WORKING PAPER · NO. 2024-159

Consumers' Mental Representation of Expenditures: Implications for Spending and Savings Decisions

Lin Fei, Daniel M. Bartels, and Walter Zhang

SEPTEMBER 2024

THE UNIVERSITY OF CHICAGO
BECKER FRIEDMAN INSTITUTE FOR ECONOMICS

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Consumers' Mental Representation of Expenditures: Implications for Spending and Savings Decisions

ABSTRACT

People's mental representation of expenditures is crucial to how they budget. We propose that much like how people represent natural kinds (e.g., animals and plants), people represent expenditures in a hierarchical taxonomy. Across seven studies, supported by six norming studies and three pilots, we found evidence of a hierarchical representation of expenditures. We first recover people's mental representations using a successive pile-sort method that asks people to form hierarchies of categories with common expenditures (e.g., rent, dining out, etc.). We found that there is consensus in people's hierarchical representations of expenditures and that their representations are relatively stable over time. Further, we found that people's adjustment in their spending behavior can be predicted by the distance between items in their representation. Specifically, when people overspent on an item, they were more likely to spontaneously adjust spending for items closer in representation than those further away. We examine this spontaneous adjustment behavior using both lab studies and field data with 6.5 million grocery shopping trips over twelve years. The findings highlight the connection between mental representation and consumer behavior, and they emphasize the importance of studying concepts and categories in the context of consumption.

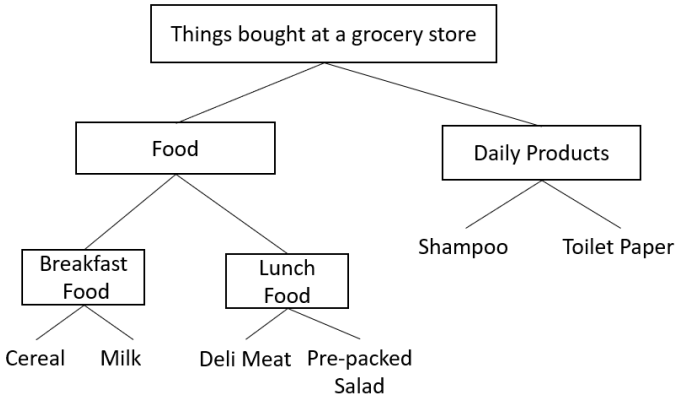
Keywords: Budgeting, Mental Accounting, Cognition, Financial Decision-Making

Consumers often maintain budgets. That is, they set restrictions on how much they wish to spend, and they behave as though money is nonfungible (Thaler 1985; 1999) and earmarked for specific functions (e.g., gas, food; Antonides et al. 2011; Hasting and Shapiro 2013, 2018; Heath and Soll 1996; Cheema and Soman 2006; Soman and Cheema 2011; Thaler 1999). For example, consumers might set a budget specifically intended for gas. When gas prices drop, consumers, who now have a gas-budget surplus, may use it to purchase premium gas (Hastings and Shapiro 2013). In a recent survey, 80% of people reported that they have formally or informally budgeted at some point in their lives, and the practice of budgeting is positively correlated with financial wellbeing (Zhang et al. 2020). Therefore, understanding how budgeting works is essential for understanding consumer financial decision making and consumer welfare.

Research on budgeting mostly theorizes that consumers group expenditures into a single level of budgeting categories (Heath and Soll 1996; Cheema and Soman 2006; Zhang et al. 2020; cf. Henderson and Peterson 1992). Some consumers might group items into relatively general categories such as “food” and “entertainment”, while others group to more detailed levels such as “groceries”, “dining out”, and “movies” (Zhang et al. 2020).

In this paper, we propose that instead of a single level, people mentally represent expenditures in hierarchical taxonomies. In other words, people represent expenditures in multiple, nested levels of categories where lower levels of the categories are more specific and higher levels are more general. For example, people might think of cereal as “breakfast food” before they think of it more generally as “food” (Markman, Brendl, and Kim 2007), and hence the “breakfast food” category is nested within the higher-level “food” category (Figure 1).

FIGURE 1: POSSIBLE REPRESENTATION OF EXPENDITURES



We hypothesize that the taxonomic distance between items is important for people's spending decisions in situations where they deviate from their budget. We define taxonomic distance as the level at which the expenditures are categorized together in consumers' taxonomies. We predict that when people's spending on an item deviates from their budget, they will adjust more on the items that are taxonomically closer to the purchased item. This prediction is important for marketers who are interested in understanding the type of products for which consumers adjust spending and the degree of adjustment when consumers encounter spending deviations such as holiday splurges. Further, past research that hypothesizes single level budgeting categories cannot account for this adjustment. In this paper, we adopt a method from cognitive anthropology to approximate the represented taxonomy of expenditures. Then, we use the taxonomy to predict how people adjust their spending in various contexts, including real shopping behaviors.

THEORETICAL BACKGROUND

Connecting Mental Representation to Mental Accounting

Mental accounting is grounded in the way that people categorize and represent concepts (Henderson and Peterson 1992). Mental accounting refers to the cognitive operations by which

prospects are evaluated with respect to some specific "topical" account instead of a "comprehensive" account (Thaler and Johnson 1990; Kahneman and Tversky 1984; Soman 2004; Thaler 1999; Zhang et al. 2020). For example, people are more likely to travel to save $5 on a calculator that originally cost $15 than when it costs $125. This is likely because they evaluate the $5 with respect to the total "topical" expenditure ($15 or $125) but not their total wealth. In doing so, people are associating money with different expenditure concepts, and how people represent the concepts will consequently influence how they spend the resources.

Researchers have connected mental representation with mental accounting (Arkes et al. 1994; Levav and McGraw 2009; Evers, Imas, and Kang 2022), and have provided insights on how people form and use mental accounts. For example, the source of money influences the types of mental accounts that are constructed. Money associated with negative emotions is less likely to be spent on hedonic purchases, and money associated with brand-specific gift cards is more likely to be spent on brand-typical items (Levav and McGraw 2009; Reinholtz, Bartels, and Parker 2015). Representation of expenditures also influences the membership of expenditures with respect to a mental account. Expenditures that have a high degree of feature-based (i.e., attribute) similarity are often integrated together in evaluating flows of gains and losses (Evers et al. 2022).

Similarly, ideas about mental representation have provided some insights on how people mentally budget (Henderson and Peterson 1992; Heath and Soll 1996). For example, researchers have found that people spontaneously categorize expenditures and organize them in categories when prompted to think about budgets (Cheema and Soman 2006; Sussman and Alter 2012). The expenditures in a category are often similar to each other in some respects. They may share similar features (e.g., electronics) or fulfill similar goals (e.g., entertainment; Felcher, Malaviya, and McGill 2001). Expenditures that are grouped together in a budgeting category have graded membership—the typicality of an expenditure within a category can differ (Reinholtz et al. 2015), which is another

prediction from research on mental representation. For example, shirts can be more typical of the clothing category than gloves (Heath and Soll 1996).

However, previous research has not connected the hierarchical nature of representation with mental accounting. To the authors' knowledge, only one paper has provided preliminary tests that people represent mental accounts hierarchically (Henderson and Peterson 1992). Specifically, Henderson and Peterson (1992) mentioned that people have a larger budget for higher-level categories but did not provide predictions on how people make spending decisions when there are hierarchical budgets. Additionally, the researchers did not make concrete predictions on how the hierarchy influences budgeting behavior. In this paper, we attempt to connect spending behavior with a taxonomic representation. We develop and introduce a method that can recover people's hierarchical taxonomy, and we generate and test hypotheses using the taxonomy recovered. In the sections below, we detail the rationale for proposing a hierarchical representation of expenditures, and we highlight additional contributions of introducing a hierarchical representation in budgeting.

Hierarchical Representation of Expenditures

There are several reasons why we think consumers group expenditures hierarchically. First, research suggests that people represent concepts of the natural world and of their everyday lives in taxonomies (Berlin 1992). People may group cats and dogs into the category "mammals" and then group mammals together with birds and fish into the superordinate category "animal". Organizing concepts in nested levels is cognitively efficient for storing and accessing information related to different objects (Collins and Quillian 1969; Smith 1978) and for generalizing category-level information to specific concepts (e.g., a dog inherits the properties of a mammal; Osherson et al. 1990). Research suggests that people hierarchically represent natural kinds, such as animals (Lopez et al. 1997), trees (Collins and Quillian 1969; Medin et al. 1997), as well as artifacts like furniture (Rosch et

al. 1976), and we suspect that people represent concepts of expenditures in terms of multiple, nested categories as well.

People represent objects in a way that reflects their experience and their knowledge of the world (Markman 2002; Medin and Atran 2004). For example, the taxonomy for trees that landscapers use tends to reflect differences between trees in shape and appearance, while botanists focus more on the scientific taxonomy. Each group's taxonomy reflects the knowledge of trees that is most useful for the way that each group typically interacts with and thinks about trees (Medin et al. 1997). Similarly, people's interactions with consumer goods likely also shape the categories they use to represent them. For example, grocery items are often subdivided into aisles that serve different functions (e.g., breakfast foods etc.) and further contains smaller groupings of products (e.g., cereal and breakfast bars). As consumers navigate through these shopping environments, they will likely internalize this organizational scheme into their representation. Several studies have suggested that consumers represent products in taxonomies (Ratneshwar and Shocker 1991; Nedungadi, 1990). Specifically, Ratneshwar and Shocker (1991) recovered a hierarchical structure from collecting pairwise similarity ratings for snack foods, and Nedungadi (1990) suggested that people represent fast-food brands in a taxonomy.

More generally, we think that people mentally organize expenditures into a hierarchical taxonomy by assessing the associations between expenditures. For example, similarity can influence the taxonomy. Ratneshwar and Shocker (1991) inferred the taxonomy of snack foods using pairwise similarity ratings between items. Notably, similarity is a notoriously multi-faceted construct that changes across contexts (Goodman 1972; Tversky and Gati 1979). Most theories of similarity, in the concepts and categories literature, treat similarity as an increasing function of shared features (e.g., like two dogs having floppy rather than pointy ears; Rosch and Mervis 1975) or relations (e.g., if Doug gives Dedre a gift, and Lance gives Reid a gift, Doug and Lance are relationally similar because they

occupy the same causal role; Gentner and Markman 1997; Hummel and Holyoak 2003) or both (Medin, Goldstone, and Gentner, 1993).

In the case of expenditures, while people may still rely on features or relations, people likely also represent expenditures based on how well the expenditures meet a particular consumption goal (i.e., ad hoc categories). For example, while potato chips and hash browns share the featural similarity of being made from potatoes, potato chips and popcorn might instead be closer in one's taxonomy as they often serve the same goal of snacking while hash browns might be closer to scrambled eggs. So, the mapping of similarity to one's taxonomy is not always straightforward, and, in particular, many mental accounts seem to be structured around goals more like ad hoc categories than like natural categories defined in terms of features and relations (see, e.g., Reinholtz, Bartels, and Parker 2015).

Contributions of a Taxonomy-Based Theory of Mental Budgeting

The notion that expenditures are represented in terms of a taxonomy addresses some limitations in the existing literature on mental budgeting. First, the current budgeting literature often assumes that people have predefined budgeting categories. Papers that investigate the type of budgets people construct often elicit those categories in a top-down fashion, asking directly for the budgeting categories people have in mind (Heath and Soll 1996; Zhang et al. 2020). However, people who do not formally budget and track their spending might not have constructed these categories. Relatedly, previous papers on mental budgeting also often prompted people with pre-defined categorical labels when investigating budgeting behaviors (Heath and Soll 1996; Cheema and Soman 2006). For example, participants would be examining their budgets for "food" and for "entertainment", when they are deciding how to categorize their expenditure of "dining out". These are researcher-defined categories and not every person will construct these specific budgeting categories (Zhang et al. 2020).

Therefore, we have limited insights into how people track and control their spending in absence of these given categories.

In addition, much of the budgeting literature suggests that expenditures have predefined membership within budgeting categories. Consumers tend to adjust more on items that are typical members of these categories (Heath and Soll 1996; Reinholtz et al. 2015) and are worse at evaluating expenditures that are exceptional (Sussman and Alter 2012). However, it is still unclear how consumers define the typicality or exceptionality of an expenditure. Further, consumers often are flexible in the way they categorize expenses (Cheema and Soman 2006) and might categorize a specific expenditure differently depending on the context ("dining out" can belong to "food" or "entertainment"). This suggests that consumers might not have clear category memberships for all expenditures.

A taxonomic representation, on the other hand, puts less emphasis on how expenditures relate to categories and focuses more on how expenditures relate to each other. In this paper, we recover the representation through a bottom-up approach that asks people to group expenditures into a hierarchy. With this method, we highlight that taxonomic distance influences budgeting without assuming anything about the categories consumers construct or the membership of expenditures. We expect these principles to apply to both people with formal budgets and to those who informally attempt to control their spending. Additionally, distances in a hierarchical representation can also reveal expenditures that are clustered together. This complements the current literature by providing insight into the expenditures that are often thought of together and the categories that people naturally construct. With a hierarchical representation, the recovered structure may also reveal which items are exceptional (e.g., items that are not grouped with others) and/or items that can be flexibly categorized (e.g., items that border two or more categories), which provides additional insight into people's budgeting behaviors.

The ideas in this paper add new, refined predictions about how people adjust their spending when they have overspent or saved money on an item. First, we extend the framework of budgeting beyond binary category membership. Heath and Soll (1996) theorized that consumers set budget categories, and then track and post spending with respect to the category. They assumed binary categories: When consumers overspent from a budgeting category, they were more likely to adjust spending on other items within the category than those outside of the category. However, people's purchase behaviors could be influenced by their out-of-category purchases (Chintagunta and Haldar 1998; Manchanda, Ansari, and Gupta 1999). For example, the purchases of soup and yogurt, which are presumably of different product categories, influence each other (Chintagunta and Halder 1998). Therefore, a framework intended to explain people's budgeting should extend beyond binary category membership terms, and our proposed hierarchical representation allows us to examine expenditures and spending adjustment more continuously.

More generally, our paper aims to expand on the theorization of mental accounting beyond the binary distinction of "topical accounts" and "comprehensive accounts" (Kahneman and Tversky 1984; Thaler 1999). Mental accounting is often seen as the behavior that "relates the consequences of possible choices to a reference level that is determined by the context within which the decision arises" (Thaler 1999). For example, people evaluate a discount of a calculator against the topical account of a calculator purchase rather than against their comprehensive wealth. These insights have proven extremely important, but our goal in this paper is to incorporate the idea of a taxonomy. A taxonomy implies that there could be multiple nested reference accounts and that people might recruit different reference levels depending on the expenditures they are evaluating. In other words, a more continuous spectrum than just "topical" and "comprehensive" accounts may be referenced if consumers have nested categories of expenditures.

Our prediction not only extends the budgeting literature but also provides insight into consumer research related to cross-product elasticities. Much research on cross-product elasticities theorizes that cross-product elasticities are the result of substitutable and complementary purchases (Kőszegi and Matějka 2020). The construct of taxonomic distance is relevant but independent from the constructs of substitutes and complements. When it comes to expenditures, people group items together for a variety of reasons, and each contributes to taxonomic distance. For example, people often group together complements—items consumed together (e.g., steak and steak sauce, pork ribs and barbecue sauce, airplane tickets and hotel bookings, bread and meat, shower gel and a loofah). They sometimes group substitutes together—where one expense could replace another (e.g., steak and pork ribs, steak sauce and barbecue sauce, rent and mortgage, shower gel and bar soap, movie tickets and Netflix). Sometimes, items are grouped together for other reasons, like retail adjacency or the degree to which two things relate to a single event (frozen pizza and ice cream, paper plates and facial tissue, cake baking mix and birthday candles).

We note that the theory of a hierarchical representation in budgeting is an expansion and is not an alternative account to the current budgeting literature. We replicate and extend the findings that consumers construct budgeting categories and that, within a predefined category, expenditure characteristics like typicality matter for spending behaviors. Our paper tests whether consumers also naturally construct categories in a nested fashion, and the taxonomies that consumers provide to us allow us to examine and address open questions within the budgeting literature.

Hypotheses and Experiments

We propose that consumers represent expenditures hierarchically and consequently adjust their spendings hierarchically when they deviate from their budget. Study 1 aims to inform the first part of

our claim: people represent expenditures hierarchically. While we cannot directly measure the structure of people's mental representation, we test some implications of this claim.

First, we test whether consumers have a shared understanding that some items are more closely related than others. For example, we could evaluate whether consumers share the understanding that shampoo and sunscreen are grouped together at a more specific, lower level than shampoo and movie tickets. If people provide similar taxonomies in the hierarchical sorting task, it supports the idea that they can represent expenditures this way, indicating some common understanding of why and how some items are closer than others.

H1: Consumers represent common expenditures similarly and reach consensus across their individual taxonomies.

We test this commonality between people's understanding by testing whether there is statistical consensus among their representations. The term "consensus" is adopted from the Cultural Consensus Model (CCM; Romney, Weller, and Batchelder 1986; Medin et al. 1997), a tool used in cognitive anthropology to test for the presence/absence of statistical agreement or commonality across individuals' representations. The CCM establishes consensus by testing whether participants' representations are positively correlated, and whether the first latent root of the interpersonal correlation matrix is large (Weller 2007). Statistically establishing consensus among participants would suggest that this representation is relatively natural to the majority of people. Past research has found that members of a culture have a consensual understanding of animal categories (Lopez et al. 1997) and consumers mostly agree on nested categories of snack foods (Ratneshwar and Shocker 1991).

We then test the stability of these hierarchical representations. If a hierarchical representation is unnatural to people, their taxonomy might be different across time because there could be large variances in how they choose to build a taxonomy out of expenditures each time they are asked to do so. Similar hierarchical representations of expenditures over time could suggest that the

representational form is natural to people. To test this, we measure the similarity of people's representation on the same set of expenditures over a span of 3 months in study 1b.

H2: People likely have similar representations of expenditures over time.

The second part of our proposal is that a hierarchical taxonomic representation predicts how much people will adjust their spending on items. Specifically, we predict that, all else equal, consumers' spending adjustment on an item will be influenced by the taxonomic distance (i.e., the level at which items are categorized together) between the item they just spent on and an item they are considering purchasing. When a consumer deviates from her budget (i.e., overspends or underspends), she would adjust spending more for items that are more closely related than those that are distantly related. If a consumer categorizes movies and scarves together in a higher-level category as discretionary spending and puts bread into a different higher-level category—necessities—the taxonomic distance between scarves and movies would be closer than that between scarves and bread. Consequently, we would predict that the when a consumer has a choice about adjusting spending, she would adjust more on movies (i.e., the closer item) than on bread (i.e., the farther item).[1] This implication stems from our proposal of a hierarchical mental representation, and we test this hypothesis using self-reported spending adjustments (study 2), incentivized choices (study 3), and actual grocery purchase behaviors (study 4).

H3: All else equal, when consumers deviate from their budget, their taxonomy of expenditures influences their spending and savings decisions even without reminders of the categories to which the expenditures belong. Specifically, when they deviate from their budget (i.e., over- or under-spend) on an item, they are likely to adjust more on items that are closer in taxonomic distance to the purchased item.

[1] We note that this hypothesis holds regardless of whether H1 obtains. If consensus does not obtain, we can use information about each person's taxonomy of expenditures to predict her spending and savings decisions. If consensus obtains, we can use the consensus taxonomy to make predictions about a group's decisions.

One noteworthy aspect of our prediction is that people may not need to be prompted with categorical labels to represent expenditures' category memberships, in contrast to the way that related research ideas have been tested in the past (Heath and Soll 1996; Cheema and Soman 2006). That is, we predict that when they face expenditures, people spontaneously adjust their spending in accordance with taxonomic distances even in absence of predefined categories. Our prediction therefore generalizes beyond consumers who have kept formal budgets (Zhang et al. 2020) to all consumers who wish to control and limit their spending.

While we expect larger spending adjustments for items closer in taxonomic distance to hold on most consumers, the differences in the size of adjustment can vary with other elements. One element that is likely to affect this pattern is how well the item primes its category inclusion relations. If the expenditure primes the representation well—in other words, is readily associated with its categories—then it is more likely to invoke the hierarchy and amplify the expected pattern. Previous papers report findings consistent with the idea that typical items may more strongly prime associations with the category that subsumes it (Rosch and Mervis 1975; Heath and Soll 1996; Osherson et al. 1990). In the paper, we hypothesize that typical items better prime the hierarchy of categories that the item belongs to. So, when people's spending deviates from their budget on a typical item, they are more likely to adjust on other items in the categories that include it. We test this moderator in study 3a.

Next, we note some additional constraints on our predictions. First, we are not arguing that taxonomic distance is the only nor the primary factor that drives spending and savings decisions. For example, among many other factors, necessity drives spending decisions—people cannot as easily adjust spending on their rent and phone bills (Zhang et al. 2020). We argue that controlling for the other plausible factors, taxonomic distance will influence spending behaviors on items that have room for adjustment. Second, large expense shocks, especially those that come from necessities, likely constrain spending and savings decisions. If someone struggles to pay rent, then she may likely cut

back on all the expenditures she possibly can. Therefore, we follow the previous literature (Heath and Soll 1996; Cheema and Soman 2006) and focus on adjustments in discretionary spending—that is, purchases that are not absolute necessities. Third, we aim to characterize how people adjust their spending and savings when considering alternatives,[2] as in typical shopping contexts, where they evaluate alternatives on display in stores or recommendations or listings on webpages, etc. Therefore, people are presented with alternatives in our experiments (studies 1-3), and people are shopping in grocery stores in our analysis of observational data (study 4).

TABLE 1: OVERVIEW OF STUDIES

Study	N	Key Finding(s)	Results
1a	27	Finds consensus across people's representations of expenditures.	Supports H1
1b	201 (First Wave) 131 (Completed)	Finds individuals' representations are relatively stable across 3 months.	Supports H2
2a	198 (First Wave) 161 (Completed)	Uses individual-level taxonomies. Finds decreased spending on taxonomically close items when overspend.	Supports H3 for spending decreases, not increases
2b	372	Uses aggregate-level taxonomies. Finds decreased spending on taxonomically close items when overspend.	Supports H3 for spending decreases and increases
2c	376	Finds effect of taxonomic distance when controlling for substitutability and complementarity using aggregate-level taxonomies.	Supports H3 for spending decreases and increases, controlling for substitutability and complementarity.
3a	402	When the focal item is typical, people are more likely to apply discounts to taxonomically close items.	Supports H3 using a rank DV and finds typicality as a moderator.
3b	307	People choose to apply discounts to taxonomically close items.	Supports H3 using consequential choices.
4	~6.5 million	People purchase more of an item when a taxonomically close item is on sale.	Supports H3 in grocery shopping trips over twelve years.

Note: The Web Appendix and all lab study materials (pre-registration, data, and analyses) are available on OSF (https://osf.io/4d23y/?view_only=0913b6c0bbdb44fd9008bc7b7870533e).

[2] Notably, we are not offering a theory of how people generate alternatives to evaluate, which is a different, important, and burgeoning research topic (see, e.g., cf. Bear et al. 2020; Mills and Phillips 2023; Morris et al. 2021).

In what follows, we first investigate how expenditures are represented. Then, we assess whether people spontaneously adjust their spending based on their representations even when they are not primed to think about budgeting categories, as is typical of retail environments.

STUDY 1: REPRESENTATION OF EXPENDITURES

Study 1a investigates whether there is consensus in people's representations (H1) of expenditures while study 1b investigates the consistency of the representations across 3 months (H2). Regardless of whether consensus is obtained, we will, in studies 2-4, further examine implications of how consumers represent expenditures.

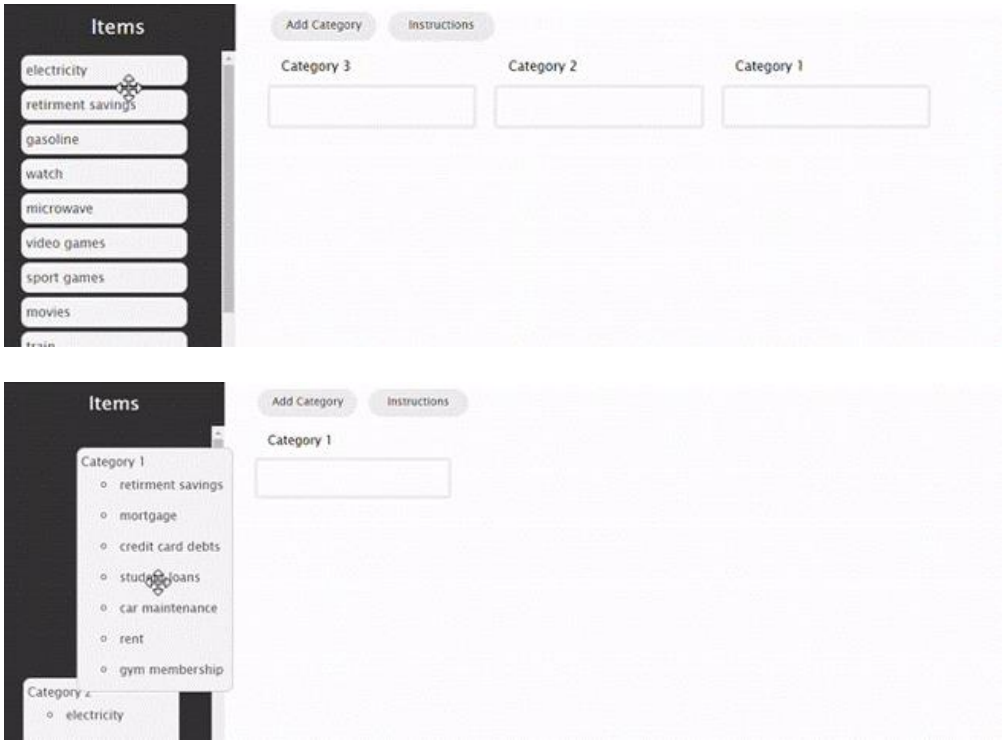
Study 1a

Participants & Procedures. Twenty-seven participants (all located in the US) were recruited on Amazon Mechanical Turk (MTurk). The average age was 38.4 and 37.0% were female. In this study, participants were prompted to think about expenditures by performing a successive pile-sort task (Boster 1994), which is widely used to examine people's representation of concepts (Lopez et al. 1997; Medin et al. 1997).

Participants performed the successive pile-sort on a web interface (Figure 2). Participants saw a set of 64 cards labeled with expenditures (e.g., rent, gas, movie tickets, etc.). The 64 expenditures were selected from a pilot study described in Web Appendix A, and the expenditures were presented in randomized order. Participants were instructed to group the expenditures into categories by which expenditures they thought would go together. They could construct a group of any size or add however many categories they would like to construct groupings natural to them. After they formed the initial level of groupings, the participants were asked to review the categories they had just put together for the next step, which was merging the initial groupings. They were asked to put together the categories that were most similar to each other, forming a higher level of less specific and larger categories. Then,

the interface implemented the splitting stage presenting all expenditures in one initial-level group. Participants were asked to split out the items from the initial grouping that were the most different from each other into smaller (lower-level) groups. They then repeated the splitting stage for all the initial groups. Participants were familiarized with the interface and the merging and splitting task with a simple set of symbol stimuli (Web Appendix B) before they proceeded to the main task.

FIGURE 2: SUCCESSIVE PILE-SORT INTERFACE



Note – Interface for the successive pile-sort task. Top: Initial and split sorting interface. Middle: Merging interface. Additional visualization videos are available on the Web Appendix.

Analysis and Results. We first translated participants' sorting into taxonomic distances. An item had a distance of 0 with itself, while the items grouped in the lower-level group (i.e., the group after the splitting stage) had a distance of 1. Items grouped in the initial group but not in the lower-level group had a distance of 2. Those that shared only the higher-level grouping had a distance of 3, while those never grouped together had a distance of 4. Each participant's sorting data were characterized as a 64 by 64 distance matrix, which represented all pairwise distances between items.

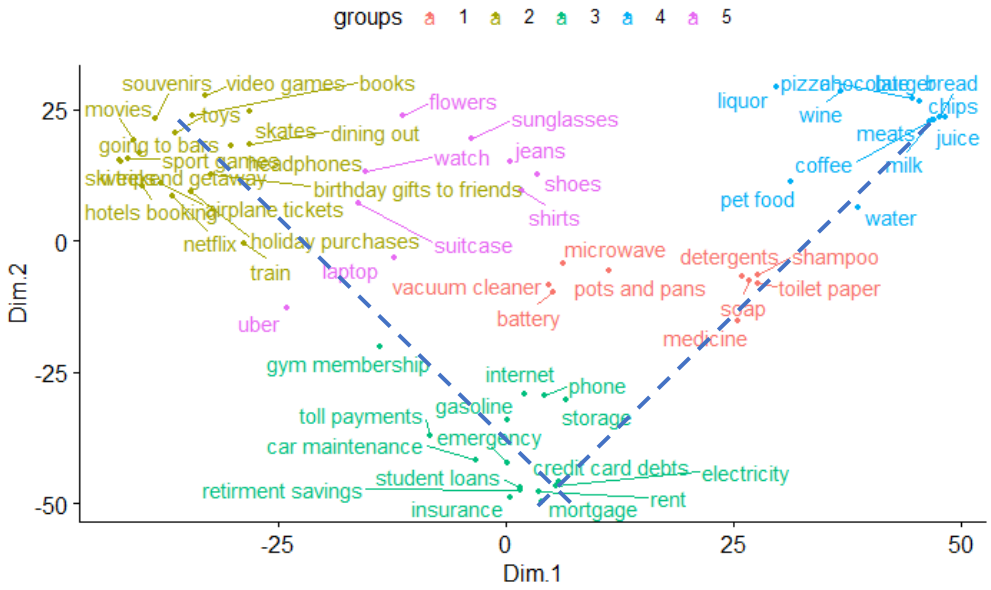
Consensus Result. We assessed whether there was consensus among the participants (H1). Consensus was assessed using the Cultural Consensus Model (CCM; Romney et al. 1986; Medin et al. 1997). Each participant's distance matrix was correlated with every other participant, which yielded a 27 by 27 correlation matrix. Then, a principal component analysis was performed on the inter-subject correlation matrix. According to CCM, there is consensus if the loadings are positive and the first latent root (eigenvalue) is relatively large compared to the rest (with a standard benchmark being 3:1; Weller 2007). We found that the loading for all participants was positive, and the first eigenvalue was 12.23, which is large relative to the second eigenvalue of 1.39. These results suggest that there is consensus in people's representations: people share the understanding that shampoo and sunscreen are closer together in representation than shampoo and movie tickets.

Aggregate Level Categorization. Because there was consensus, we further examined the aggregate representation with the implication that the structure and the distance between items would generally hold for most of our participants.

We first identified the initial categories that were constructed on the aggregate level. These categories reflect the general concepts that consumers have when budgeting, and they can be informative of the categorical labels that researchers and marketers might adopt. To obtain the categories, we first obtained the aggregate distance matrix by summing up all individuals' distance matrices. Then, we mapped the aggregate distance matrix onto two dimensions with multidimensional scaling (MDS; Kruskal 1978), which yielded the value of each item with respect to each of the two dimensions. To identify the categories, we performed k-means clustering on the MDS values. We chose five as the number of clusters for two reasons. First, five was the mode of the number of groups that participants constructed on the initial-sort (range = 3-18). Second, five clusters allowed for a small within-cluster sum-of-squares as well as variation in the categories.

Applying five clusters to the aggregate level distance matrix, we obtained a grouping presented in figure 3. The five groups that consumers readily have in mind seem to be bills (e.g., rent), groceries (e.g., juice), household products (e.g., shampoo), clothing items (e.g., shirts), and entertainment (e.g., dining out, airplane tickets). We conjecture that the two dimensions roughly correspond to 1) the typical spending on the item and 2) the necessity of spending on the item.

FIGURE 3: MULTIDIMENSIONAL SCALING REDUCTION WITH CLUSTERED GROUPS



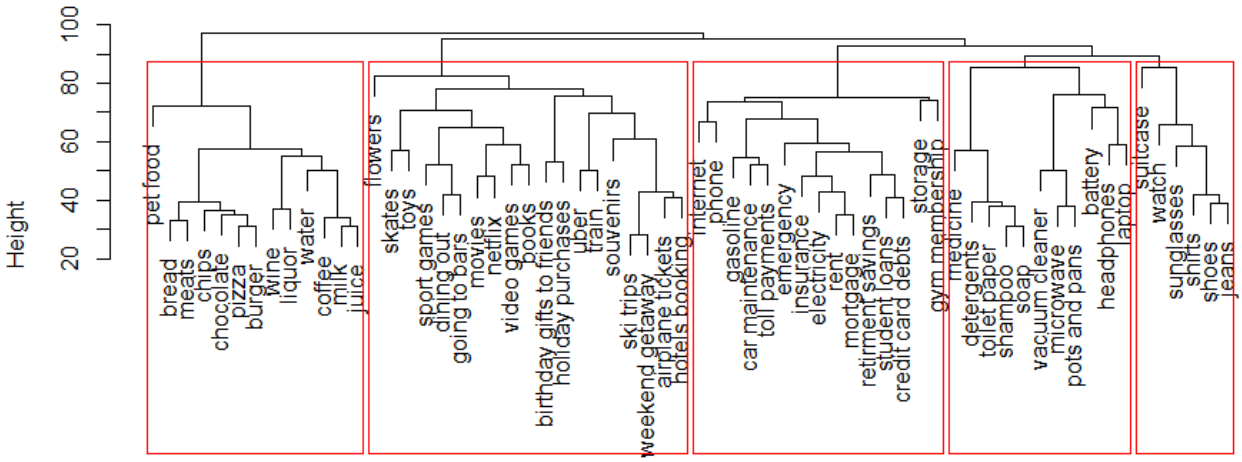
Note – Multidimensional scaling reduces the space of products to two dimensions, but the two dimensions need not be the X and Y axis. Any orthogonal directions can be interpreted as the dimensions (Cox and Cox 2008). Dashed lines represent one interpretation. The line sloping down may represent how hedonic a product is (Hirschman and Holcrook 1982) while the line sloping up may represent the amount of spending or the frequency of purchasing.

The dendrogram of the average item distances is presented in Figure 4.[3] The distances in both the dendrogram and the MDS preliminarily suggest that people's sorting of expenditures is related to but different from the concepts of substitutes and complements. For example, among the items that are revealed to be most closely associated with each other in taxonomic distance, burgers and pizzas seem to be substitutes while shirts and jeans seem to be complements. Detergents and soaps can be seen as

[3] Note that the groupings in Figure 4 are slightly different from those in Figure 3. This discrepancy is reasonable because the clustering was based on the reduced dimensions derived from MDS, while the dendrogram was based directly on the distance matrix. The two visualizations allow us to examine both the categories and the dimensions that people categorize on.

both, while sunglasses and watches are hardly either.[4] Therefore, taxonomic distance does not seem to be simply explained by substitutes or complements, which is a point we further elaborate on and systematically examine in study 2c. The dendrogram also reveals that among the groups, bills have high ratings on necessity and items in this group often have a rigid budget (Appendix C). As we noted earlier that our focus (like the other literature on budgeting) is on discretionary spending, we will concentrate on the rest of the groups when selecting stimuli for spending adjustment in the following studies.

FIGURE 4: DENDROGRAM OF THE PRODUCTS ON AGGREGATE LEVEL



Note – The red brackets represent the aggregate level categorizations. This grouping has a slight discrepancy with the MDS clustering because MDS maps the expenditures on two dimensions while the dendrogram only uses the aggregate distance matrix.

Study 1b

Participants & Procedures. Study 1b further tests if a taxonomic representation is natural to consumers by examining their representations of the same set of expenditures across 3 months. Two hundred and one participants (all located in the US) were recruited on Prolific to participate in a longitudinal study and perform the same successive pile-sort as study 1a twice, separated by three

[4] We verify the substitutes and complements relationship of our stimuli with a norming study. The details of the norming study are available in Web Appendix D.

months. The first wave of data collection was in January 2022 while the second wave was in April 2022, 13 to 15 weeks after the first wave. 146 (72.6%) participants started the second wave while 131 (65.2%) participants completed both parts. The average age in the first wave was 33.2 and 155 (77.1%) participants were female. The average age in the second wave was 34.8 and 105 (80.2%) participants were female.

Analyses & Results. We first tested whether there is selective attrition. We ran a logistic regression using our collected demographic variables (i.e., age, income, gender) to predict whether participants finished the second wave of data collection. Other than age, which marginally predicts finishing in the second wave (β = 0.03, z = 1.93, p = .053), no other collected demographic variables significantly predicted the participation in the second wave (*p*s > .05). Also, the model predicts attrition no better than a model that simply predicts the base rate of participation in the second wave (i.e., a model with only an intercept, without factoring in age, gender, and income; $\chi^2(4) = 2.98$, $p = .41$). No demographic variables predicted the correlation between representations in the two waves (|*t*|'s < .65, *p*s > .5).

Next, we examine the stability of people's representations across time. The correlation between the aggregate matrices in the two waves is 0.99. The average correlation between each individual's distance matrices is 0.54 (*SD* = 0.16, 25th Percentile = .41, 75th Percentile = .66). To interpret this value in context, we performed a permutation test. Each person who participated in the second wave was paired with a randomly drawn different person from the first wave and the correlation between their representations was calculated. The average correlation when one was not paired with herself is 0.42. We repeated this process of random pairing and correlation calculation 5000 times and obtained a bootstrapped 95% CI of [0.40, 0.43]. This indicates that people have higher consistency with themselves three months ago than with other people, which implies that the representations are relatively stable over time (H2).

Discussion

Studies 1a and 1b used a bottom-up approach to recover consumers' representations of expenditures by asking participants to successively sort various expenditures, and the results have important theoretical and practical implications. First, the recovered taxonomy of expenditures sheds light on the categories of expenditure that most consumers might entertain. Some of the five initial level categories we recovered (bills, groceries, household products, clothing items and entertainment spending) are similar to those used in previous research on mental budgeting (e.g., entertainment, food, and clothing, Heath and Soll 1996). Study 1a also provides insight into how these categories relate to each other and how items relate to categories. For example, the category of grocery foods is close to household products but far from entertainment, and suitcase is at the boundary of the "clothing" category and the "entertainment" category.

In addition, through the successive pile-sort method, we were able to gain insights into consumers' representations. We found that consumers reach a consensus in their representation of expenditures (consistent with H1). However, we note that consensus does not mean that there is no heterogeneity in consumers' groupings. In our studies, some consumers generated three initial level groups while others generated more than a dozen, but consensus can still be obtained as long as the distances between items across individuals are highly correlated. Even though participants varied in the sizes and number of categories they created, there was consensus on the relative taxonomic distances between expenditures according to CCM. Using the online successive pile-sort interface like the one we developed, we were able to assess and compare the representations of many individuals in a short amount of time.

We proposed that budgetary restraints are based, at least in part, on consumers' taxonomic representations. Specifically, we predict that when consumers deviate from their budget on a focal

item, they are likely to adjust more on expenditures that are taxonomically close to that item. Because the representation of expenditures reached consensus, we can test our prediction on both the individual level and the aggregate level. That is, how individuals represent expenditures can predict their individual spending adjustment, and the average distances between expenditures can predict the average adjustment pattern as well.[5] In the following studies, study 2a examines individual-specific spending and saving decisions. Study 2b examines aggregate spending adjustments while study 2c evaluates the influence of substitutability and complementarity. Studies 3 and 4 extend the domain and examine whether people's taxonomy can predict a variety of spending and saving behavior such as promotions application and grocery purchases.

STUDY 2: TAXONOMIC DISTANCE AND SPENDING ADJUSTMENT

Studies 2a to 2c examine the implication of a hierarchical taxonomy of expenditures on people's spending and savings decisions (H3). Specifically, we propose that after overspending or underspending on an item, people will adjust more on items that are taxonomically close. We adopt our paradigm from Heath and Soll (1996) by prompting consumers to consider a spending scenario on an item (i.e., a focal item). We then elicit their adjustment on several comparison items that have different taxonomic distances with the focal item.

Study 2a

[5] At the outset of this project, we did not know whether consensus would obtain in how people group together these expenditures. The finding of broad consensus makes it easier to study group responses, but, notably, larger aggregate taxonomic distances can come about multiple ways—for example, a distance of 2.3 might reflect (possibility #1) that nearly everyone groups two items together in the initial sort, or (possibility #2) that a large number of people group items together at the lowest level, and a large number of people group them together at a higher level or possibly not group them together at all. Our predictions about group spending and savings decisions doesn't change, but if possibility 2 (larger distances due to heterogeneity) obtains, it adds noise to the prediction.

In Study 2a, we examine whether participants adjust differently for items of different taxonomic distances using each participant's own taxonomic distances. In order to customize our stimuli to each participant's own taxonomy, we designed a two-wave survey. In the first wave, we asked participants to perform the successive pile-sort as in study 1, which allows us to customize the stimuli in the second wave to every participant's taxonomy. This allows us to remove heterogeneity on large taxonomic distances and make predictions that people adjust more for items closer in representation on the individual level.

Participants and Procedures. Study 2a was a two-wave survey. We pre-registered to recruit 200 participants for the first wave and only invited back participants who were eligible for the second wave. Specifically, only participants who established a clear four-level hierarchy were eligible for the second wave. For the second wave, we also pre-registered that we would only include participants who passed attention checks. One hundred and ninety-eight Prolific participants (all located in the U.S.) answered the first survey. Out of the 198 people, 171 (86%) people were eligible for the second wave, and 168 answered the second survey. 161 participants passed attention checks and were included for analysis. The average age in the sample was 31.4 (range:18-77), and 57.9% were female.

The first wave of study 2a used the same categorization task as study 1. Participants sorted the items into categories (initial level), merged the categories (higher level) and split the initial level categories into smaller categories (lower level). Two days after the first wave of data collection, participants who were eligible were invited back for the second wave.

In the second wave, we used a two-condition within-subject design: participants read one scenario where they underspent on a focal item and another where they overspent on the focal item. The focal item is different in each scenario. After reading that they had spent on a focal item, they indicated their desire to adjust their spending on four comparison items: a lower-level item, an initial-level item, a higher-level item, and a different-category item. A lower-level item is an item that was

categorized together by the participant with the focal item at the lower-level in the first wave (taxonomic distance = 1). An initial-level item is an item that was categorized with the focal item at the initial level but not the lower level (distance = 2), and likewise for the higher-level item (distance = 3). The different-category item was never categorized together with the focal item at any level (distance = 4).

The focal items were chosen from an ordered list: "dining out", "shirts", "microwave", and "coffee", because these items were close to the centers of four different clusters in figure 3.[6] Here, four focal items are chosen even though we recovered five groups in study 1 because, as we noted, our focus is on discretionary spending. We wanted to avoid absolute necessities (the fifth group), where people likely cannot adjust spending. Since each participant saw two scenarios, we used the first two items for which the participant generated a four-level hierarchy (i.e., there were items categorized at each level with the focal items). For each focal item, we then selected the four comparison items that are maximally comparable with each other. To do this, we first generated all possible focal-comparison combinations of the four distances for each participant. Then we selected the one combination that minimized the difference along additional norming dimensions. Norming dimensions were collected from a separate group of participants and included questions like how hedonic pending on an item is and how much one budgets for it (see Web Appendix C). To ensure there is a comparison item at every level, we did not invite participants for the second wave if none of the focal items had a four-level hierarchy (e.g., if they failed to split out an initial category).

More specifically, participants first entered a price that they usually budget for the focal item. Then, they read about a scenario where they either underspent or overspent by 30% on the item with

[6] The order corresponds to how spread out the clusters are in the MDS. “Dining out” is centered at the most spread-out cluster while “coffee” is centered at the least. Larger spread in the cluster indicates that the categorization is more varied, which will result in more differences in the comparison items chosen across individuals. Thus, we wanted to choose them more often.

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 (1- Underconsume a lot, 4 – No change, 7 – Overconsume a lot). Each participant rated their adjustments for a scenario where they overspent on the focal and another where they underspent on the focal, and the order of the scenarios was randomized. After this task, participants completed a tightwad-spendthrift measure (Rick, Cryder, and Loewenstein 2008) and a propensity to plan measure (money subscale; Lynch et al. 2010) for additional controls. They then entered their age, gender and income, and were compensated for the study.

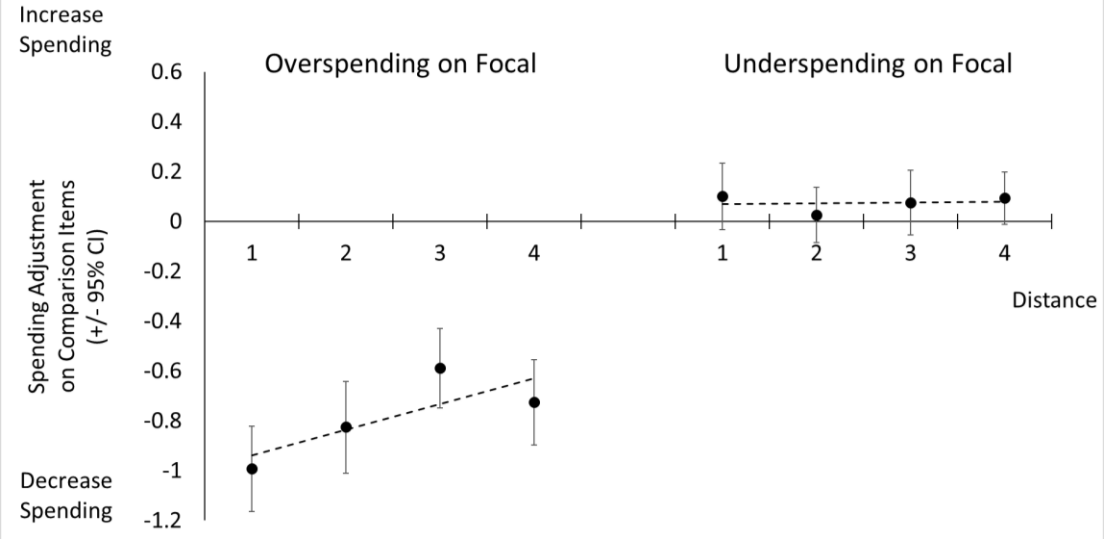
Analysis and Results. We assessed whether participants' sorts reached consensus. For all participants (N = 198) who finished the first wave (i.e., the categorization task), the first eigenvalue of the correlation matrix was considerably larger than the second eigenvalue (82.08 vs 8.50, which exceeds the benchmark of 3), and all the loadings were positive. These results suggest that there is consensus in people's representation of these expenditures (H1), replicating the finding of Study 1. And again, if consensus had not obtained, we could still make individual-level predictions about spending and savings decisions.

We then tested whether people are more likely to adjust on items at closer taxonomic distances when they deviate from the budget they set for a focal item (H2). In this case, when people overspent on the focal item, they should save more money on the items that are taxonomically close than those that are taxonomically distant. On the other hand, when people underspent on the focal item, consumers should increase their spending most on the items that are closest to the focal item.

To test the implications of taxonomic distance on spending adjustments resulting from budget deviations, we ran two regressions: one for overspending and one for underspending. We regressed consumers' spending adjustment onto the contrast-coded comparison items' taxonomic distance (1 = lower-level; 2= initial-level; 3 = higher-level; 4 = different category) while clustering the standard error

by participants. Consistent with H2, when people overspent on a focal item, they adjusted their spending less for items at greater taxonomic distances (β = .10, $t(157) = 3.57$, $p < 0.001$). However, taxonomic distance did not predict upward adjustment when there was underspending on a focal item (β = .0031, $t(157) = .13$, p =.89; Figure 5). The magnitude of overall adjustment was also smaller in the underspending condition than in the overspending condition (0.07 vs 0.78, $F(1,160) = 46.64$, $p < 0.001$), which is consistent with previous literature that there is an asymmetry in self-reported adjustment across events of overspending and underspending (Zhang et al. 2020). Specifically, people report greater intentions to reduce spending after overspending, relative to their intentions to increase spending after underspending.

FIGURE 5: SPENDING ADJUSTMENT ON COMPARISON ITEMS (STUDY 2A)



In Web Appendix E, we report the regression models with additional controls. We included individual differences measures and norming controls (e.g., how hedonic it is to spend on the item, how much one wants to budget for the item, etc. Web Appendix C). Web Appendix E also included different specifications of the models for study 2a, such as models including focal item fixed effects and individual random effects.

Consistent with H2, when people overspent on an item, they chose to save more money on expenditures that are closer in taxonomic distance. While most previous mental budgeting papers

explicitly prompt people with category labels and ask about spending adjustments, study 2a avoided the explicit prompt. But it is possible that study 2a heightened the accessibility of these categories, even though not explicitly presented, because all participants completed the categorization task before the spending adjustment task. So, in study 2b, we counterbalance the order of the categorization and spending-adjustment task. Further, because both study 1 and study 2a found consensus in participants' sortings, we use the aggregate taxonomy from study 1 to select items at different distances for the spending-adjustment task in study 2b.

Study 2b

Participants and Procedures. We pre-registered to recruit 400 participants and only included those who passed attention checks for analysis. Three hundred and seventy-two MTurk participants (all located in the U.S.) participated in the survey and passed the attention check. The average age in the sample was 33.4 (range:18-77), and 51.6% were female.

Study 2b used a 2 (spending condition: overspend vs. underspend on the focal item) by 5 (scenarios) mixed-subject design. Each participant read five purchasing scenarios, each with a different focal item. For each scenario, participants read that they either overspent or underspent on the item with respect to the budget. The budget and the respective deviation amount also varied across all five scenarios for more generalizable results.

Participants then rated their desire to adjust their spending on three comparison items on a 7-point scale (1- Underconsume a lot, 4 – No change, 7 – Overconsume a lot). The three comparison items were a close item, an intermediate item, and a far item. A close item is one that was categorized together with the focal item at the initial level by more than 90% of the participants in study 1. An intermediate item is one that was categorized together with the focal item at the initial level by 50% of the participants. A far item was never categorized together at the initial level with the focal item. The

comparison items were matched on two additional ratings collected through a norming study: (i) the likelihood of purchasing the item and (ii) the amount usually spent on the item (Web Appendix C). We included these measures because these considerations may influence people's intentions to adjust spending on a given item. The specific sets of stimuli are presented in table 2.

Participants also finished the same successive pile-sort task as study 1 in addition to the spending-adjustment task. The order in which they completed the pile-sort task and the spending-adjustment task was counterbalanced across participants. They then entered their age, gender, and income, and were compensated for the study.

TABLE 2: STUDY 2B STIMULI

Focal Items	***Comparison Items***		
	Close	Intermediate	Far
Chocolate	Pizza	Liquor	Toys
Juice	Milk	Pet Food	Jeans
Ski trips	Weekend Getaway	Holiday purchases	Medicine
Jeans	Shoes	Skates	Meats
Microwave	Pots and pans	Laptop	Weekend Getaway

Notes. – 5 sets of stimuli used in study 2. Norming ratings for each item are in Web Appendix C.

Analysis and Results. We first check whether the stimuli we selected from study 1 map onto participants' taxonomy in study 2b. In other words, we check whether this group of participants categorized the close item with the focal item most of the time, intermediate item half of the time, and far item almost never. Averaging across five scenarios, we found that the focal items were categorized with the same-category items 83.2% of the time, the marginal-category items 54.8% of the time, and the different-category items 12.8% of the time. The pattern was qualitatively the same for all five sets. This suggests that the taxonomic relationship between stimuli is consistent with study 1.

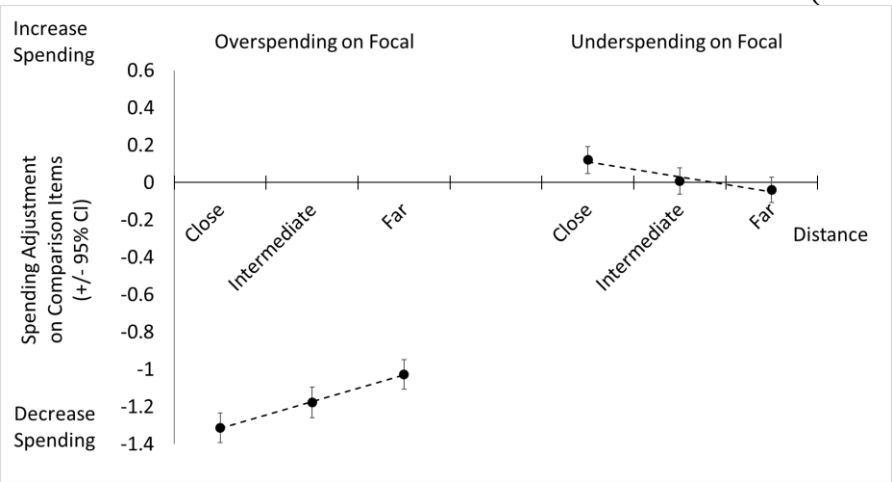
To test the effect of different levels in the hierarchy on spending adjustment (H3), we ran two regressions, one for overspending and one for underspending. We regressed consumers' spending adjustment onto the comparison items' taxonomic distance (contrast-coded; -1 = close; 0=

intermediate; 1 = far), order of the tasks (categorization-first = 0.5, rating-first = -0.5), and the interaction between the two, while clustering standard errors by participant.

When people overspent on the focal item, they reduced their consumption more for items that are more likely to be grouped together with the focal item (β = 0.14 , $t(360) = 5.99$, $p < .001$; Figure 6). This result conceptually replicates the past findings (Heath and Soll 1996) that people restrict their use of money for purchases within the same category. Similarly, when people underspent on their focal item, they adjusted consumption upward for items that were more likely to be grouped together with the focal item (β = -0.08, $t(358) = -4.24$, $p <.001$). The magnitude of adjustment in the underspending condition is smaller than those in the overspending condition (0.03 vs 1.17, $F(1,353) = 137.4$, $p < 0.001$), which is consistent with study 2a. Participants compensated for spending less or more than usual on focal items by spending or saving more on the taxonomically close items.

These results are unaffected by the order of tasks (main effects and interactions, $|t|$s < 1, ps >0.15). Further, they remained significant when we controlled for additional norming measures and individual random effects (Web Appendix E). We also report results using different ways of calculating distances including the distances from the multidimensional scaling and hierarchical clustering, as well as rank transformed distances. The qualitative results remain the same across changes in model specifications, controls and ways of computing distances (Web Appendix E).

FIGURE 6: SPENDING ADJUSTMENT ON COMPARISON ITEMS (STUDY 2B)



We also investigated whether people spontaneously adjusted their spending based on the taxonomic distance even in absence of the taxonomy-elicitation task. Specifically, we repeated the same analyses on the group that completed the spending adjustment task before the categorization task. As indicated by the lack of order effects reported above, the pattern was the same: people adjusted more when the comparison items were taxonomically closer to the focal item (Overspending: β =.12 , $t(192) = 3.75$, $p < .001$; Underspending: β = -.08, $t(192) = -2.76$, $p = 0.003$). In other words, even when people are not explicitly prompted to think of a taxonomy of expenditures, taxonomic distance relates to people's spending and savings decisions.

Because we also collected each individual participant's pile-sort, we further verified whether an individual's own taxonomy predicts her adjustment pattern. We recoded each comparison item with the taxonomic distance from each participant's own sorting (1 = lower-level; 2 = initial-level; 3 = higher-level; 4 = different category). Then, we regressed spending adjustment in overspending (and underspending) onto distances from individualized taxonomy, while controlling for the additional norming measures we collected. The result is the same: participants adjusted their spending more for taxonomically close items (Overspending: $\beta = 0.05$, $t(357) = 2.61$, $p = .005$; Underspending: β = -0.04, $t(355) = -1.89$, $p = 0.058$).

Study 2c

Study 2c examines the possibility that taxonomic distance is simply capturing substitutive relationships between products, as products that substitute for each other produce positive cross-product elasticities, like those we observed in study 2a and 2b. We theorize that taxonomic distance is distinct from substitutability, and study 2c assesses this claim empirically. Here we note again that similar usage occasions can contribute to perceived similarity of items within a category (Ratneshwar and Shocker, 1991), this version of the construct of "similarity" is nearly identical to substitutability.

But recall that people might group together not only substitutes (e.g., wine and liquor, train fare vs. ride share fare). They might also group together complements (e.g., sunglasses, sunblock, and beach towels, gasoline and car maintenance, headphones and laptops, or shoes, shirts, and jeans) and items found together in retail spaces (e.g., baby powder and children's Tylenol, peaches and bok choy).

Note that the pattern of adjustment that one would expect for complements is the opposite pattern of adjustment that one would expect for substitutes. Our proposal is that however people choose to group together expenditures, the closer the taxonomic distance, the more adjustment, and Study 2c was designed to discern the unique contributions of distance, substitutability, and complementarity. Study 2c uses a similar design as study 2b with a larger set of stimuli (24 focal products total), and controls for substitutability and complementarity ratings across product pairs.

Participants and Procedures. We pre-registered to recruit 400 participants and only included those who passed attention checks for analysis. Three hundred and seventy-six Prolific participants (all located in the U.S.) participated in the survey and passed the attention check. The average age in the sample was 37.88 (range:18-80) and 48.4% were female.

Prior to running study 2c, we conducted a norming study to collect the substitutability ratings and complementarity ratings (Web Appendix D). In this norming study, we first selected pairs of products that are often grouped together at the lowest level in previous studies (21 pairs). Then, we treated all of the items as focal and chose comparison items—one at each level that were maximally comparable to each other on the norming dimensions. We then collected substitutability and complementarity ratings for all product pairs (92 pairs total).

For study 2c, we selected three unique product pairs that are very high on rated substitutability, three that are very low on substitutability, three that are very high on complementarity, and three that are very low on complementarity, creating a total of 12 product pairs. We used all the items in the 12 pairs as the focal item for study 2c, resulting in a total of 24 focal items (i.e., spending scenarios). This

selection aimed to maximize variance in substitutability and complementarity ratings among items often grouped together at the lowest level, to best capture the effect of these cross-product relationships (full stimuli in Web Appendix D). Among the products we selected, there is a negative correlation between taxonomic distance and substitutability ratings (Pearson's $r = -0.72$, $p < .001$) and a negative correlation between taxonomic distance and complementarity ratings (Pearson's $r = -0.38$, $p = .002$). In other words, taxonomic distance is related not only to how much a pair of purchases substitute for each other, but also to the extent these purchases complement each other. Therefore, the variation along taxonomic distances does not simply equate to the variation in substitutability and complementarity patterns.

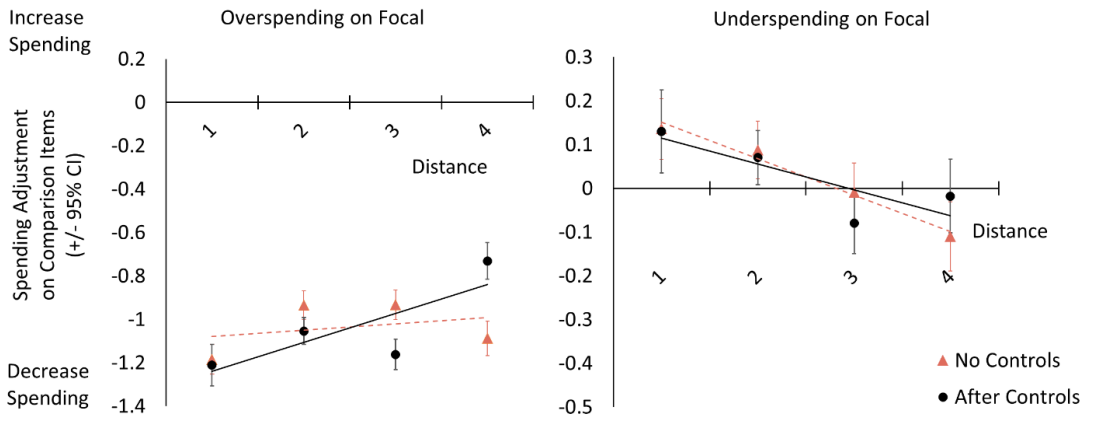
Study 2c used a 2 (spending condition: overspend vs underspend on the focal item) by 4 (set of 6 purchasing scenarios) between-subject design. Each participant read a set of six purchasing scenarios, randomly drawn from four sets. We predefined the sets such that all the focal items in a set were relatively different from each other. Each participant read that they either over or underspent on all the focal products in their past: "You realized that the [focal product] you bought last week was a lot more expensive (a lot cheaper) than you thought it was." Similar to study 2b, participants then rated their desire to adjust their spending on three comparison items on a 7-point scale (1- Underconsume a lot, 4 – No change, 7 – Overconsume a lot). The three comparison items were a close item, an intermediate item, and a far item.

Analysis and Results. For each spending condition, we preregistered two regressions. One predicts the likelihood-to-adjust ratings with 1) the taxonomic distance between the focal and 2) substitutability of the target and focal, controlling for several other dimensions of norming data that we collected (such as how hedonic the spending was and how much people wanted to control the spending on the item). The other regression is the same specification with just replacing the substitutability rating with the complementarity rating. Taxonomic distance was significant in all four regressions in the

expected direction. Specifically, when participants overspent on the focal, they adjusted down more on items closer in distance (Controlling for substitutability: $\beta = 0.167$, $t(176) = 4.68$, $p < .001$; Controlling for complementarity $\beta = 0.173$, $t(176) = 7.86$, $p < .001$; Controlling for both $\beta = 0.152$, $t(175) = 4.25$, $p < .001$). When they underspent on the focal, they adjusted up more on items closer in distance (Controlling for substitutability: $\beta = -0.107$, $t(188) = -3.60$, $p < .001$; Controlling for complementarity $\beta = -0.081$, $t(188) = -4.32$, $p < .001$; Controlling for both $\beta = -0.112$, $t(187) = -3.73$, $p < .001$) (full models in Web Appendix E).

Figure 7 presents the raw ratings (peach triangles) and the estimated marginal mean ratings (black circles) after controlling for both substitutability and complementarity, showing that even in this experiment, which was designed to powerfully manipulate substitutability and complementarity so as to detect their effects, taxonomic distance remains a uniquely significant predictor of people's behaviors.

FIGURE 7: RAW RATINGS AND ESTIMATED MARGINAL MEANS OF STUDY 2C



Discussion

Consistent with H2, studies 2a-c find that after people spent on an item, they intend to adjust more on expenditures that are taxonomically closer. This result holds regardless of whether we use stimuli selected from a participant's individual taxonomy or from an aggregate taxonomy, whether

people complete the categorization task before making spending adjustments, and whether we control for substitutionary and complementary relationships (study 2c). These findings are consistent with the notion that people reference a taxonomy of expenditures when making decisions about spending on alternatives, and the findings are not explained by the substitutability and complementary relationships between the expenditures.

The results of our studies extended the budgeting literature in a number of ways. First, in these studies, the categories that participants appear to be using were not prompted in the immediate context of the task. This suggests that the categories of expenditures we recover align with how consumers represent and restrict their use of money. Second, all studies used deviation from budget restraint in both directions (i.e., both overspending and underspending) and found patterns of adjustment in both directions. This finding further suggests that both spending and savings decisions can be predicted by people's taxonomies.

Finally, our studies provide unique insights on budgeting behavior beyond dichotomous categorical relationships for budgets. In standard treatments of mental accounting, theorists presume that consumers are applying either a topical account (i.e., when two things are in the same category) or a comprehensive account (i.e., when they're not). Our studies find that people make differential adjustments on spending depending on the level that the items are categorized together, which can be at the lowest level or at a higher level (study 2a).

STUDY 3: TAXONOMIC DISTANCE AND POTENTIAL MODERATORS

Study 2 examined the connection between people's represented taxonomy and their spending adjustment intentions, but all studies still asked participants to categorize the expenditures.[7] In study 3,

[7] For half of Study 2b's participants, the categorization task followed the decision task. For these participants, the categorization task could not have had a causal impact on their spending adjustments.

we examine people's spending adjustment behaviors across a variety of spending decisions without ever doing the categorization task. Additionally, we aim to test several potential moderators of the patterns we observe, and we incentivize the choices that the participants are making. Note that in studies 2a and 2b, we found relatively limited increases in spending intentions when people saved money on the focal item. To further investigate these settings, we now focus our investigation on behaviors after underspending on the focal.

In studies 3a and 3b, participants read that they purchased a focal item, and the focal item came with a discount that they could apply to a comparison item of their choice. We hypothesize that a product purchase would prime the more specific budgeting categories associated with the purchase. Consequently, the purchase would more strongly influence the savings decisions of a close item than a far item, and people would be more likely to apply promotions to a close item. Many papers in consumer decision research find that people's preferences are significantly affected by elicitation methods (Bettman, Luce, and Payne 1998; O'Donnell and Evers 2019). So, as a test of generality, whereas Studies 2a-c used ratings, Study 3a measures preferences through ranking, Study 3b uses consequential choices.

To select expenditure items that people can apply promotions on, we ran two pilots on a variety of items that people can purchase in large wholesale stores and identified 50 new items to use as stimuli for study 3 (Web Appendix F). Then we ran a norming study to establish the aggregate distance between the items (Web Appendix G).

Study 3a

Participants and Procedures. Four hundred and two MTurk participants (all located in the US) responded to the survey while 356 passed the attention check. The average age in the sample was 40.6 (range: 18-73), and 44.1% were female.

Participants assumed that they were purchasing all items on a shopping list of four items. They learned that one of the items (the focal item) that they purchased comes with a promotion—a price discount to be applied to another item on the list. The three other items on the list are of close, mid, and far distances to the focal item. The distance label is defined the same way as study 2b: a close item is one that was categorized together with the focal item at the initial level by more than 90% of the participants in the norming study. An intermediate item is one that was categorized together with the focal item at the initial level by 50% of the participants. A far item was never categorized together at the initial level with the focal item. Then, they ranked the items by how much they wanted to apply the promotion to each item (1 - I want to apply the promotion to this item the most; 3 - I want to apply the promotion to this item the least). We expected most participants to apply the promotions to the comparison item that is the closest in their representation of these expenditures.

Study 3a used a 2 (promotion magnitude: large – 40% off vs small – 10% off) by 5 (scenarios) by 2 (focal typicality: typical vs atypical) mixed-subject design. Each participant read five scenarios, and the promotion magnitude and the focal typicality were randomized for each scenario. Each scenario consisted of a focal item that participants had already purchased and three comparison items that they could apply a discount promotion to: a close item, an intermediate item, or a far item. We collected typicality ratings in a separate norming study, where participants rated the typicality of expenditures with respect to the initial categories they constructed in a successive pile-sort task (Web Appendix G).[8] While we expect that a more typical focal product would better prime the taxonomy, we also include the manipulation of the promotion magnitude because a large promotion (i.e., underspending from budget) might also activate the taxonomy more strongly.

TABLE 3: Study 3a STIMULI

Focal	***Comparison***

[8] Once again, we did not prompt participants with categories, but due to the high level of agreement in categorization, some items emerged as more typical than others on average.

	Close	Mid	Far
gourmet chocolate truffles ***meats***	wine	gas	toys
sunglasses ***jeans***	backpack	lamp	restaurant gift cards
potted plants ***lamp***	pots and pans	books	sunglasses
diapers ***toilet paper***	detergents	stationery	pizza
restaurant gift cards ***movie tickets***	streaming services	gourmet chocolate truffles	pet food

Notes. – There are 5 sets of stimuli used in study 3a. Red indicates that the focal item is typical of its category.

Analysis and Results. To test the effect of taxonomic distance on people's preferences for applying the promotions, we ran an ordinal logistic regression (McCullagh 1980). We regressed consumers' rankings of the items onto the items' aggregate distance from the focal (centered at 0), contrast-coded magnitude (0.5 = 40% off; -0.5 = 10% off), contrast-coded typicality (0.5 = typical focal; -0.5 = atypical focal), and the interaction between distance and magnitude as well as distance and typicality. Consistent with our hypothesis, people preferred to apply promotions to taxonomically-close comparison items: there was a positive effect of distance on ranking (β = .17, Wald's Z = 6.56, p < .001), and more "closest comparison" items were ranked the highest (42.5%, vs 26.8% (mid distance) and 30.6% (far distance), $\chi^2(2)$ = 106.19, p <.001, figure 7). We also observed a marginally significant tendency such that when the focal item is more typical, the effect of taxonomic distance is more prominent (interaction between distance and typicality β = .09, Wald's Z = 1.67, p = .09). The effects were consistent regardless of the promotion magnitude (interaction between distance and magnitude β = -.04, Wald's Z = -0.78, p = .43, ns). These patterns persist when we include how substitutable the items are for each other (see Web Appendix H and the results of Study 2c).

Study 3b

Whereas study 3a asked participants to rank comparison items, study 3b tested a different dependent variable: consequential choice. Participants chose one comparison item that they would like to get for free when the focal item was on sale, and we used a random lottery incentive where participants had a chance to realize their choices.

Participants and Procedures. We pre-registered that we would collect 300 participants. Three hundred and seven MTurk participants (all located in the U.S.) responded to the survey and 302 people passed the attention check. The average age in the sample was 38.8 (range: 20-70), and 51.7% were female.

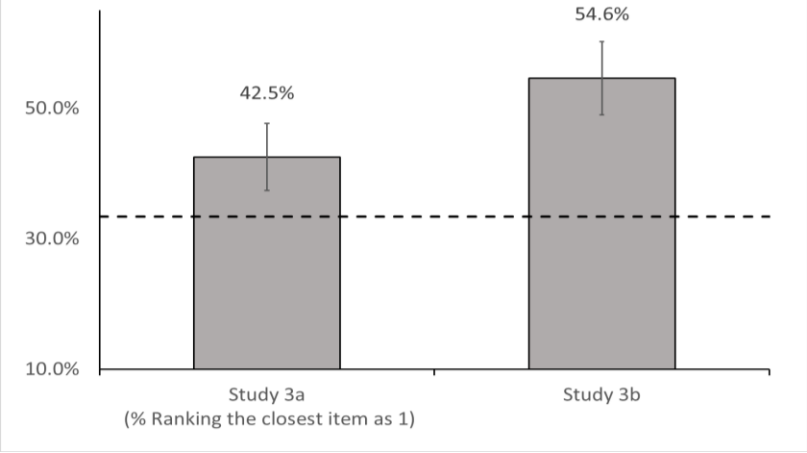
Participants were randomly assigned to one of two groups: they either assumed that they bought one of two focal items (coffee or toilet paper; forced-focal condition), or chose between the two items to buy (choose-focal condition). We included a choice group in this experiment because, whereas we have been presenting with participants focal items in our previous studies, in most real-world decisions, people choose the focal item that they purchase. Participants read that they spent $15 when purchasing the focal product, which was selected to be around the average price consumers usually spend on these products (Web Appendix F). Then, they learned that the focal item that they purchased comes with a buy-one-get-one-free promotion. They could apply the promotion to one of the three comparison items: detergents, pizza, and stationery products, all of which were priced around $15. The three comparison items are held constant in the study, but the distance with the focal item varied as revealed in our norming study (Web Appendix F). When the focal item is toilet paper, detergents share the closest distance, pizza shares the farthest distance, and stationery products are in the middle. When the focal item is coffee, pizza shares the closest distance, stationery products share the farthest distance, and detergents are in the middle.

After participants made their choices, they could opt in to provide a link to select a comparison product that they usually purchase. Participants then answered how much they needed each of the

comparison products in their daily life (1 – Don't need this at all; 7 – Need this a lot), and entered their age, gender, and income to complete the study. After the study was complete, three participants were selected at random and given a $15 MTurk bonus with a message that reminded them of their product choice (and the product link if provided).

Analysis and Results. Across two conditions, the proportion of choosing the closest item was significantly larger than the mid-distance item (54.6% vs 16.6%, $X^2(1) = 93.86$, $p < 0.001$, $\phi_{cramer} = 0.53$; figure 8), and the far-distance item (54.6% vs 28.8%, $X^2(1) = 40.37$, $p < 0.001$, $\phi_{cramer} = 0.31$). This is consistent with the notion that when people spent on a focal item, they wanted to save on the closest item. Unexpectedly, we also observe that people chose the far-distance item more often than the mid-distance item ($X^2(1) = 12.23$, $p < 0.001$, $\phi_{cramer} = 0.27$). The pattern did not differ between the forced-focal condition and the choose-focal condition ($X^2(1) = 1.75$, $p = 0.42$, $\phi_{cramer} = 0.08$).

FIGURE 8: % CHOOSING THE CLOSEST COMPARISON PRODUCT IN STUDY 3



Note: - (Left) Proportion of closest comparison products ranked 1st to apply promotions to in study 3a. (Right) Proportion choosing the closest comparison product to apply promotions to in study 3b. Horizontal line indicates the expected proportion if participants were choosing at random (33%).

Discussion

Consistent with H3, studies 3a and 3b found that in the context of promotions, consumers' savings decisions are influenced by the taxonomic distances between items. Specifically, when

consumers bought a focal item that comes with a discount, they were more likely to apply it to taxonomically close purchases. This pattern generalized across ranking (3a) and consequential choices (3b). This pattern is not consistent with simple ideas about substitutes and categories—if the item closest in taxonomic distance is just the closest substitute, people would not choose it after having just purchased the focal item. Notably, consumers appear to have made these spending and savings decisions by spontaneously recruiting categories of expenditures: all participants in study 3 responded to the surveys without performing the categorization task.

Study 3 further tested several potential moderators. Study 3a found that spending adjustments based on the taxonomic distance were more pronounced when the focal item was typical, which suggests that deviations from those typical expenditures better prompt people to adjust based on taxonomic distance. This finding also conceptually replicates the finding from Heath and Soll (1996), which supports our theorizing that budgeting behaviors are grounded in categorization principles and reinforces the current literature with consistent findings. Also, the effect of taxonomic distance remained constant with small and large discounts of the focal item (study 3a) and with pre-determined and participant-chosen focal items (study 3b).

Our results also hint that the predictive power of taxonomic distance could diminish as the distance between the items gets larger. Though we did not observe a consistent effect across all our empirical studies (see the internal meta-analysis in Web Appendix J), it seems that the shape of the effect of taxonomic distance is less clear, in particular, for distances greater than 2.[9] Recall that items with distance = 2 were grouped together in phase one of the pile-sort—it seems likely that we should observe our predicted effects there and at smaller taxonomic distances. Items at distances greater than

[9] Although the pattern that we predicted (applying the discount to taxonomically closer items, particularly at distances 1 and 2) generalized across two kinds of elicitations in Study 3 (ranking and consequential choice) and ratings in Study 2, the effects of further taxonomic distances are neither consistent across studies nor easily interpretable based on any plausible theories about how elicitation methods affect responses.

two share increasingly superordinate category membership or possibly no common category membership at all. Spending on an item may be associated with the spending on more mid-level or subordinate level categories more easily than a superordinate level category. When the taxonomic distance is larger than the standard, baseline level people usually adopt (as revealed in phase one sorts), people might evaluate spending relative to a comprehensive mental account like total wealth (Morewedge, Holtzman, and Epley 2007) and think less about the taxonomic distance.

Our findings have implications for marketing managers. First, marketers can leverage our findings to design product bundles and promotions that are more appealing to consumers. For example, a discounted bundle of toilet paper and detergents might be more appealing than a bundle of detergents and instant coffee. Similarly, marketers can improve the appeal of promotions by promoting taxonomically close items. When a promotion is contingent on purchasing another good, consumers might be more willing to claim the deal when the two items are taxonomically close to each other. Further, marketers can learn about consumers' mental representations of expenditures and design their research with an approach similar to the successive pile-sort in order to recover the consumers' representation between products.

STUDY 4: EXAMINATION OF GROCERY PURCHASE

In study 4, we examine the effect of taxonomic distance in a real-world shopping environment. Specifically, we investigate how spending on comparison items changes when focal items of different taxonomic distance are on sale. In study 4, we examine consumers' grocery purchase data from twelve years' worth of shopping trips (~6.5 million trips; Consumer Panel Data) collected and maintained by NielsenIQ.[10] Specifically, we expect two patterns of behavior. First, we expect consumers to increase

[10] Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets. The conclusions drawn from the NielsenIQ data are those of the

their spending on comparison items when they purchase an on-sale focal item because we expect people to increase their spending when they gain slack in their budget. Second, we expect the adjustment on comparison items to be larger when the focal item on sale is taxonomically close to the comparison item.

Data and Stimuli Selection

The Consumer Panel Data contains a longitudinal panel (from 2004 to date) of approximately 40,000-60,000 U.S. households who continually provide information about their purchases. The dataset contains information about the shopping trips, such as the price paid and quantity bought for each product, and whether each purchased product was on sale. It also contains information about each purchasing household such as their income, household size, and the number of children in the household. We used twelve years' worth of data—approximately 6.5 million shopping trips, from 2008-2019[11]—to examine the effect of a focal item's discount on the amount spent on the comparison item.

To examine spending adjustment on products of different taxonomic distances, we structured the data as follows. First, we collapsed across all sizes and brands of the same type of product and only recorded the product type. For instance, Tide Free and Gentle detergents 92 oz. and Arm & Hammer 122 oz. would both be recorded as detergents. This is because we only recovered the participants' taxonomies of the expenditures, but not specific products. If one of the brands or sizes was on sale, we marked the expenditure item as on sale. We then take the sum of the price purchased for the products as the total amount spent on the expenditure item.

---

researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[11] We focus our panel from 2008 to 2019 because the earlier years (2004-2006) have substantively fewer panelists, and we are concerned about the impact of Covid-19 in 2020. The results for all the years (2005-2020, ~8.1 million shopping trips) are reported in Web Appendix I. The patterns and the study's conclusions are the same for the focal analysis and the larger analysis reported in the Web Appendix.

Second, we selected the focal and comparison product sets presented in Table 4. The products were selected from the same set of expenditures that we used to recover distances for studies 3a and 3b (Web Appendix F). Specifically, we selected items that were comparable in price and not rated as substitutes or complements with each other based on the norming study we collected (Web Appendix G). Further, we deliberately chose items that are unlikely to appear near each other (e.g., in the same aisle or on the same shelf) in a retail space to minimize the influence of spatial proximity.

Finally, we selected households that have purchased the comparison item together with the focal item at least once in that calendar year. We attempt to investigate households that have budgets for the focal and comparison items we chose. Therefore, the items should not be exceptional (Sussman and Alter 2012) or one-time purchases for the household.

TABLE 4: STUDY 4 PRODUCT SETS (2008-2019)

Comparison	***Focal***	
	Close	**Far**
Detergents	Toilet Tissue (N = 2,244,990)	Coffee (N = 1,479,929)
Pizza	Coffee (N = 1,214,767)	Toilet Tissue (N = 1,183,760)
Detergents	Disposable Diapers (N = 177,780)	Pet Food (N = 61,501)
Pizza	Pet Food (N = 43,363)	Disposable Diapers (N = 116,388)

Notes. – Product sets used in study 4. The number below each focal item indicates the number of trips that were used in the analysis.

Model and Analysis

To test the effect of the focal item's deal on the amount spent on the comparison product, we estimated the two models, one for the close focal item and one for the far focal item. Both models have the following structure:

$$ln\left(1 + Y_{CompPurchased}\right) = \beta_0 + \beta_{deal}X_{focalDeal} + \beta_{household}\gamma + \beta_{compFE}C + \beta_{yearFE}T + \varepsilon.$$

$ln\ (1 + Y_{CompPurchased})$ is the log of the total amount spent on the comparison product. $X_{focalDeal}$ is an indicator variable of whether the focal product is on sale, and the two models would yield estimated coefficients for both the close focal item (β_{deal_close}) and the far focal item (β_{deal_far}). γ is a vector of household characteristics which include demographics (income level, household size, type of residence, and race) and past purchasing behavior (average past year's trip spending for the comparison item with and without promotions, basket size with and without the promotions, and past promotional values). C indicates the fixed effect for each comparison item, T represents the fixed effect for each year in the data, and ε represents the error term.[12]

Our theoretical framework generates two predictions. First, consumers would spontaneously increase their spending on the comparison items when a focal item is on sale. This implies that the β_{deal} coefficients should be positive for both close focals and far focals. Even though the expenditures vary in taxonomic distance, they are all expenditures incurred in one grocery trip, and grocery is a category recovered in study 1. Therefore, consumers would perceive that they have budget left for the grocery trip when an item is on sale and spend it on other items in the trip. Second, we predict that the effect should be greater when the focal item is taxonomically close to the comparison product, which means β_{deal_close} should be greater than β_{deal_far}. We emphasize that we focus on comparing the two coefficients instead of estimating a precise effect of the promotion; our primary interest is in seeing how comparison items of different distances are impacted by the same focal item.

Consistent with our predictions, we found that across the four sets, people increase spending on the comparison products both when taxonomically close items and taxonomically far items are on sale ($\beta_{deal\_close} = 0.042, t(3{,}680{,}862) = 68.413, p < .001$, and $\beta_{deal\_far} = 0.021, t(2{,}841{,}540) =$

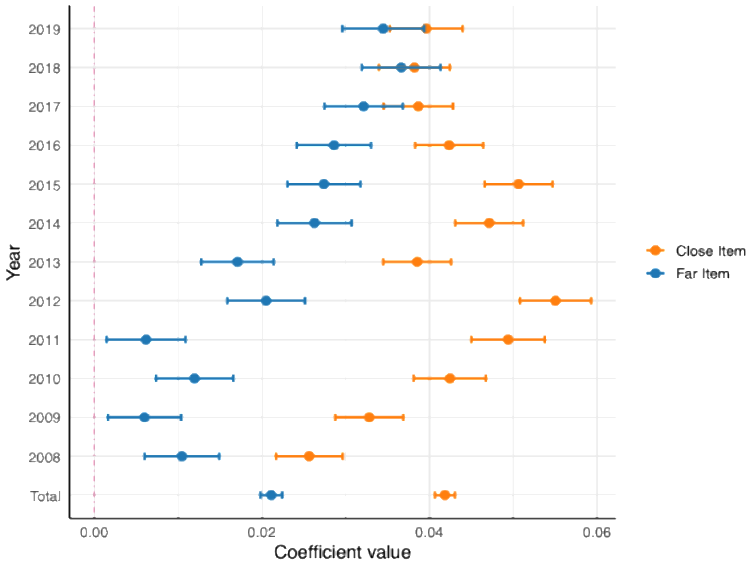
[12] We omit year-level fixed effects in the specification when we report the coefficients by each year of the data in Figure 9.

$31.627, p < .001$).[13] Web Appendix I supplies the regression tables. In addition, people increase spending more on the comparison product when the item on sale is taxonomically close (i.e., β_{deal_close} is significantly greater than β_{deal_far}, t = 22.88, p < 0.001). The pattern remains the same when we run the model each year (Figure 9).

Discussion

Study 4 applied our findings to consumer's actual purchase behaviors. Our results are consistent with H3: consumer's spending and savings decisions for everyday products are related to taxonomic distances between items. If the price drops on a given product, we can expect the demand on a taxonomically close item to increase. This finding further suggests that taxonomic distances may relate to, for example, cross-product elasticities, in addition to more widely studied factors like shelf placement and other marketing mix variables (Bezawada et al. 2009).

FIGURE 9: EACH YEAR'S FOCAL COEFFICIENT BY FOCAL DISTANCE (2008-2019)



Notes. – Focal coefficient by focal distance. Error bars represent 95% CI.

[13] The semi-log regressions have the following interpretations: for focally close items, a promotion leads to approximately a 4.2% increase in spending for the comparison item. For focally far items, a promotion leads to approximately a 2.1% increase in spending for the comparison item.

Some caveats are worth mentioning. We must be cautious in interpreting the $\beta_{deal\_close}, \beta_{deal\_far}$ coefficients, as there are omitted variables. The dataset contains no information, for example, about the spatial proximity of the items (i.e., how close two items are) in the physical stores. If the focal and the comparison items appear next to each other, it is somewhat plausible that receiving a deal on one would cause splurging on another. However, we think spatial proximity has limited influence on our results in study 4. Whereas all detergents are likely to be placed together, the products we chose (e.g., coffee, toilet paper) are likely to vary in their aisle placement across retailers. In addition, we chose products that we suspect are not right next to each other in the store (e.g., pizza is usually in the freezer and not next to coffee or toilet paper). Also, importantly, we cannot control for the firms' (unobserved) promotional and pricing strategies as they are optimizing their marketing mix (Hitsch et al. 2021). In any case, the results of this study need to be evaluated together with the experimental results as they relate to our hypotheses.

GENERAL DISCUSSION

In this paper, we investigated how people mentally represent expenditures and how taxonomic distance influences people's spending and savings behavior. We developed and introduced a successive pile-sort interface that can be used to quickly recover people's representation for many items. Using this interface, study 1a and 1b found that consumers share consensus on a taxonomic representation and the representation is stable over time. Because there was consensus, we were able to test the effect of the taxonomy on consumers' spending and savings decisions on both the individual (study 2a) and the aggregate level (study 2b). We found that after spending on a given item, people adjust their consumption more on items that are closer in distance to the given item, and this effect could not be explained by substitutability and complementarity patterns (study 2c). This adjustment appears to be spontaneous. People were presented with spending alternatives, without prompting an

explicit a budgetary category, and they still made adjustments based on taxonomic distances for various spending activities like promotion applications (study 3a and 3b) and grocery purchases (study 4). The results are consistent with our proposal that people represent expenditures hierarchically and recruit those representations in their spending decisions.

Limitations and Future Directions

We found that taxonomic distances predicted people's spending and savings decisions across multiple domains, but there are several limitations to note and questions left to explore. First, the framework we use for this paper primarily concerns what is likely a relatively stable taxonomy (study 1b) that consumers recruit when they make spending adjustments. We think that a stable taxonomy is a somewhat reasonable assumption given that consumers often interact with similarly structured shopping environments (e.g., groceries are separated into similar aisles). Not all taxonomies are this stable—many are not—and in many cases, consumers likely construct topical accounts in an ad-hoc fashion given the context (Barsalou 1983, 1985; Reinholtz et al. 2015; Sussman and Alter 2012). For example, rideshare fare and the cost of a meal might not be close in one's taxonomy, but they could be temporally close and related to a common event of dining out. Therefore, experiencing an unexpectedly high ride fare might lead to ordering a less expensive meal. This is an important phenomenon in mental accounting that lies beyond the scope of the current paper. We predict that holding temporal proximity and goal-relevance constant, people would likely adjust more on expenditures that are taxonomically close.

Relatedly, future work is needed to examine in detail the factors that may temporarily or permanently change people's taxonomy. We presume that consumers' taxonomies of expenditures are multiply determined and influenced by many sources, and we have only limited insight into the experiences and information-processing tendencies that formed the taxonomies we observe in this

project. Past research has identified a plethora of factors that affect the formation and use of categories. For example, the frequency with which two concepts co-occur (Barsalou 1985), goals that objects relate to (Ross 1997), and resemblance of features (Rosch and Mervis 1975) all influence the categories people construct. We conjecture that all of these elements play a role in how people represent expenditures. In addition, people could be influenced by knowledge specifically related to the organization of expenditures. Such knowledge includes typical groupings seen in stores and websites, along with standard budget categories they have been instructed to construct and track, as the characteristic ways that people tend to interact with objects in the world influences how they are sorted (Medin et al. 1997). Future studies can investigate how these myriad factors interact to shape people's taxonomies of expenditures.

Future work could also examine further moderators that impact the taxonomy's effect on spending decisions. Our studies focused on one factor that moderated the effect: typicality. Specifically, study 3a finds that when the focal item is typical (and is thus more likely to prompt the categorical structure), people are more likely to adjust according to taxonomic distance. Other potential indicators might include individual differences: people who budget much might be more affected by the taxonomy. Our studies also hint that the predictive power of taxonomic distance seems to decrease as distance increases (study 3). Although results from our studies were inconclusive (see Appendix J), it is conceivable that if two items are very unrelated (with a large taxonomic distance), the category that includes both might be inaccessible or so expansive as to be best characterized as a "comprehensive" account. Taxonomic distance in these cases may not predict spending well.

Finally, this work focuses on the situations where people are in shopping environments and thinking about spending adjustments among alternatives with the idea of budgeting in their minds. We highlight that in these situations where alternatives are present, the taxonomic relationship between purchases matters. One other important step to budgeting could be the planning stage, during which

people formally set a budget by pre-defining the items or the category and deciding how much money they'd allocate to each category. Future work on budgeting could also focus on the behavior in such planning stages and how people generate alternatives to evaluate (Bear et al. 2020; Mills and Phillips 2023; Morris et al. 2021).

Theoretical Contributions

Our investigation offers a number of theoretical contributions. First, our approach introduces one way to connect people's mental representation with important marketplace behaviors: we first expand our understanding of how knowledge is organized in people's minds and then develop educated hypotheses based on the new understanding. Expenditures are often goal-oriented, and categories constructed from expenditures are often *ad hoc* (e.g., "things that can be purchased at a supermarket"). It was not obvious *ex ante* whether people represent them the same way that people represent other objects, like trees or animals. Therefore, we started with extracting the representation with the online successive pile-sort interface, which helps recover the representation for a large group in a short time. The recovered taxonomy then provides a disciplined basis to further develop and test our hypotheses as it makes clear which items are closer in the taxonomy. Our framework is novel in the context of understanding of mental accounts and should be generalizable to future research that investigates representations.

Our recovered representations connect to many existing concepts such as typical expenses, exceptional expenses, and malleable expenses. Because we mapped out relationships between expenditures and recovered categories as clusters (study 1), we can make predictions about how typical a given expenditure is for a category (Heath and Soll 1996; Reinholtz et al. 2015). Specifically, items closer to the center of a cluster should be more typical of their categories. Similarly, items in-between two clusters could be more malleable and could be assigned to multiple categories (e.g., "dining out" as

"food" or "entertainment"; Cheema and Soman 2006). For example, rent would be typical to the "bills" category whereas gym membership is more malleable between "bills" or "entertainment expenditures". Similarly, we could also identify items that are likely to be treated as exceptional expenses (Sussman and Alter 2012) as items that do not seem to be categorized with others in the dendrogram, like flowers and suitcases. We hope that this work serves as a starting point for more research that combines uses insights from theories of concepts and categories to better understand budgeting and consumer decision making, more generally.

Further, we identify a downstream effect of people's mental representations. Much research on mental representation investigated how people's representations influence mental processes such as the way they think (Collins and Quillian 1969) and make inferences (Randall 1976; Osherson et al. 1990). Meanwhile, less of this basic research in cognitive science has focused on connecting people's mental representation to their observable behaviors (cf. Atran et al. 1999). Our work helps bridge this gap by suggesting that people recruit their mental representation when they are making spending and savings decisions, which are concrete, observable, important behaviors. Specifically, people's mental representation of expenditures helps determine which items they choose to adjust their spending on as well as the magnitude of their spending adjustment. They adjust the most on items that are taxonomically closest to the focal item. These results in turn emphasize the importance of studying people's representations (cf. Reinholtz et al. 2015) and highlight the connection between one's cognitive processes and consumption behaviors (Bartels and Johnson 2015).

Finally, our proposed structure contributes to the current budgeting literature by relaxing the often-used binary distinction of "within-category" and "out-of-category" (Heath and Soll 1996). As our result reveals, a binary view of category inclusion can be restrictive and produce inaccurate predictions of how people adjust their spending. If people followed a binary distinction for mental accounts, they would likely only adjust spending on detergent when toilet paper is on sale but not when pizza is on

sale. However, as revealed in study 4, a sale on pizza impacts the amount purchased for detergents, likely because they are categories together at a higher level of "grocery purchases". This suggests that budgeting decisions more likely recruit a taxonomy and items further away in the taxonomy can still impact each other (although the degree may vary, see Appendix J). Because we investigated and revealed people's representation of expenditures, we were able to expand beyond the "topical" versus "comprehensive" dichotomous view of mental budgeting to generate more refined hypotheses around how people spend and budget.

Marketing Implications

Consumers often deviate from their budgets. High inflation, unemployment, and other income and price shocks have likely caused more people to budget and consequently adjust their spending and savings decisions. Our work provides some insights on how marketers can approach the questions of predicting and helping with consumers' spending adjustments. Credit card companies and personal finance apps can promote better consumption and budgeting behaviors when they better understand people's representation of expenditures. For example, if these companies want to decrease overspending (e.g., to lower default rates), they could label expenditures at a level that is more specific to people's natural level of representation, because prompting people to consider smaller, lower-level categories may help consumers restrict spending (Thaler and Shefrin 1981; Krishnamurthy and Prokopec 2010). By eliciting multiple levels of expenditure grouping (figure 4; dendrogram in Study 1), we provide inputs for how to design labels that are more or less specific to people's natural grouping (e.g., whether "fast food" is indeed more specific than how consumers naturally think about expenditures).

Relatedly, taxonomies like the ones we have recovered might provide important insights on directory and menu designs. Companies and marketers that organize items for consumers can benefit

from understanding how consumers naturally think about expenditures. They can better structure or position the products into a hierarchy that is consistent with consumers' representation for improved product search experience. For example, we found that consumers represent food and daily products closer than food and clothing (study 1). Supermarkets can benefit from this knowledge and organize their search directory or aisles accordingly (i.e., putting foods and daily products closer) because a well-organized directory can reduce search complexity (Jacko and Salvendy 1996; Miller 1981).

In addition to probing how consumers think about expenditures, we highlight the approach of the successive pile-sort method of investigating relationships between concepts relevant to marketing. Marketers can use successive pile-sort to understand consumers' representations for brands and their products. For example, marketers can use the successive pile-sort to generate conceptual maps for brands (Hauser and Koppelman 1979; Negunadi 1990) and better understand the demand relationships between them: brands organized together at low levels may likely be substitutes. Alternatively, a brand considering launching a new product (e.g., product extensions) can use the successive pile-sort method to examine whether the product fits the brand (Aaker and Keller 1990; Bottomley and Holden 2001). Specifically, the brand can use the successive pile-sort method to assess the distance between current products and the new product. This allows the brand to further gauge whether the new product should be introduced (Parker et al. 2018) and what the marketing message should be (Moreau, Markman and Lehmann 2001).

Finally, even though participants in our studies reached a consensus in their sorting, marketers can still benefit from the cases when there is a lack of consensus in consumers' representations. When consumers' representations of products or brands diverge, it implies that people have heterogeneous associations for a given product or brand. Marketers could consider using these differences to characterize different consumer segments. For example, consumers who more readily group together wine and cheese might be systematically different from consumers who more readily group together

milk and cheese. Understanding how other differences across consumers relate to differences in sorting might allow marketers to better target their consumers.

The current paper develops a framework from theories of categories and concepts and applies a method of investigation from that literature to ask new questions about mental accounting and mental budgeting. The studies investigated consumers' representation of expenditures and related the recovered taxonomic distances to spending and savings decisions. We suspect the methods here could be used to investigate hypotheses beyond budgeting-related decisions. By connecting people's cognitive processes with consequential behavior, we hope the paper invites more investigations that apply methods and theoretical perspectives from cognitive science in their study of consumer behavior (Bartels and Johnson 2015).

REFERENCES

Aaker, David A. and Kevin L. Keller (1990), "Consumer Evaluations of Brand Extensions," *Journal of Marketing*, 54 (January), 27–41.

Antonides, Gerrit, I. Manon De Groot, and W. Fred Van Raaij (2011), "Mental Budgeting and the Management of Household Finance," *Journal of Economic Psychology* 32(4), 546-55.

Arkes, Hal R., Cynthia A. Joyner, Mark V. Pezzo, Jane Gradwohl Nash, Karen Siegel-Jacobs, and Eric Stone (1994), "The Psychology of Windfall Gains," *Organization Behavior and Human Decision Processes*, 59 (3), 331 – 47.

Atran, Scott, Douglas Medin, Norbert Ross, Elizabeth Lynch, John Coley, Edilberto Ucan Ek, and Valentina Vapnarsky (1999), "Folkecology and Commons Management in the Maya Lowlands," *Proceedings of the National Academy of Sciences*, 96(13), 7598-603.

Barsalou, Lawrence W. (1983), "Ad Hoc Categories," *Memory and Cognition*, 11 (3), 211–27.

———. (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), 629–54.

Bartels, Daniel M., and Eric J. Johnson (2015), "Connecting Cognition and Consumer Choice," *Cognition*, 135, 47-51.

Bear, Adam, Samantha Bensinger, Julian Jara-Ettinger, Joshua Knobe, and Fiery Cushman (2020), "What Comes to Mind?" *Cognition*, 194, 104057.

Berlin, Brent (1992), *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies*. Princeton University Press.

Bettman, James R., Mary Frances Luce, and John W. Payne (1998) "Constructive Consumer Choice Processes." *Journal of Consumer Research*, 25(3), 187-217.

Bezawada, Ram, Subramanian Balachander, Pallassana K. Kannan, and Venkatesh Shankar (2009), "Cross-Category Effects of Aisle and Display Placements: A Spatial Modeling Approach and Insights," *Journal of Marketing*, 73(3), 99-117.

Boster, James (1994), "The Successive Pile Sort," *Cultural Anthropology Methods*, 6(2), 11-12.

Bottomley, Paul A., and Stephen JS Holden (2001), "Do We Really Know How Consumers Evaluate Brand Extensions? Empirical Generalizations Based on Secondary Analysis of Eight Studies," *Journal of Marketing Research*, 38(4), 494-500.

Cheema, Amar and Dilip Soman (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.

Chintagunta, Pradeep K., and Sudeep Haldar (1998), "Investigating Purchase Timing Behavior in Two Related Product Categories," *Journal of Marketing Research*, 35(1), 43-53.

Collins, Allan M., and M. Ross Quillian (1969), "Retrieval Time from Semantic Memory." *Journal of Verbal Learning and Verbal Behavior*, 8, 240-47.

Cox, Michael AA, and Trevor F. Cox (2008), "Multidimensional Scaling." In *Handbook of Data Visualization*, Springer, Berlin, Heidelberg, 315-47.

Evers, Ellen RK., Alex Imas, and Christy Kang (2022), "On the Role of Similarity in Mental Accounting and Hedonic Editing," *Psychological Review*, 129(4), 777-89.

Felcher, E. Marla, Prashant Malaviya, and Ann L. McGill (2001), "The Role of Taxonomic and Goal-Derived Product Categorization in, Within, and Across Category Judgments," *Psychology & Marketing*, 18(8), 865-87.

Gentner, Dedre, and Arthur B. Markman (1997), "Structure Mapping in Analogy And Similarity," *American Psychologist*, 52(1), 45-56.

Goodman, Nelson (1972), "Seven Strictures on Similarity," *Problems and Projects*, 437-47.

Hastings, Justine S. and Jesse M. Shapiro (2013), "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks", *The Quarterly Journal of Economics*, 128(4), 1449-498.

Hastings, Justine, and Jesse M. Shapiro (2018), "How Are SNAP Benefits Spent? Evidence from a Retail Panel." *American Economic Review*, 108(12), 3493-540.

Hauser, John R., and Frank S. Koppelman (1979), "Alternative Perceptual Mapping Techniques: Relative Accuracy and Usefulness," *Journal of Marketing Research*, 16(4), 495-506.

Heath, Chip and Jack B. Soll (1996), "Mental Budgeting and Consumer Decisions", *Journal of Consumer Research*, 23(1), 40-52.

Henderson, Pamela W., and Robert A. Peterson (1992), "Mental Accounting and Categorization," *Organizational Behavior and Human Decision Processes*, 51 (2), 92–117.

Hirschman, Elizabeth C., and Morris B. Holbrook (1982), "Hedonic Consumption: Emerging Concepts, Methods and Propositions," *Journal of Marketing*, 46(3), 92-101.

Hitsch, Günter J., Ali Hortacsu, and Xiliang Lin(2021), "Prices and Promotions in U.S. Retail mMarkets," *Quantitative Marketing and Economics* (19), 289–368.

Hummel, John E., and Keith J. Holyoak (2003), "A Symbolic-Connectionist Theory of Relational Inference and Generalization," Psychological Review, 110(2), 220-64.

Jacko, Julie A., and Gavriel Salvendy (1996), "Hierarchical Menu Design: Breadth, Depth, and Task Complexity," *Perceptual and Motor Skills*, 82(3), 1187-201.

Kahneman, Daniel and Amos Tversky (1984), "Choices, Values, and Frames," *American Psychologist*, 39, 341 – 50.

Kőszegi, Botond, and Filip Matějka (2020), "Choice Simplification: A Theory of Mental Budgeting and Naive Diversification," *The Quarterly Journal of Economics*, 135(2), 1153-207.

Krishnamurthy, Parthasarathy, and Sonja Prokopec (2010), "Resisting That Triple-Chocolate Cake: Mental Budgets and Self-Control," *Journal of Consumer Research*, 37(1), 68-79.

Kruskal, Joseph B. (1978), *Multidimensional Scaling*. No. 11. Sage.

Levav, Jonathan and A. Peter McGraw (2009), "Emotional Accounting: How Feelings About Money Influence Consumer Choice," *Journal of Marketing Research,* 46, 66–80.

Lopez, Alejandro, Scott Atran, John D. Coley, Douglas L. Medin, and Edward E. Smith (1997), "The Tree of Life: Universal and Cultural Features of Folkbiological Taxonomies and Inductions," *Cognitive Psychology*, 32, 251-95.

Lynch Jr, John G., Richard G. Netemeyer, Stephen A. Spiller, and Alessandra Zammit (2010), "A Generalizable Scale of Propensity to Plan: The Long and the Short of Planning for Time and for Money," *Journal of Consumer Research,* 37(1), 108-28.

Manchanda, Puneet, Asim Ansari, and Sunil Gupta (1999), "The "Shopping Basket": a Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18(2), 95-114.

Markman, Arthur B. (2002), "Knowledge Representation."

Markman, Arthur B., C. Miguel Brendl, and Kyungil Kim (2007), "Preference and the Specificity of Goals," Emotion 7(3), 680-84.

Medin, Douglas L., Robert L. Goldstone, and Dedre Gentner (1993), "Respects for Similarity." *Psychological Review*,100(2), 254-78.

Medin, Douglas L., and Scott Atran (2004), "The Native Mind: Biological Categorization and Reasoning in Development and Across Cultures," *Psychological Review*, 111(4), 960-83.

Medin, Douglas L., Elizabeth B. Lynch, John D. Coley, and Scott Atran (1997), "Categorization and Reasoning Among Tree Experts: Do All Roads Lead to Rome?" *Cognitive Psychology,* 32, 49-96.

McCullagh, Peter (1980), "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2), 109-127.

Mills, Tracey, and Jonathan Phillips (2023), "Locating what comes to mind in empirically derived representational spaces." *Cognition*, 240, 105549.

Miller, Dwight P (1981), "The Depth/Breadth Tradeoff in Hierarchical Computer Menus," *Proceedings of the Human Factors Society Annual Meeting*, 25(1), 296-300.

Moreau, C. Page, Arthur B. Markman, and Donald R. Lehmann (2001), ""What is It?" Categorization Flexibility and Consumers' Responses to Really New Products," *Journal of Consumer Research*, 27(4), 489-98.

Morewedge, Carey K., Leif Holtzman, and Nicholas Epley (2007), "Unfixed Resources: Perceived Costs, Consumption, and the Accessible Account Effect," *Journal of Consumer Research*, 34(4), 459-67.

Morris, Adam, Jonathan Phillips, Karen Huang, and Fiery Cushman (2021), "Generating Options and Choosing Between Them Depend on Distinct Forms of Value Representation," *Psychological Science*, 32(11), 1731-746.

Nedungadi, Prakash (1990), "Recall and Consumer Consideration Sets: Influencing Choice Without Altering Brand Evaluations," *Journal of Consumer Research*, 17(3), 263-76.

Osherson, Daniel N., Edward E. Smith, Ormond Wilkie, Alejandro Lopez, and Eldar Shafir (1990), "Category-Based Induction," *Psychological Review*, 97(2), 185-200.

O'Donnell, Michael, and Ellen RK Evers (2019). "Preference Reversals in Willingness to Pay and Choice," *Journal of Consumer Research*, 45(6), 1315-30.

Parker, Jeffrey R., Donald R. Lehmann, Kevin Lane Keller, and Martin G. Schleicher (2018), "Building a Multi-Category Brand: When Should Distant Brand Extensions Be Introduced?" *Journal of the Academy of Marketing Science*, 46(2), 300-16.

Randall, Robert A (1976), "How Tall is a Taxonomic Tree? Some Evidence for Dwarfism," *American Ethnologist*, 543-53.

Ratneshwar, S. and Allan D. Shocker (1991), "Substitution in Use and the Role of Usage Context in Product Category Structures," *Journal of Consumer Research,* 28 (August), 281–95.

Reinholtz, Nicholas, Daniel M. Bartels, and Jeffrey R. Parker (2015), "On the Mental Accounting of Restricted-Use Funds: How Gift Cards Change What People Purchase." *Journal of Consumer Research*, 42 (8), 596-614.

Rick, Scott I., Cynthia E. Cryder and George Loewenstein (2008), "Tightwads and Spendthrifts", *Journal of Consumer Research,* 34(6), 767-782.

Romney, A. Kimball, Susan C. Weller, and William H. Batchelder (1986), "Culture as Consensus: A Theory of Culture and Informant Accuracy," *American Anthropologist*, 88, 313-38.

Rosch, Eleanor, and Carolyn B. Mervis (1975), "Family Resemblances: Studies in the Internal Structure of Categories," *Cognitive Psychology*, 7(4), 573-605.

Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem (1976), "Basic Objects in Natural Categories," *Cognitive Psychology*, 8, 382–439.

Ross, B. H. (1997), "The use of categories affects classification," *Journal of Memory and Language*, 37, 240–267.

Smith, Edward E. (1978), "Theories of Semantic Memory," *Handbook of Learning and Cognitive Processes*, 6, 1-56.

Soman, Dilip (2004), "Framing, Loss Aversion, and Mental Accounting," *Blackwell Handbook of Judgment and Decision Making, ed. Koehler Derek J. Harvey Nigel, Oxford, U : Blackwell,* 379 – 98

Soman, Dilip, and Amar Cheema (2011) "Earmarking and Partitioning: Increasing Saving by Low-Income Households," *Journal of Marketing Research,* 48, S14-S22.

Sussman, Abigail B., and Adam L. Alter (2012), "The Exception is the Rule: Underestimating and Overspending on Exceptional Expenses." *Journal of Consumer Research,* 39(4), 800-14.

Thaler, Richard H., and Hersh M. Shefrin (1981), "An Economic Theory of Self-Control." *Journal of Political Economy*, 89(2), 392-406.

Thaler, Richard H. (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 27 (1), 15–25.

———. (1999). Mental Accounting Matters. *Journal of Behavioral Decision Making*, 12(3), 183-206.

Thaler, Richard H., and Eric J. Johnson (1990), "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice," *Management Science*, 36(6), 643-60.

Tversky, Amos, and Itamar Gati (1979), "Studies of Similarity," *Cognition and Categorization*, 79-98.

Weller, Susan C (2007), "Cultural Consensus Theory: Applications and Frequently Asked Questions," *Field Methods*, 19(4), 339-68.

Zhang, C. Yiwei, Abigail B. Sussman, Nathan Wang-Ly, and Jennifer Lyu (2020), "How Consumers Budget." Available at SSRN 3739543.