design. Each set of scenarios consisted of a focal item that they had "spent" on and four comparison items, one item each that (i) shared a lower-level category with the focal item, (ii) shared an initial category with the focal item, (iii) shared a higher-level category with the focal item, and (iv) was never categorized together with the focal item. We used the set focal items similar to the set in Study 2, and we generated comparison items that are maximally comparable to each other on additional norming measures we collected (e.g., how hedonic it was to spend on the item, how often do people budget this item, etc.).

Participants were not invited for the second part if none of the focal items had a four-level hierarchy (e.g., if they merged all the initial categories into one big category, or if they failed to find the two that could be merged). Eligible participants were invited back for the second part one day after the first part. They first were prompted for a price that they usually budget for the focal item. Then, they read about a scenario where they underspent [overspent] 30% on the item with respect to their budget. They then rated their likelihood of adjusting their spending on the four comparison items (lower, initial, higher-level, and different-category item) on a 7-point scale as in Study 2. Participants repeated this process for both focal items. After this task, participants completed some additional control measures that did not influence the main results.

Analysis and Results

Results from the overspending conditions mirrored those of Study 2. When people overspent on a focal item, they adjusted their spending less for items at greater taxonomic distances (b = 0.10, t(642) = 2.66, p < 0.01). Adjustment for the "higher-level" and "different" items was significantly less than that for the "lower-level" items (b's = 0.40, t(640) = 2.19, p < 0.01; b = 0.27, t(640) = 2.13, p = 0.03), and the pattern that the higher-level was adjusted marginally more suggested that money might have regained fungibility on the higher-level (M(high) = 3.41 vs M(different) = 3.27, t(160) = 1.54, p = 0.13). Spending on initial-sort items was directionally adjusted less, but this difference was not significant (b = 0.16, t(640) = 1.32, p = 0.20). For underspending, we did not observe differences across levels (F(3,640) = 0.31, p = 0.82; Figure 4).

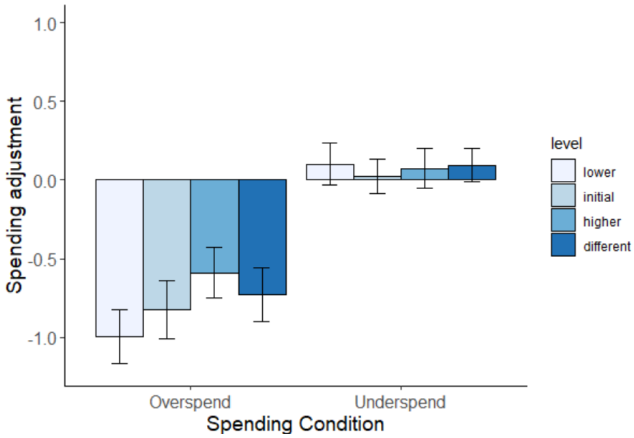

Figure 4: Spending adjustment in Study 3. Negative values indicate the spending will be adjusted downward while positive values indicate the spending will be adjusted upward. Error bars represent 95% confidence intervals.

Discussion

Study 3 found partial support for the hypothesis that people's spending adjustment differs across taxonomical distances. Specifically, adjustments were larger for items that are closer, but the effect is significant only when people overspent on focal items. The lack of a significant effect in underspending could be because it is generally difficult for people to report that they intend to spend more than they intend to spend (Zhang et al., 2020). This difficulty was also observed in Study 2 in that the slope for underspending was less than for overspending. Consequently, much larger power is needed for detecting the detailed differences across distances when there is underspending.

General Discussion

Across three studies, we investigated how people represent budget items with respect to each other. The categories we recovered from people in Study 1 were related to how people restrict the use of money (Study 2), and the strength of budgetary restrictions was a function of the taxonomic distance (Study 3).

Our methodology and findings deepen the understanding of how people represent expenditures in several ways. First, we recover taxonomies that correspond to standard budgetary categories in sensible ways (e.g., household products are considered more similar to clothing). This allows us to make more refined predictions on spending adjustment across budgeting groups. Second, our elicited hierarchies can translate into distance between items, allowing us to test for spending adjustment on items without imposing budget category labels. This approach better approximates how people typically think about spending and subsequent budget adjustments in reality, as they often are not reminded of the category in which spending belongs to.

We also hope that this research inspires more investigation of how people represent functional objects and how such representation affects downstream inferences and decisions, as several questions remain open and our studies have their limitations. One potential direction of future studies is to compare our recovered categories with other imposed categories and how the different categorical labels influence the magnitude or the direction of spending adjustment. Notably, in asking participants to make distinct groupings for all items, we did not allow cross-categorization, which is multiple categorical membership for each item (Ross & Murphy, 1999). This is a limitation of our studies, because expenditures and products might serve different functions in different contexts. Future studies could investigate when and how priming different usage occasions could influence people's representation and consequently spending adjustment. Additionally, we observed consensus among participants' categorization dimensions, yet we note that this result is not equivalent to the lack of heterogeneity in the grouping. People could still be generating categorizations of different sizes on the initial sort. In fact, in our data, people generate vastly different numbers of initial-sort categories (from 3 to 18, with a mode at 5-6). Addressing such heterogeneity has important implications as category sizes might influence people' spending behavior.

The current paper develops its theoretical framework from the categorization theories and applies the investigation methodology to a mental budgeting context. Providing an approach to investigating people's categories of expenditures, such methods could be extended beyond budgeting to countless other decisions about spending and saving, and potentially illuminate other structures of, for example, ad-hoc categories, with implications for better understanding downstream decisions.

Acknowledgements

This research is funded by the University of Chicago Booth School of Business. We thank hocag, several students, and several faculty members at Booth for their feedback and CogSci reviewers for detailed comments.

References

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

Boster, J. (1994). The successive pile sort. *Cultural Anthropology Methods*, *6*(2), 7-8.

Cheema, A., & Soman, D. (2006). Malleable mental accounting: The effect of flexibility on the justification of attractive spending and consumption decisions. *Journal of Consumer Psychology*, *16*(1), 33-44.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240-247.

Cox, M. A., & Cox, T. F. (2008). Multidimensional scaling. In *Handbook of data visualization* (pp. 315-347). Springer, Berlin, Heidelberg.

Heath, C., & Soll, J. B. (1996). Mental budgeting and consumer decisions. *Journal of Consumer Research*, *23*(1), 40-52.

Henderson P. W., & Peterson, R. A. (1992). Mental accounting and categorization. *Organizational Behavior and Human Decision Processes*, *51*(1), 92-117.

Hirschman, E. C., & Holbrook, M. B. (1982). Hedonic consumption: emerging concepts, methods and propositions. *Journal of Marketing*, *46*(3), 92-101.

Lopez, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology*, *32*(3), 251-295.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, *32*(1), 49-96.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, *35*(1), 113-138.

Murphy, G. L. (2002). *The big book of concepts*. MIT press.

Ratneshwar, S., & Shocker, A. D. (1991). Substitution in use and the role of usage context in product category

structures. *Journal of Marketing Research*, *28*(3), 281-295.

Reinholtz, N., Bartels, D. M., & Parker, J. R. (2015). On the mental accounting of restricted-use funds: How gift cards change what people purchase. *Journal of Consumer Research*, *42*(4), 596-614.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, *88*(2), 313-338.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382-439.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, *38*(4), 495-553.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, *35*(1), 1-33.

Weller, S. C. (2007). Cultural consensus theory: Applications and frequently asked questions. *Field methods*, *19*(4), 339-368.

Zhang, C. Y., & Sussman, A. B. (2017). The role of mental accounting in household spending and investing decisions. *Forthcoming in C. Chaffin (Ed.), Client psychology*, 19-07.

Zhang, C. Y., Sussman, A. B., Wang-Ly, N., & Lyu, J. (2020). How consumers budget. *Available at SSRN: https://ssrn.com/abstract=3739543*.