Budgeting Increases Reliance on Category-Level Evaluations

David Dolifka and Stephen A. Spiller

UCLA Anderson School of Management

Last updated 06/22/2023

CONSUMER RELEVANCE AND CONTRIBUTION STATEMENT

There is a recent and growing interest in understanding how consumers budget and the implications of budgeting for consumption (Amasino, Dolgin and Huettel 2023; Choe and Kan 2021; Lukas and Howard 2023; Zhang et al. 2022; Zhang and Sussman 2018). Much of the budgeting literature considers how the use of budgets affects consumers' overall level of spending (Larson and Hamilton 2012; Lukas and Howard 2023; Thaler 1985, 1999; Thaler and Shefrin 1981; Wertenbroch 1998). The present paper sets the issue of overall level of spending aside and focuses on allocation. We ask "what drives budget allocations?" We propose allocating engages a different psychological process than does purchasing. Specifically, budget allocation increases the reliance on category-level evaluations (assessing a category as a whole). As a result, budgeters are more sensitive to summary representations of a category (i.e., a category's average value) than non-budgeters. We contribute to the theoretical literature on budgeting and resource allocation by identifying how allocating (vs. purchasing) impacts how consumers perceive value, without necessarily changing their underlying preferences. This has important implications for spending because consumers tend to follow the budgets they set. When allocations are pulled towards budget categories that appear more valuable on average, subsequent spending follows. This suggests the mere act of budgeting will shift consumption towards higher-average categories and that consumers with identical preferences will spend their funds differently, depending on whether they budget in advance of spending. These findings may have practical implications for the design of budgeting applications and tools.

ABSTRACT

Consumers frequently use budgets to manage their spending. Budgets are consequential, as money in budgets is treated as though it is not fungible, so budget allocations matter. How do consumers set budget allocations? Consumers could use item-level evaluations, and set budgets in accordance with their end purchases, or category-level evaluations, and set budgets in accordance with their overall evaluations of the category. We propose that relative to purchase decisions, allocation decisions increase consumers' reliance on category-level evaluations. As a result, budget allocations are more sensitive to category summary representations (i.e., the average value of a category) than are purchase decisions. Because budget allocations causally impact spending, the act of budgeting shifts spending toward categories with higher average value. This implies consumers with identical preferences and identical budget levels will spend differently, depending on whether they allocate in advance of spending. We present evidence of these patterns in three preregistered experiments including a mouse-tracking study and an incentivized, data-rich experimental game.

Keywords: budgeting; resource allocation; evaluation mode

Budgeting can be a powerful tool for consumers to manage their personal finances. The act of budgeting involves allocating resources across categories and then making spending decisions within those categories. As a result of this two-stage process, budget allocations are consequential: They affect both what consumers buy and how much they spend. This central role of budget allocations in consumer spending calls attention to a key question: What determines budget allocations? The present research seeks to further our understanding of how allocating funds across categories relates to the underlying values of those budget categories.

We propose that setting a budget leads consumers to evaluate a set of potential purchases differently than purchasing does. Specifically, we propose that relative to purchasing, allocating leads to a greater focus on *category-level evaluations*. When evaluating categories, people often rely on a category's average as a summary representation. Therefore, we propose that relative to purchase decisions, allocation decisions are more sensitive to the average value of a budget category. As a result, the mere act of budgeting—by virtue of engaging greater use of category-level evaluations—can affect the composition of spending. This implies consumers with identical preferences may systematically consume different bundles, depending on whether they budget.

We begin by defining budgeting and discussing prior research on consumer budgeting and budget allocation. We then discuss convergent findings regarding people's tendencies to use different modes of evaluation when encountering individual items versus sets of items. We connect this work to the psychology of allocation, which we argue encourages consumers to adopt a category-level mode of evaluation when making allocation decisions. We then present three experiments. Experiment 1 tests whether the sensitivity to category evaluations differs between consumers who allocate versus those who make purchases. Experiment 2 explores process by using mouse-tracking to record how participants search for items arranged in a grid of

consumption options, and how this search differs for allocation vs. purchase. We use this mouse tracking measure as a proxy for category-level evaluations and examine its implications for allocations and purchases. Experiment 3 uses a tightly controlled budgeting and spending game to manipulate average category values while holding alternative forms of value (i.e., marginal value) constant. The set of results indicates that allocation is uniquely sensitive to category-level evaluations, such as a category's average value. This has important implications for consumers' use of budgets to manage their finances, because allocating a budget may shift consumption in previously unforeseen ways.

THEORETICAL DEVELOPMENT

Budgeting as a Two-Stage Process

Budgeting is a two-stage process involving: (1) the allocation of funds, and (2) the subsequent spending of those funds. In the first stage, allocation represents the division of funds between distinct accounts. Allocation makes money non-fungible, as specific funds become linked to specific usages. In the second stage, previously allocated funds are used in a manner consistent with the account's rules (Heath and Soll 1996; Lukas and Howard 2023; Thaler 1985; Zhang and Sussman 2018).

Budgets can be used to manage both the total level of spending as well as the composition of spending across budget categories. A representative survey conducted by Zhang et al. (2022) finds both motivations are important to consumers. Whereas prior consumer research has considered how the act of budgeting affects consumption levels (Larson and

Hamilton 2012; Lukas and Howard 2023; Thaler 1985, 1999; Thaler and Shefrin 1981; Wertenbroch 1998), we consider how the act of budgeting affects the composition of spending. Our focus is on how consumers distribute their funds across categories and the bundles they ultimately consume, rather than the total amount of resources budgeted and spent. In our own replication and extension of the Zhang et al. (2022) budgeting survey, we find the majority (58%) of budgeters claim one of the main reasons for budgeting is to manage spending across various categories (web appendix A). We seek to explore how budgeters divide their funds (and ultimate consumption) across such categories.

The effect of budgeting on consumption. Budgets matter because consumers prefer to spend within their budget allocations. Once allocated, money in budgets is treated as though it is no longer fungible: Money budgeted for one purpose is less likely to be used for a different purpose (Hastings and Shapiro 2013, 2018; Heath and Soll 1996; Soman and Cheema 2011; Sussman and O'Brien 2016; Thaler 1985; Zelizer 1997). As a result, budgeting has numerous effects on consumption. Having pre-established budgets affects how consumers respond to price and income shocks (Du and Kamakura 2008; Hastings and Shapiro 2013, 2018), and consumers can use budgets strategically to reduce consumption of goods they seek to limit due to self-control considerations (Krishnamurthy and Prokopec 2010). Ironically, under certain circumstances, using budgets can lead to unintentional increases in spending. For example, the use of budgets might reduce the focus on minimizing costs, conditional on remaining under budget (Larson and Hamilton 2012), and consumers who set budgets too early might habituate to a higher level of consumption and find it harder to regulate the spending of previously allocated funds (Choe and Kan 2021). Depending on whether a limited budget (e.g., a weekly happy hour

budget) or expansive budget (e.g., a monthly food budget) is more accessible can impact the perceived costliness of different expenditures, thereby affecting consumption (Morewedge, Holtzman and Epley 2007).

As these examples demonstrate, allocation has direct consequences for spending. So, what affects allocation? Prior research highlights several key inputs into the budget allocation decision.

Predicted spending. One key input is predicted spending: When people believe they will spend more, they tend to allocate more money to that budget (Howard et al. 2022; Lukas and Howard 2023; Peetz and Buehler 2009; Stilley, Inman and Wakefield 2010a; Stilley, Inman and Wakefield 2010b; Sussman and Alter 2012; Ülkümen, Thomas and Morwitz 2008). People are not always well-calibrated: Their predictions are often underestimates of true spending for a variety of reasons. But in categories in which consumers expect to spend more, they tend to set larger budgets (Howard et al. 2022; Lukas and Howard 2023).

Self-control. Budget allocations are also often intertwined with self-control considerations. Budgets enhance self-control (and reduce consumption) when avoidance aspects of the consumption experience are highly salient and consumption monitoring is feasible (Krishnamurthy and Prokopec 2010). As a result, consumers may strategically set budgets lower than predicted spending in such contexts (Thaler 1985, 1999; Thaler and Shefrin 1981; Wertenbroch 1998). Budgets can also help constrained consumers navigate trade-offs they might otherwise avoid, thereby reducing dysfunctional behavior (Fernbach, Kan and Lynch 2015).

Incidental factors. Beyond predicted spending and self-control considerations, a number of incidental factors affect budget allocations. These are factors which ought to be irrelevant by most accepted normative standards but nevertheless shape the allocations that consumers make. Budget allocations depend on arbitrary groupings of budget categories, consistent with the broader literature on partition dependence (Bardolet, Fox and Lovallo 2011; Jia, Li and Krishna 2020; West et al. 2022). For example, consumers may allocate more money to movies if they have two budgets devoted to movies and food (where food encompasses both groceries and dining out), than if they have three budgets devoted to movies, groceries, and dining out. In addition, consistent with a broader literature indicating that attention affects choice, exogenous factors that call greater attention to a budget category lead to greater prioritization of that budget category (Mrkva and Van Boven 2017).

In each of these cases of predicted spending, self-control, and incidental factors, a key underlying assumption is that consumers allocate based on where they perceive value. That is, each of these literatures implicitly or explicitly acknowledge that consumers allocate more to budget categories that are evaluated as more valuable. But how do consumers assess the value of their budget categories?

Evaluation Mode

In assessing a set of options, consumers might focus their evaluations on individual items or on the category as a whole. For example, a consumer might consider each item and allocate funds according to the most-attractive individual items. Alternatively, they might assess the value of a category directly and holistically, possibly without scrutinizing the values of

individual items. We refer to the two styles as *item-level* and *category-level* evaluations, respectively. Applied to consumption contexts, an item-level evaluation focuses on the aspects of individual items; whereas a category-level evaluation focuses on what is represented by a category of items. This distinction between item-level and category-level evaluation modes relates to the distinction between derived and direct evaluations (Sood, Rottenstreich and Brenner 2004) and bears similarities to analytic versus holistic thinking (Masuda and Nisbett 2001; Nisbett et al. 2001).

Task cues evaluation mode. Prior work on processing styles suggests different evaluation modes can be cued by the task at hand (e.g., Hossain 2018; Monga and John 2008). When making choices between stores, consumers are more likely to rely on direct evaluations of sets whereas when making choices between unorganized sets of options, consumers are more likely to rely on derived evaluations of sets from items (Sood et al. 2004). When a task benefits from a narrow focus and attention to detail (e.g., identifying individual features embedded within a drawing), people adopt an analytic approach. When a task benefits from a broader view (e.g., writing a story about the drawing), people take a holistic approach (Monga and John 2008). Purchasing and allocating represent fundamentally different types of tasks. Purchasing involves identifying the most valued option(s) from a set of alternatives and can often be reduced to a choice of *which* item to buy. Allocation requires anticipating the distribution of future purchases across various categories and represents a question of *how much* to spend in each category. We propose that an item-level evaluation closely fits the task of deciding which item to purchase, and a category-level evaluation better suits the task of anticipating how much will be purchased in various budget accounts.

Category-level evaluations. Prior research on perception, cognition, and social cognition finds consistent evidence for how people assess sets, categories, and groups of people. Across a variety of stimuli and contexts, people accurately and automatically extract summary information from sets, but not necessarily the properties of individual members within sets (Ariely 2001; Haberman and Whitney 2009; Whitney and Yamanashi Leib 2018; Yamanashi Leib et al. 2020). People are particularly adept at extracting the mean of a category as its representation. Early tests examined this tendency in the context of basic visual perceptions involving size, colors, and motion paths (Ariely 2001; Chong and Treisman 2003; Watamaniuk and Duchon 1992). Remarkably, people also extract category means from complex assortments (Whitney and Yamanashi Leib 2018; Woiczyk and Le Mens 2021), such as facial features and emotional expressions (Haberman and Whitney 2007, 2009).

Such ensemble perception also occurs in consumer judgments: Consumers can accurately extract the average value of a set of products, even when they are unable to remember individual products within the set (Yamanashi Leib et al. 2020). Consumers may also base decisions on extracted averages. For example, when reporting their willingness to pay for a choice set, adding a less-attractive alternative decreases willingness to pay (Le Lec and Tarroux 2020). Consumers are less willing to pay for a medium of exchange which does versus does not have additional less-attractive uses associated with it (Spiller and Ariely 2020). Even when considering relatively simple gambles, adding a dominated option decreases the proportion of occasions on which consumers choose that choice set (Smith and Spiller 2023).

Complexity elicits averaging. Why do people extract averages from categories? This is presumably an adaptive response of the visual processing systems to reduce complexity in the

environment (Ariely 2001; Whitney and Yamanashi Leib 2018). But the use of averaging to reduce complexity extends beyond visual processing. In economic decisions, consumers tend to smooth out non-linearities, effectively using average costs as a basis for consumption decisions (Liebman and Zeckhauser 2004). These tactics are observed for judgments and decisions in the face of tax schedules (de Bartolome 1995; Rees-Jones and Taubinsky 2020), price schedules (Gottfries and Hylton 1987; Ito 2014; Shin 1985), and credit card repayments (Gathergood et al. 2019). Through reinforcement learning, people tend to make choices over time in proportion to the average long-run benefits they receive from those choices (Davison and McCarthy 1988; Herrnstein et al. 1993; Herrnstein and Prelec 1991; McDowell 2013; Rachlin and Laibson 1997). Compared to individual purchasing decisions, budget allocation is—or at least feels like—a more complex task. Therefore, we suspect the added complexity is an additional reason that consumers may use category-level evaluations to extract a category average, without focusing on each option within a budget category.

Thus, we propose that compared to purchasing, allocation will be more likely to engage a category-level mode of evaluation. This is because the allocation task inherently operates at the level of categories, with consumers forced to evaluate how much they value each category.

H1: Compared to purchasing, allocation induces a greater relative focus on category-level evaluations (vs. item-level evaluations).

As detailed above, people tend to automatically extract the mean when evaluating categories. Therefore, we propose a category's average[1] value will guide allocation decisions more than purchase decisions. This is because allocating leads consumers to form category-level evaluations, which should reflect a category's average value.

H2a: Allocation decisions are more sensitive to a category's average value than are purchase decisions.

H2b: The differential sensitivity to a category's average value between allocation and purchasing is explained by differences in evaluation modes.

An implication of H2a is that allocations and purchases (without prior allocation) will differ, even among consumers with identical preferences and budget levels. Therefore, if prior allocations guide subsequent spending (Heath and Soll 1996; Lukas and Howard 2023; Thaler 1985; Zhang et al. 2022), consumers who budget should ultimately spend differently than those who do not. This is not because they are spending different amounts in total. Rather, consumption differences should arise because allocations are especially drawn towards budget categories with a higher average value, and these allocations guide later purchasing.

H3: Because budget allocations are sticky, the effect of a category's average value on spending will be greater for those who budget than for those who do not budget.

---

[1] Perceptions of average may include shifts in the mode, extreme values, or most recent values (Howard and Shiri 2022; Kahneman et al. 1993). Holding all else constant, any such change may result in a change in the perceived average; however, it is possible different shifts may have different influences. For expository simplicity, we refer to any such shift in the mass of the distribution as shift in the average and leave further distinctions for future research.

We test and provide evidence for these hypotheses in three experiments. In experiment 1, we test H2a by asking whether allocators are more sensitive to a category's average value than purchasers. Experiment 2 explores the proposed process based on evaluation mode (H1). We use mouse-tracking to document the search pattern over various options and construct a proxy for category-level evaluation. We test whether the differential sensitivity to category average values between allocators and purchasers can be explained by differences in evaluation mode (H2b). Experiment 3 introduces an incentivized budgeting and spending game with imputed values. This design allows us to address an alternative interpretation that a category's average value merely reflects marginal value. We rule out this alternative explanation by holding the marginal value constant while manipulating the average value of budget categories. We observe spending decisions for all participants, enabling a test of whether budgeters and non-budgeters spend differently (H3).

All preregistrations, materials, data, and code are available at
https://researchbox.org/353&PEER_REVIEW_passcode=MIJYNO.

EXPERIMENT 1

Are allocators more sensitive to category values than purchasers? We propose they are (H2a), because allocating encourages category-level evaluations, and category-level evaluations automatically extract information about the category average. Therefore, the current study considers the interaction between category values and whether participants allocate vs. purchase.

Method

Participants. 800 Amazon Mechanical Turk (AMT) participants (M_{age} = 41; 54% female) completed this study[2].

Design and stimuli. Participants considered various potential discretionary dining and entertainment expenditures, along with their associated costs, for the coming month. In the instructions, we emphasized that necessities had already been financially accounted for, so participants were to focus entirely on discretionary dining and entertainment purchases. There were a total of 15 dining and 15 entertainment options available for purchase, several of which were adapted from prior work (Cheema and Soman 2006; Heath and Soll 1996; Sussman and Alter 2012). Each option was available for purchase in the coming month and had an associated cost ranging from $5 to $100. The two categories were constructed to have similar distributions of costs (M_{Dining} = $M_{Entertainment}$ = $31.67). The option-cost pairings were held constant for all participants and were presented in side-by-side lists. The order of elements within each list was randomized across participants (figure 1).

---

[2] There were 801 complete observations; however, two were associated with the same participant identifier. We removed the second observation from this participant to preserve naïveté.

FIGURE 1

EXAMPLE OF RANDOMLY ORDERED OPTIONS

Dining Options	Entertainment Options
An expensive modern restaurant dinner $100	Tips for a street performer $5
Snacks at a convenience store $7	Going to an amusement park $40
A meal kit delivered to your door $15	Entrance to a state park $8
Ice cream $6	A baseball game $18
Treating friends to a restaurant dinner $75	A day at the spa $100
Ordering smoothies for your family $28	Going to the museum $15
A bag of beef jerky $5	Parking to explore a new area $5
Grabbing takeout for lunch $10	Purchasing a game for your phone $7
Drinks with some friends $30	A music concert $90
Switching to organic groceries $35	Mini golf $13
A fancy date night at a restaurant $95	A ceramics class to make a teacup $95
Getting morning donuts for the office $18	Movie tickets $11
A flavored latte at Starbucks $6	Streaming subscription $13
A nice bottle of wine $25	Camping at the fairgrounds $10
Steak instead of hamburger this week $20	A boat tour $45

Note—A randomly ordered list of dining options (*left*) and entertainment options (*right*). The options and their associated costs were held constant across all participants.

Participants were randomly assigned to one of two conditions. In the *allocate* condition, participants' task was to divide $250 between the two categories. In the *purchase* condition, participants' task was to make hypothetical purchases by selecting desired options from the lists. In the purchase condition, participants could click on an item to indicate they wished to purchase it. To encourage spending based on preferences rather than mathematical convenience, participants in the purchase condition could spend any amount between $200 and $300. After making the allocation or purchase decisions, all participants rated how much they valued each category, on average. The value of each category, dining and then entertainment, was assessed

using the prompt: "On average, how much value do you get out of this category of purchases?"

Responses were collected on a 1-7 scale, anchored on 1 = "Not very much" and 7 = "A lot".

Results

Analysis plan. The dependent measure was constructed as the percentage[3] of total dining

and entertainment consumption that was allocated to dining, which we call the "Dining Share".

In both conditions, this was calculated as [dining dollars / (dining dollars + entertainment

dollars)] x 100%. In the allocate condition, (dining dollars + entertainment dollars) was always

equal to $250. In the purchase condition, (dining dollars + entertainment dollars) could vary

slightly depending on total spending. On average, consumers should consume a greater dining

share when they value dining more than entertainment; however, we predict this tendency will be

stronger for allocators than for purchasers (H2a). To test this, we first constructed a single

measure to reflect the relative preference for the dining category, compared to the entertainment

category. This was constructed as the average value of dining – the average value of

entertainment. We call this variable "Dining Preference". We also include a control variable

called "Aggregate Liking", calculated as the sum of the two category average value ratings. This

variable is not of theoretical interest and is merely used to maintain statistical equivalence to a

model using raw value ratings rather than difference measures. We dummy-coded the condition

variable as 1 for the allocate condition and 0 for the purchase condition.

We regressed the dining share on the dining preference, aggregate liking, condition, and

the two-way condition interactions. The coefficient on dining preference represents the simple

---

[3] We preregistered this as a proportion in [0, 1], rather than a percentage in [0, 100]. We rescaled this variable into a percentage for reporting convenience, which has no impact on the statistical tests or inferences.

slope on dining preference among purchasers. We expect this coefficient to be positive, as consumers should purchase more in categories they value more highly. The key test is the interaction between the dining preference and condition. This term reflects the extent to which category value differentially matters when allocating compared to purchasing. We expect this coefficient to be positive, reflecting a larger sensitivity to category value when people allocate compared to when they do not (H2a).

Dining share. Participants consumed more dining than entertainment ($M = 53.84$, $SD = 18.96$), as indicated by the mean dining share exceeding 50% ($t(799) = 5.73$, $p < .001$). There were no differences in the dining share across conditions ($M_{allocate} = 53.71$, $SD_{allocate} = 18.12$, $M_{purchase} = 53.96$, $SD_{purchase} = 19.78$; $t(798) = 0.18$, $p = .857$), suggesting allocating vs. purchasing did not have a main effect on relative dining consumption.

Among purchasers, there was a positive simple slope on relative dining preference in predicting dining share ($b = 3.01$, $se = 0.50$, $t(794) = 6.03$, $p < .001$). However, the simple slope in the allocate condition ($b = 5.33$, $se = 0.40$, $t(794) = 13.25$, $p < .001$) was significantly larger, as indicated by the interaction in the full model ($b = 2.31$, $se = 0.64$, $t(794) = 3.61$, $p < .001$, Cohen's $f = 0.13$). In other words: dining share was more strongly related to relative dining preference among allocators than among purchasers. This implies that for a given set of category preferences, allocating and purchasing lead to different outcomes. This was apparent in a model examining the simple effect of allocation versus purchasing on dining share among three subsets of participants (those who preferred dining to entertainment, those who preferred entertainment to dining, and those with no preference). Among consumers who rated the dining category higher than the entertainment category, allocating increased the share of dining options relative to purchasing ($M_{allocate} = 65.55$, $M_{purchase} = 59.09$, $t(794) = 3.04$, $p = .002$, Cohen's $d = 0.34$).

Among consumers who rated the dining category lower than the entertainment category, allocating decreased the dining share relative to purchasing ($M_{allocate}$ = 43.33, $M_{purchase}$ = 49.14, $t(794)$ = -2.94, p = .003, Cohen's d = 0.31). And among consumers who rated the dining category the same as the entertainment category, allocating and purchasing did not differ ($M_{allocate}$ = 53.78, $M_{purchase}$ = 54.44, $t(794)$ = -0.27, p = .787, Cohen's d = 0.03).

Discussion

Experiment 1 considers the link between category evaluations and hypothetical consumer decisions. Allocators were more sensitive to their category evaluations than purchasers, as indicated by the condition-by-dining preference interaction. This provides initial support for H2a. Notably, this arises in an experimental paradigm that presents categories of options by default. In both conditions, all possible expenditures were equivalently displayed and organized into categories. This suggests it is not merely the presence of categories that makes allocators more sensitive to a category's average value. Instead, we propose the key distinction is the mode of evaluation. Allocators (vs. purchasers) are hypothesized to make more category-level evaluations, even when options are already organized within clearly defined categories. To test this more directly, we introduce a modified experimental design in experiment 2.

EXPERIMENT 2

The prior experiment found evidence that allocators are more sensitive than purchasers to the average value of expenditure categories. Though we hypothesized this was due to allocating

vs. purchasing engaging different modes of evaluation, our prior data could not address this directly. The current experiment incorporates mouse-tracking to record the search patterns over various dining and entertainment options. We use this data to construct a measure of search, which serves as a proxy for evaluation mode.

Method
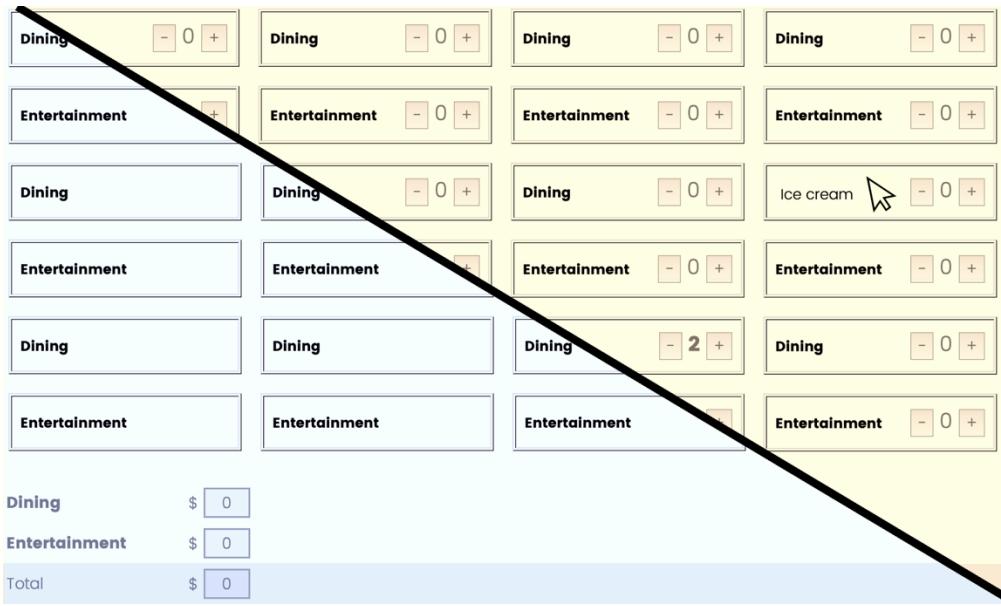
We ran three separate mouse-tracking experiments over the span of two months. These experiments were highly related—differing only in surface-level features—and all were conducted on AMT (see web appendix B for additional details). We observed variability in estimates across these experiments, such that any study in isolation (or different groupings of studies) might lead to a somewhat different statistical inference. To avoid selective reporting, we combine all three experiments into a single dataset. We emphasize this approach does not strengthen our inferences relative to selectively reporting the experiment with the most supportive data. Rather, because we observed surprising variability across experiments, we felt obligated to report all the data. Therefore, the combined analysis resembles a form of meta-analysis across our own experiments. We include separate analyses for each independent experiment version in web appendix B.

All experiment versions used the same core design and all had identical preregistrations (differing only in sample size). As in experiment 1, participants encountered various dining and entertainment options. These items were similar or identical to those used in study 1, and the full option list is provided in web appendix B. There were 12 of each type of option, arranged in a 6 x 4 grid; see figure 2. Within each row, all options were of the same type (all dining or all

entertainment). Adjacent rows never contained the same type of options. All options were hidden by default, with only the category label displayed (dining or entertainment). Options were temporarily revealed when the mouse cursor entered the box and concealed when the mouse cursor exited the box, as in MouseLab (Payne, Bettman, and Johnson 1988). This allowed us to track the sequence in which participants searched for options. Option locations were randomized within the appropriate spaces designated for either dining or entertainment.

FIGURE 2

ILLUSTRATION OF MOUSE-TRACKED OPTIONS GRID, SPLIT BY CONDITION

Note—An illustration of the 6x4 option grid, including both the allocate condition (bottom left, highlighted in blue) and the purchase condition (top right, highlighted in yellow). Each box within the grid displayed the category label by default until moused over. Mouse hovering temporarily revealed the specific option within each box (see third row of last column, in which "Ice cream" is displayed as a dining option). Allocators divided $150 between two budgets, as shown in the bottom left of the image. Purchasers made $150 of purchases using the [-] and [+] buttons. Note: color highlights included for illustrative purposes and were not included in the actual study.

As in the prior experiment, participants were randomly assigned to either the allocate or purchase condition. In the allocate condition, the goal was to divide $150 between a dining and entertainment budget. In the purchase condition, the goal was to make $150 of purchases. All options cost $10, and each was available up to four times. The cover story was that participants should allocate for or purchase discretionary items for the coming month, and every option was available a maximum of once per week. Allocators made their decisions by setting their budgets at the bottom of the page, and purchasers made their decisions by clicking buttons contained within each box (figure 2). As in the prior experiment, the allocations or purchases were used to construct the dining share as the dependent variable. We calculate this as [dollars to dining / total dollars ($150)] x 100%, thus reflecting the dining share.[4]

After making this decision, all participants evaluated the dining and entertainment categories separately. Using the same prompt as in experiment 1, they indicated how much value they got out of each category on average, on a 1-7 scale anchored on "Not very much" and "A lot".

Mouse-tracking. We observed the mouse-tracked sequence in which participants explored options. In addition to the sequence, we also captured dwell times, measured in hundredths of a second. We used this search data to construct a proxy for evaluation mode. First—to reduce noise—we followed our preregistered plan to remove any mouse-overs less than one tenth of a second. This threshold has been previously used in other mouse-tracking paradigms as the minimum threshold for processing information (Payne et al. 1988). Next, we adopt the approach

---
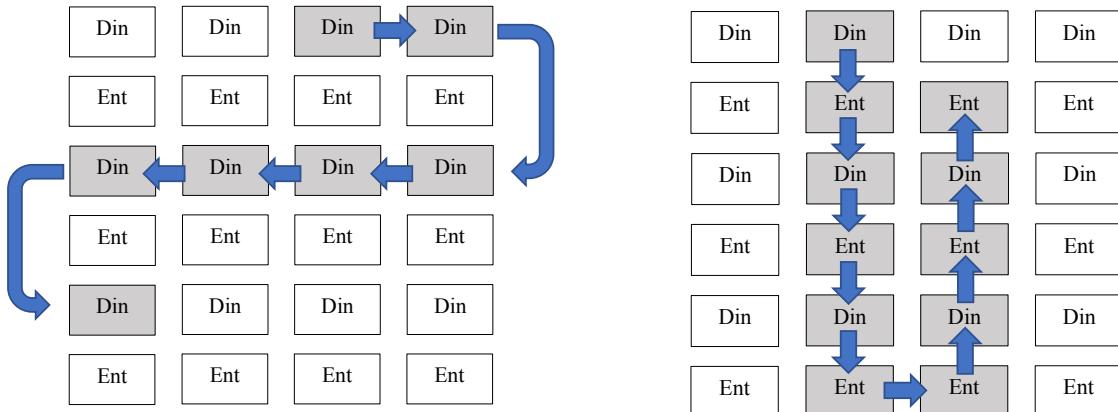
[4] Our preregistration specified using (Dining – Entertainment) rather than (Dining / (Dining + Entertainment)) x 100%. Given the fixed $150 total across the two categories, one is simply a linear transformation of the other, so no substantive nor statistical inferences change as a result of this change. It simply eases exposition.

from Payne et al. (1988) to construct a search index of the relative tendency to make within- vs.

between-category transitions. Transitions are defined as moving from any given option to a

different option. The within-category search index is constructed as (within-category transitions

– between-category transitions) / total transitions. This index therefore ranges from -1 (only

between-category transitions) to +1 (only within-category transitions).

Because this search index measures the relative tendency to search within categories vs.

between categories, it serves as a proxy for evaluation mode. Within-category search aligns with

category-level evaluations; between-category search aligns more closely with item-level

evaluations. Figure 3 illustrates two different hypothetical search patterns corresponding to these

different modes of evaluation.

FIGURE 3

EXAMPLE OF TWO HYPOTHETICAL SEARCH PATHS

NOTE—Two hypothetical mouse-tracked search patterns. The left pattern reflects all
within-category transitions and would score very high on the within-category search index ((6-
0)/(6+0) = +1). The right pattern reflects a majority between-category transitions and would
score very low on the search index ((1-9)/(1+9) = -0.8).

22
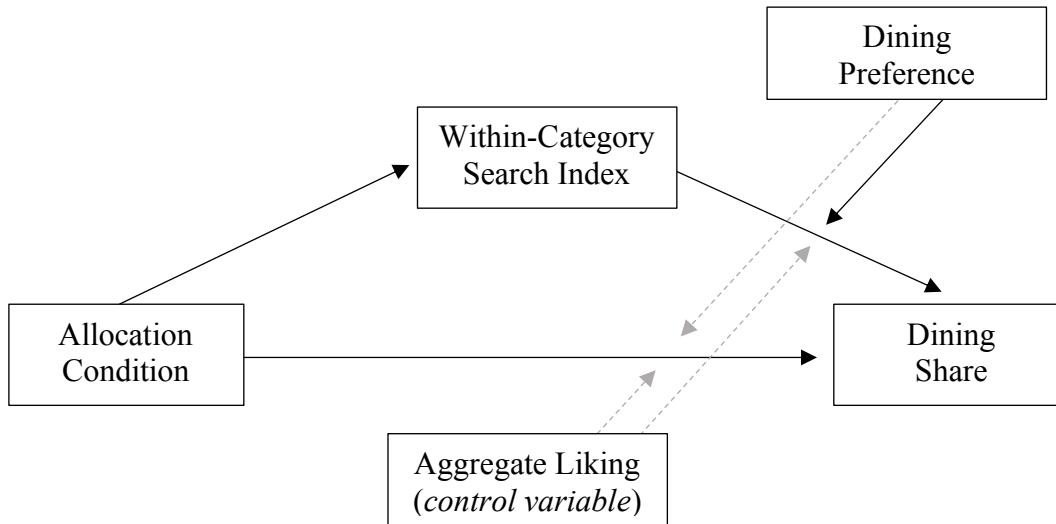
Results

A total of 3,810 complete observations were collected across the three versions of this experiment (N_1 = 1,003; N_2 = 1,404; N_3 = 1,403). Two observations were removed for being linked to an AMT identifier with a prior complete or incomplete observation. This resulted in a pre-exclusion sample of 3,808 unique observations. Following our preregistered plan, we excluded any participant who self-identified as completing the study on a mobile phone or tablet. The studies were clearly advertised on AMT as being for "computers only, no phones". This exclusion resulted in a final sample of 3,694 observations (N_1 = 972; N_2 = 1,369; N_3 = 1,353). This is an important criterion because the mouse-tracking does not work properly on these devices. Additionally, the custom JavaScript code used in this study may have mechanically prevented mobile device users in the allocation condition from progressing beyond the main decision page. The mechanical blocking of certain mobile device users may partially explain the differential attrition we observed across the three experiments. Subjects assigned to the allocate condition were less likely to complete the experiment than those in the purchase condition ($\chi^2(1)$ = 84.82; $p < .001$). Analyses of attrition across all experiments are provided in web appendix E.

Analysis plan. Like the prior study, we again focus on the dining share as the dependent measure [dollars to dining / total dollars] x 100%. Again, the relative dining preference is the average value of dining – the average value of entertainment. To improve the interpretability of main effects, we contrast-code the condition variable as +1 for the allocate condition and -1 for the purchase condition. The new variable in this study is the within-category search index, which ranges from -1 (indicating all between-category search) to + 1 (all within-category search).

The current design allows us to address two untested research questions. First, is there more within-category search in the allocate condition? If so, this would support the hypothesis that allocating encourages a category-level mode of evaluation (H1). Second, does the extent to which the within-category search index mediates the relationship between condition and the dining share depend on the relative preference for the dining category (H2b)? The underlying intuition is straightforward. When participants focus on categories (as proxied by the within-category search index), they will be more sensitive to how they perceive a category's average value. For two people with equivalent and positive preferences for dining (dining category's average value > entertainment category's average value), we would expect the person with a higher within-category search index to consume *more* dining, other things equal. This is because the person who searches within-category is more likely to consider the value of the dining category overall rather than individual options. For two people with equivalent and negative preferences for dining (dining category's average value < entertainment category's average value), we would expect the person with a higher within-category search index to consume *less* dining. This implies a statistical test of moderated mediation, asking whether the indirect effect through the search index depends on relative dining preferences. We formally test this using a Process 17 macro in R (Hayes 2012) to bootstrap the index of moderated mediation. As depicted in figure 4, we include both dining preference and aggregate liking as potential moderators and allow both variables to interact with the mediated pathway and the direct pathway. As in experiment 1, aggregate liking (the sum of both category average value ratings) is included as a control to maintain statistical equivalence to a model using only raw scores and is not of interest.
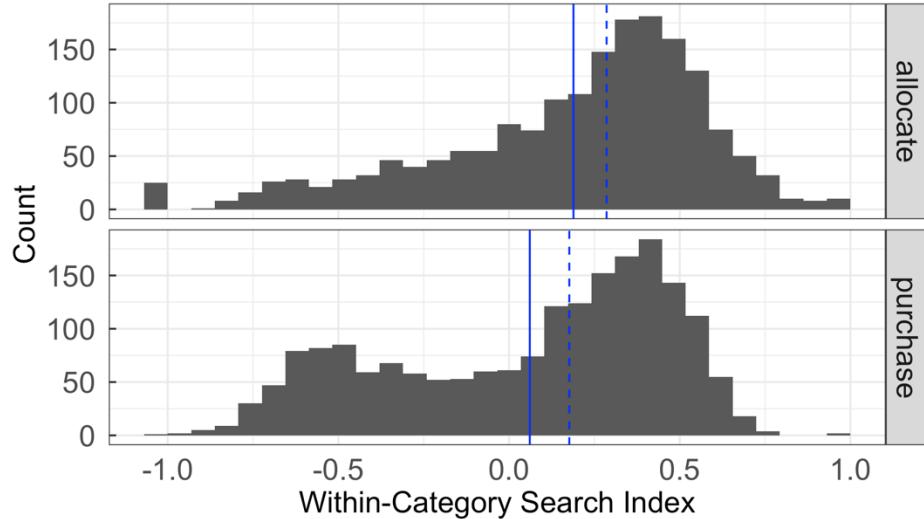
FIGURE 4

MODERATED MEDIATION MODEL

Note—Conceptual model of moderated mediation pathways. This is statistically analyzed using a Process 17 model (Hayes 2012). Aggregate Liking is a control variable used to maintain model equivalency, as discussed in experiment 1. Dotted arrows are not of theoretical interest.

Within-category search index as a proxy for evaluation mode. Participants in the allocate condition made proportionally more within-category transitions than those in the purchase condition (figure 5). This was measured as higher search indices in the allocate condition ($M =$ 0.19, $SD = 0.39$), compared to the purchase condition ($M = 0.06$, $SD = 0.41$; $t(3680) = 9.68$, $p <$ .001, Cohen's $d = 0.32$). We interpret this result as evidence of greater category-level evaluations among allocators than purchasers (H1).

FIGURE 5

WITHIN-CATEGORY SEARCH INDEX DISTRIBUTIONS, BY CONDITION

NOTE—The within category search index ranges from -1 (only between-category transitions) to +1 (only within-category transitions). Solid blue lines represent condition means. Dashed blue lines represent condition medians.

Search index-by-dining preference interaction. The effect of condition on the search index suggests the measure is a useful proxy for category-level evaluations. Therefore, we expect the link between the search index and the dining share to be moderated by the relative dining preference. Participants who focus on categories (as proxied by the search index) should be more sensitive to their relative dining preferences when deciding upon their dining share. This appears to be the case, as regressing dining share on the search index, dining preference, aggregate liking, and the two-way interactions with search index indicates a significant search index-by-dining preference interaction ($b = 0.824$, $se = 0.268$, $t(3676) = 3.07$, $p = .002$, Cohen's $f = 0.05$).

Conditional indirect effects. Taken together, (a) the effect of condition on search index and (b) the search index-by-dining preference interaction suggest the indirect effect of condition

on dining share through search index may be moderated by dining preferences. We test this moderated mediation using a Process 17 macro in R (Hayes 2012), which reveals a significant index of partial moderated mediation ($b = 0.050$, $CI_{95\%} = [0.012, 0.090]$). This indicates the indirect effect through the within-category search index depends on the relative dining preferences. We interpret this result as evidence that the differences in sensitivity to average value (captured by the dining preference) among allocators vs. purchasers is partially explained by evaluation mode (as proxied by the search index). Therefore, this provides evidence for H2b.

Model without mouse-tracked search index. We can also test for differential sensitivity to averages by condition using the same model from experiment 1 (H2a). This analysis does not include the search index, but rather regresses the dining share on condition, dining preference, aggregate liking, and the two-way condition interactions. We do not find compelling evidence for the expected condition-by-dining preference interaction, which is directionally consistent but not statistically significant ($b = 0.121$, $se = 0.111$, $t(3687), = 1.09$, $p = .275$).

Discussion

The purpose of this experiment was to explore how allocation affects evaluation mode. We find allocating (vs. purchasing) leads to more category-level evaluations, as proxied by within-category comparisons. This provides evidence for H1. We hypothesized evaluation mode is important because it impacts the sensitivity to category evaluations, like a category's perceived average value. Indeed, participants who score higher on the within-category search index are more sensitive to their dining preference when deciding upon the dining share. We consider the

entire process—how allocation affects the search index, which in turn interacts with dining preferences to affect the dining share decision—through a model of moderated mediation. Consistent with H2b, we find evidence that differential sensitivity to average value between allocators vs. purchasers can be partially explained by differences in evaluation mode.

To our surprise, we did not detect the condition-by-dining preference interaction that was observed in study 1 (H2a). Recall, in the prior study, allocators were more sensitive to their relative dining preference than purchasers. In the present study, we find only directional evidence for this relationship. Upon deeper reflection, there may be reasonable explanations for this discrepancy. Compared to experiment 1, the present design made it much more difficult to assess category value while making the dining share decision. Whereas all options were displayed in experiment 1, options were only temporarily displayed when hovered over in experiment 2. Therefore, in order to roughly equate the two paradigms on the dimension of available information, participants in experiment 2 would need to both (a) explore every option and (b) perfectly remember them. To the extent most participants did not explore every option (43% of participants did *not* examine every option) and most people cannot hold anywhere near 24 items in working memory (Miller 1956), it seems reasonable that category impressions might be considerably noisier signals for allocators in experiment 2 than they were in experiment 1.

Thus far, experiments 1 and 2 have provided evidence that compared to purchasing, allocating shifts the evaluation mode towards categories rather than items (H1). As a result, allocators are more sensitive to category evaluations than purchasers (H2a) in experiment 1, though not clearly in experiment 2. (Experiment 3 will provide an additional test of this hypothesis.) This potential difference in sensitivity appears to be partially attributable to differences in evaluation mode (H2b). Taken together, these findings expand our theoretical

understanding of *allocation*; however, they do not directly speak to the substantive implications for *spending*. Therefore, experiment 3 considers the complete budgeting process (both allocating and spending) in an incentivized paradigm.

EXPERIMENT 3

Experiment 3 has three main goals. First, we seek to explore the downstream effects of allocation on *spending*. This will allow us to consider how budgeting as a two-stage process (allocating followed by spending) depends on category evaluations. Furthermore, this will enable a direct comparison of spending between those who budget vs. those who do not, rather than comparing allocations to purchases. Second, we will induce values, rather than elicit them. Whereas the prior studies relied on measuring category evaluations, the current study will directly manipulate values. This allows us to explore our theory in a paradigm that permits clear causal inferences and is less susceptible to measurement error. Third, we aim to address an alternative interpretation for the results in both experiments 1 and 2.

Alternative interpretation of sensitivity to average category value. In both designs, we relied on the relative dining preference, which was constructed from evaluations of the average value of both dining and entertainment categories. It is possible, however, that these category evaluations were not holistic evaluations (i.e., a representation of average), but instead reflected a different dimension of value, like marginal value. For example, participants might have based their category evaluations on the value of the marginal purchase in each category. To the extent the average value and marginal value of a category may covary, we cannot completely rule out

this alternative interpretation in our prior studies. (We suggest that item-level evaluations lead to a greater relative focus on marginal value, so the reason we expect a difference between allocation and purchasing is because there is a wedge between average value and marginal value. If the measures were inadvertently capturing marginal value instead, though unlikely, the results could in principle indicate that allocation increases sensitivity to marginal value.)

We conducted a supplementary study in an attempt to disentangle average and marginal value. 606 AMT participants performed a hypothetical allocation task, in which they divided funds between a dining and entertainment account. We elicited the average value for each category by asking participants to list three typical purchases in each category and then rate how much they valued those purchases, on average. We elicited value at the margin by asking participants to identify the purchase they would most-like to make in each category if they had the additional funds and then rate its value. Regressing the dining share on the average values, *controlling for marginal value*, revealed significant coefficients on average value (both *p*s < .001; see web appendix C). While this supplementary study provides encouraging evidence that allocators are sensitive to average value above and beyond marginal value, this simple study may be susceptible to measurement error (e.g., perhaps participants were not identifying the true marginal purchase). Our experiment 3 paradigm provides a very controlled and precise design to separate the average category value from that category's marginal value.

To advance these three goals, we developed an incentivized budgeting game in which participants repeatedly allocated and spent funds to score as many consumption points as possible. This paradigm allows us to (a) observe incentivized spending for all participants (b) in a paradigm with induced values where (c) we can carefully hold constant the points of the marginal good.

Method

Participants. 970 participants (M_{age} = 41; 52% female) were recruited from AMT

completed this experiment[5].

Design overview. The budgeting game was an incentivized, multi-period decision game

in which participants spent their money on items that varied in points. All items had the same

cost but differed in the points they awarded. As in experiments 1 and 2, these items belonged to

either a dining or an entertainment category. The goal of the game was to accumulate as many of

these consumption points as possible, and this was directly incentivized with a modest additional

payout. We manipulated whether participants allocated funds prior to spending and which

category (dining or entertainment) had a higher average value.

The game was structured as occurring over a sequence of simulated weeks, though the

entire game took place in a single experimental session lasting approximately 25 minutes.

Participants played five practice weeks (which we do not analyze) and then five incentivized

weeks (which we do). Each week, participants had $230 to spend on items costing $10 each. A

set of 16 items was displayed each day (Monday through Friday), such that an entire week was

comprised of 80 total options (16 per day for 5 days). Of the 80 total options, 40 were dining

(indicated by an image of food) and 40 were entertainment (indicated by an image of event

tickets). Every participant received a single draw of 80 options, and the order of these options

---

[5] There were originally 1007 complete observations. In two cases, a single participant identifier had two complete
observations; we kept the first response from each such pair for analysis. 35 observations were excluded for having a
previous or concurrent incomplete response from the same participant identifier in the dataset, meaning the
completed observation may not have been naïve. This resulted in the final sample of 970 naïve participants. All of
our focal results (i.e., those involving the high vs. low dining average distribution) replicate if we include all 1005 or
1007 observations instead.

was randomized every week (including the practice weeks). Using the same draw of 80 items made learning the game and the point value distributions more tractable. The basic structure of a game week is depicted in figure 6.
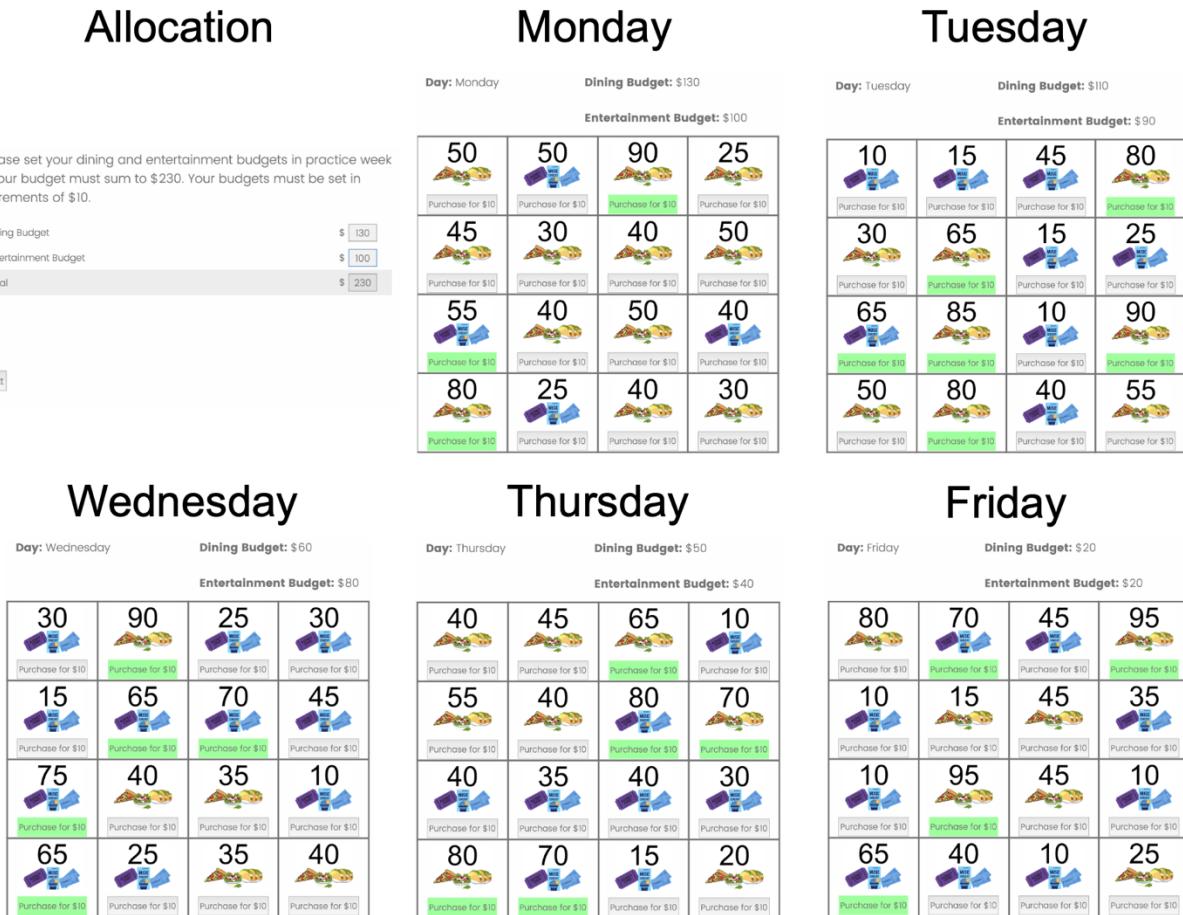
Participants used their weekly money to buy dining and entertainment items. Each item cost $10 and was worth the number of points indicated on the item, ranging from 10 to 95. Each simulated day, Monday through Friday, participants encountered a 4 x 4 grid of 16 items as depicted in figure 6. Participants could purchase as many items as they had money available; they could not exceed $230 in weekly spending. After making decisions for one day, participants were shown their purchased items and then continued to the next day's selection. Participants were not permitted to revise previous decisions. Unused money carried over from day to day within each week but did not carry over from one week to the next. After five practice weeks there were five incentivized weeks with total incentives averaging approximately 20% of overall compensation. Realized bonuses among non-excluded participants ranged from $0 to $1.25, with a median of $0.80. Bonuses were paid in addition to a fixed $3.25 participation payment.

Budget manipulation. Participants were randomly assigned to either the budget or no-budget condition. In the budget condition, participants started each week by allocating their $230 between a dining and entertainment budget (in $10 increments). They then encountered five sequential days, during which they could spend up to $230. Expenses were automatically tracked to the appropriate budget, and participants could see the remaining balance in each budget; however, allocations were non-binding. Participants knew they were allowed to disregard their allocations, so long as their spending did not exceed the weekly constraint of $230. In the no-budget condition, participants did not allocate weekly funds prior to encountering the five days

in which they could spend. Identical to budgeters, the only spending constraint was that non-

budgeters could not exceed $230 of weekly spending.

FIGURE 6

OVERVIEW OF THE BUDGETING GAME STAGES

Note—Each of five practice weeks and five game weeks followed these steps. First, participants in the budgeting condition allocated $230 across a dining budget and an entertainment budget. Next, all participants made purchase decisions each day for each of five days. Items were randomized across days, and participants faced each day's screen in sequence. On some days, participants may have needed to scroll to the right to see all budget information. For presentation purposes, this information has been condensed into this image.

Category average value manipulation. All items ranged in value from 10 to 95 points (in

5-point intervals). We manipulated the distributions of points within the dining and

entertainment categories, such that either dining or entertainment had a higher average value. Rather than manipulating the entire range of the distribution (e.g., from 10 to 95), we instead divided this range into the high-point items (those worth 60 or more points) and the low-point items (those worth less than 60 points). This allowed us to independently manipulate whether dining had a higher or lower average value within both point regions. The threshold of 60 points was deliberate, as this was precisely the value of the marginal purchase for both categories, as will be discussed shortly.

The high-point distributions were manipulated to be either higher for dining (realized dining vs. entertainment means across participants: 81 vs. 70) or lower for dining (74 vs. 82). Similarly, the low-point distributions were manipulated to be either higher for dining (38 vs. 26), or lower for dining (25 vs. 37). We preregistered a clear interest in the effect of the high-point distribution condition as these were the items most likely to be purchased and be included in the evaluation set (Bettman and Park 1980; Payne 1976). Therefore, for ease of explication, we focus on reporting and interpreting the results of manipulating average values in this high-point region, which we will refer to as the "dining average" manipulation. Complete analyses and additional discussion of the low-point distribution are included in web appendix D.

Disentangling average and marginal values. We designed the distributions with one additional key feature in mind. Specifically, we held constant the number of items belonging to each category in the high-point and low-point distributions. The high-point distribution always had exactly 23 items worth 60 points or more (14 dining and 9 entertainment). The low-point distribution always had exactly 57 items worth less than 60 points (26 dining and 31 entertainment). Therefore, given the $230 of weekly funds, an omniscient player would always allocate for and/or purchase the 23 items with point-values of 60 or higher. This holds the point-

value of the marginal purchase across categories constant at 60. Regardless of whether the average value of dining is higher or lower than entertainment in this region, dividing the $230 between 14 dining items and 9 entertainment items is guaranteed to result in the most points. As detailed in web appendix D, we take additional steps to construct a distribution that effectively holds this marginal value constant, even as participants slightly deviate from the 14/9 split.

The game was designed to test whether budgeters vs. non-budgeters exhibit different sensitivities to a category's average value within an engaging and incentivized paradigm. The design provides a high degree of experimental control, enabling us to cleanly separate the effect of average value from marginal value in category evaluations. Whereas the prior experiments were limited in comparing allocations to purchases (H2a), the current experiment also compares spending between budgeters and non-budgeters. Specifically, because all participants spend their money (differing only in whether they are assigned to allocate funds in advance of spending), we can analyze the downstream spending effects. Budgets were not binding, meaning participants could ignore them or deviate from their allocations. Despite this, we expect spending to be more sensitive to a category's average value for budgeters who previously allocated than people who spend without allocating (H3).

Summary. To recap, participants in the budgeting condition repeatedly allocated funds between two budgets, and all participants purchased items to earn points. The distributions of items were structured such that the dining category had either a higher or lower average value, but the marginal value was equated across dining and entertainment. Participants were well-informed (five comprehension questions, reported in web appendix D) and well-trained in the paradigm (five practice weeks). Participants knowingly faced the same weekly distribution of items for the entire session of practice and incentivized weeks to facilitate learning. Within this

incentivized game, we examine whether budgeters vs. non-budgeters respond differently to category average values when marginal values are held constant.

Results

Because of the potential for noise and extreme responses, we preregistered that we would exclude participants who failed to buy at least 50% of the most-valuable options. Of 970 participants, 149 purchased fewer than 50% of these items across all 5 game weeks, likely indicating inattentiveness or misunderstanding, and were thus excluded, leaving a final sample of 821. The interpretation of the preregistered analyses does not meaningfully change if noise participants are included (see web appendix D).
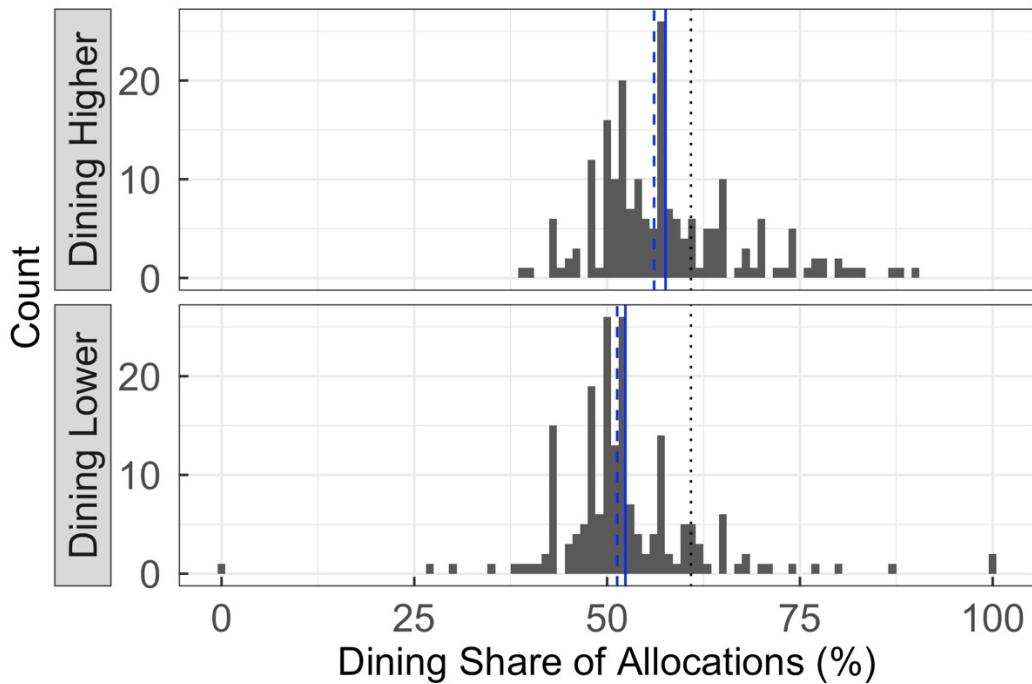
Dining share measure. We construct two different measures of the dining share: one for allocations and one for spending. In both cases, the dependent variable is calculated as [dollars of dining / (dollars of dining + dollars of entertainment)] x 100%[6]. For example, $150 to dining and $80 to entertainment equates to a dining share of 65%. For those in the budget condition, we can examine the dining share for both allocation and final spending. For those in the no-budget condition, we only consider the dining share of spending.

---

[6] Our preregistration specified (Dining – Entertainment) rather than (Dining / (Dining + Entertainment) x 100%). Because some participants did not exhaust their budget, these two measures are not perfectly deterministic transformations of one another. They are, however, extremely highly correlated ($r = .994$), neither is clearly dictated as a preferred measure, and none of our key results hinge on which metric we use. We use dining share to maintain consistency across studies for the sake of readability.

Analysis plan. We preregistered two levels of analysis. First, we consider only participants in the budget condition and regress their dining share of allocation on dining average condition (+1 = dining high, -1 = dining low), the low-point dining average condition (+1 = dining high, -1 = dining low), and their interaction. We are interested in whether category average value (specifically among the high-point items, as captured by the dining average condition variable) affects the dining share in an environment in which marginal value is held constant across categories. The second analysis considers all participants and examines the dining share of spending. Here, we are specifically interested in whether budgeters and non-budgeters differ in their sensitivity to the dining average condition. We regress the dining share of spending on dining average, the low-point dining average, budget (+1 = yes, -1 = no), and all two- and three-way interactions.

First analysis: sensitivity to average among allocators. The first analysis considers only participants in the budget condition. As previously noted, we are interested in the effect of the dining average condition (reflecting the average value among the high-point items that are more frequently purchased). Participants allocated a greater dining share when the dining average was higher ($M = 57.58$, $SD = 9.47$), compared to lower ($M = 52.37$, $SD = 9.90$; $b = 2.61$, $se = 0.49$, $t(390) = 5.35$, $p < .001$, Cohen's $d = 0.54$; figure 7). This indicates allocators are sensitive to the average value of budget categories, even when the marginal value of each category is held constant.

FIGURE 7

BUDGETS BY DINING AVERAGE CONDITION

NOTE—The dependent measure, the dining share of allocations, by dining average distribution. Solid blue lines represent condition means. Dashed blue lines represent condition medians. Dotted black lines represent the value-maximizing allocation ($140 to dining and $90 to entertainment; a 61% dining share).

Though not our focus, we also consider the effect of manipulating average value in the low-point region of the distribution (options with less than 60 points, which were less likely to be purchased). Participants assigned to see higher dining averages in this region allocated marginally more funds to dining than entertainment ($M = 55.88$, $SD = 10.33$) than those in the low average condition ($M = 54.21$, $SD = 9.66$; $b = 0.83$, $se = 0.49$, $t(390) = 1.71$, $p = .089$, Cohen's $d = 0.17$). This effect was significantly smaller than the focal effect of the dining average condition in the high-point region ($t(390) = 2.57$, $p = .010$).
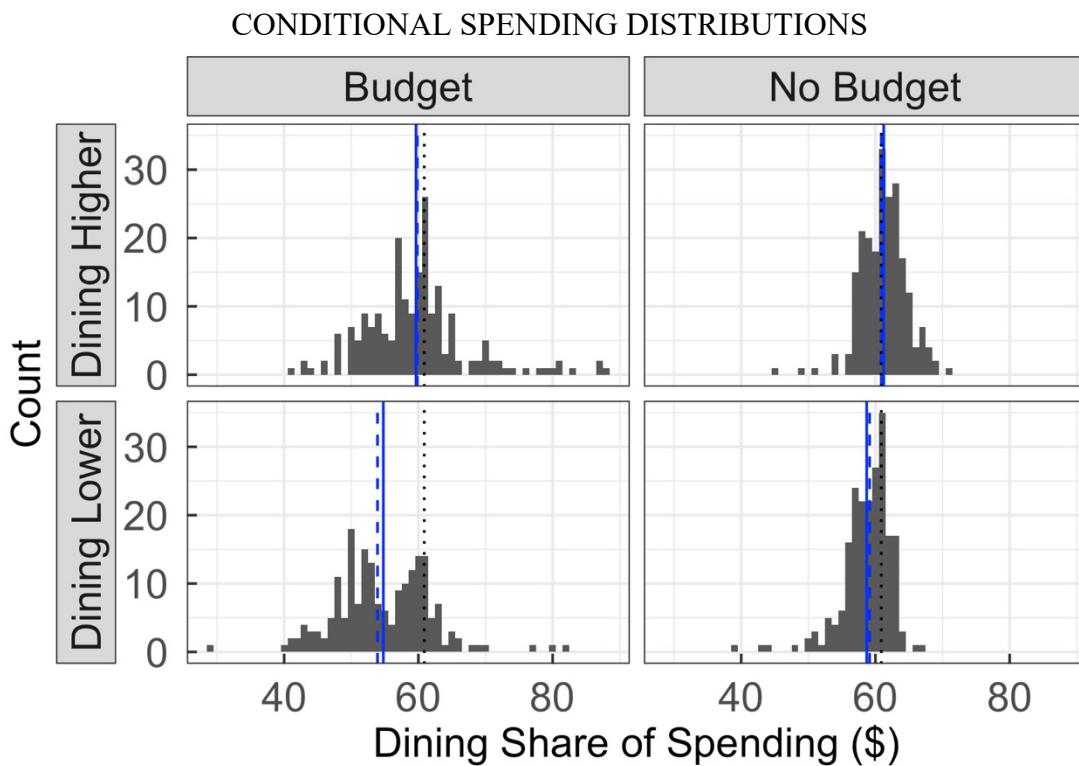
Prior to presenting the second analysis and set of results, we present supplementary evidence that budgets predict spending. This exploratory analysis bridges the first finding (allocators are sensitive to average value, holding marginal value constant) and the next analysis, which considers how budgeters vs. non-budgeters spend differently. In the budget condition, people who allocated relatively more to dining budgets than entertainment budgets spent relatively more on dining than entertainment ($b = 2.50$, $se = 0.12$, $t(392) = 20.39$, $p < .001$). This is not attributable merely to individual differences. An item-level analysis including a rich set of controls including participant fixed effects reveals an additional $10 in a dining budget is associated with an additional 3.2 percentage point increase in the probability of purchasing a dining item but only a 2.0 percentage point increase in the probability of purchasing an entertainment item. Conversely, an additional $10 in an entertainment budget is associated with only an additional 2.2 percentage point increase in the probability of purchasing a dining item, but a 3.2 percentage point increase in the probability of purchasing an entertainment item (difference in differences: $t(65) = 7.78$, $p < .001$; see web appendix D for details). Therefore, budget allocations appear to be sticky, and we should expect any differences in allocation (based on the dining average condition) will have downstream impacts on spending.

Second analysis: budget-by-dining average interaction. Recall, we regressed the dining share of spending on the budget condition, the dining average condition, the low-point dining average condition, and all two- and three-way interactions. The preregistered goal of this analysis was to test whether the effect of dining average was different for budgeters vs. non-budgeters. As expected, the effect of dining average varied depending on the presence of budgets ($b = 0.54$, $se = 0.20$, $t(813) = 2.78$, $p = .006$, Cohen's $f = 0.10$). This provides direct support for

H3, which predicts spending will be more sensitive to average value for budgeters than non-budgeters. Specifically, participants who set budgets were more sensitive to the dining average (M_{higher} = 59.65, SD_{higher} = 7.87 vs. M_{lower} = 54.78, SD_{lower} = 7.13) than those who did not set budgets (M_{higher} = 61.23, SD_{higher} = 3.40 vs. M_{lower} = 58.70, SD_{lower} = 3.61). See figure 8. Budgeting vs. not budgeting did not affect sensitivity to the average values in the low-point condition (b = -0.30, se = 0.20, $t(813)$ = -1.53, p = 0.126) and was significantly smaller than the effect of the focal dining average ($t(813)$ = 3.12, p = .002)[7]. All other interactions were non-significant (ps > .1).

FIGURE 8

CONDITIONAL SPENDING DISTRIBUTIONS

NOTE—Dining share of spending by dining average distribution. Solid blue lines represent condition means. Dashed blue lines represent condition medians. Dotted black lines represent value-maximizing spending ($140 to dining and $90 to entertainment; a 61% dining share).

---

[7] Given difference in variance across conditions, we repeated this analysis with robust standard errors. No substantive nor statistical conclusions changed.

Additional analysis: comparing allocations to spending. This experiment also permits a rather imperfect and non-preregistered test of H2a, which was tested in experiments 1-2. Recall, experiment 1 found allocators are more sensitive to a category's average value than purchasers; however, this relationship was not significant in experiment 2. While our experimental game was not specifically designed to test H2a, one can compare the dining share of allocations for budgeters with the dining share of spending for non-budgeters. There is a main effect of dining average ($t(813) = 8.04$, $p < .001$) and a main effect of budgeting ($t(813) = -10.11$, $p < .001$). These are qualified by the expected interaction, such that the effect of dining average is greater for allocations than spending ($t(813) = 2.56$, $p = .011$).

Discussion

Experiment 3 was designed with three main goals in mind. First, the study allowed us to compare spending across budgeters and non-budgeters in our incentivized game. We found budgeters' spending was more sensitive to a category's average value than non-budgeters' spending was (H3). Second, we manipulated averages through imputed values, thus reducing prior concerns about the measurement error associated with self-reported category values. Third, we observed sensitivity to average value—and importantly, differential sensitivity between budgeters and non-budgeters—while holding marginal value constant. Taken together, this experiment advances our understanding in finding that consumers with identical preferences (to maximize point values) are differentially sensitive to category average values depending on whether or not they budget, thus leading to differences in spending.

The game design is quite distinct from the other experiments. This experiment provided an engaging and incentivized repeated decision task in which there was ample opportunity to learn, and comprehension was generally high (see web appendix D). The participant experience in this experiment is clearly different than in our other studies. Consistent findings across such varied designs provides a reassuring signal of the generalizability of the findings across contexts.

GENERAL DISCUSSION

Consumers' budget allocations matter because they affect spending. The current research explores the unique psychology of resource allocation. Compared to making purchases, allocation induces a relatively greater focus on category-level evaluations. As a result of focusing on categories, consumers who allocate appear especially sensitive to a category's average value. Consumers generally adhere to the budgets they set, so the sensitivity to average value at the time of allocation affects downstream spending. Therefore, an important contribution of this work is that budgeting (vs. not budgeting) changes the composition of spending even if and when it does not change the amount of spending.

Future Directions

Measuring central tendency. In experiments 1 and 2, consumers reported their own assessment of average category value; in experiment 3, we manipulated a distribution of values. Across these studies, it is possible that participants reported or attended to different measures of central tendency (e.g., mean, median, mode). Though ensemble perception has traditionally

focused on the arithmetic mean of groups and sets (Ariely 2001; Haberman and Whitney 2009; Whitney and Yamanashi Leib 2018; Woiczyk and Le Mens 2021), we acknowledge there are a multitude of possible measures of average.

Recent work suggests there may be meaningful distinctions between such metrics in certain contexts (Howard et al. 2022), but such distinctions are unlikely to qualitatively impact our findings. First of all, our core interest is not in distinguishing between related measures of central tendency (Howard and Shiri 2022), but rather to examine sensitivity to a category's summary representation. Furthermore, category value represented by the mean, median, and mode are all distinct from marginal value, which we hold constant in experiment 3. Across experiments, we demonstrate the robustness of average value (as a proxy for central tendency) by measuring it in different ways. In experiments 1 and 2, we ask participants for their own evaluations of the average value of a category of purchases. We do not specify a specific functional form, but rather observe the response to the participant's interpretation of "average". In experiment 3, we calculate an arithmetic mean from distributions of values. The same key pattern of results exists across both approaches. In other consumer contexts, there may be situations in which it is more useful to distinguish between measures such as mean, median, and mode. For example, a consumer who splurges on a rare, extravagant vacation might become more likely to perceive differences between the mean and median value of their entertainment purchases. This provides a potential area for future research.

Evaluating budget categories. The preceding section discussed how consumers might apply different functional forms to summarize a distribution of values. Another area for future research is the extent to which various purchases are or are not evaluated within a distribution of

values (regardless of how central tendency is assessed). In experiment 3, we find participants are more sensitive to average value for high-point purchases than low-point purchases. This raises the possibility that consumers put differential weight on values across the distribution (Bear et al. 2020) or edit out low-value options from consideration (Kahneman and Tversky 1979). Further understanding of this topic may be especially important when consumers have budget categories spanning a wide array of possible values (such as in our experiment 3). Finally, valuable experiences which are highly accessible may encourage consumers to allocate more money to a budget, even if the value of that experience is unlikely to meaningfully affect the value of consumption offered by a budget category.

Constraining consumption. Self-control considerations are key motivating reasons for budgeting (Krishnamurthy and Prokopec 2010; Thaler 1980, 1999; Wertenbroch 1998). If consumers are concerned that their short-run selves will selfishly overconsume at the expense of their long-run selves, they may seek to constrain short-run spending opportunities by setting strict budgets. This characterization emphasizes a potential factor missing from our current analysis: There can be multiple dimensions of value which can be realized over different time horizons and are sometimes in conflict with one another (e.g., short-run value, like taste, vs. long-run value, like health). Our inquiry has collapsed value into a single dimension, and thus does not speak to such self-control issues. Future research could address this by considering domains with different short-run and long-run benefits and orthogonally manipulate the average value of each.

Additional predictors of evaluation mode. We find allocating (vs. purchasing) induces greater relative focus on category-level evaluations than on item-level evaluations. However, there are likely additional predictors of evaluation mode. As previously discussed, category- vs. item-level evaluation modes somewhat resemble distinct processing styles, such as holistic vs. analytic thinking (Nisbett et al. 2001). To the extent these processing styles may be relatively stable across people and cultures, (Hildebrand, Harding and Hadi 2019; Li et al. 2018; Masuda and Nisbett 2001; Monga and John 2007) there may be meaningful and identifiable individual differences in category- vs. item-level evaluations. Furthermore, among allocators, the nature and the complexity of a task may further reinforce a given evaluation mode. We suspect complexity (e.g., the accessibility of possible purchases, the number of possible purchases, the number of budgets, the duration of budgeting periods, etc.) will encourage category-level thinking as a simplification strategy. Therefore, both individual tendencies to prefer category-level evaluations as well as task complexity might increase the sensitivity to a category's average value.

Heterogeneity in sensitivity to averages. In addition to the important inter-group differences, we also observe considerable intra-group differences in budget allocation, suggesting the presence of meaningful heterogeneity in allocation decisions (see figure 7). What drives this heterogeneity? Prior examinations of cost-benefit reasoning have examined education and training in economics (e.g., Larrick, Nisbett and Morgan 1993), suggesting they may be plausible contributors. We conjecture that forward-thinking consumers (e.g., those who plan ahead or consider potential outcomes; Lynch et al. 2010; Nenkov, Inman and Hulland 2008) may

be less likely to be sensitive to the average when budgeting, as planners are more likely to

consider their opportunity costs (Bartels and Urminsky 2015; Fernbach et al. 2015; Spiller 2011).

Implications

Budgeting patterns. A subtle implication of the current findings is that consumers may

allocate too much (from a value-maximization perspective) to categories from which they

perceive the greatest average value, all else equal. Consider experiment 3, in which we held

constant the set of items that would earn the most points (and the largest real bonus payment).

Allocating or spending in line with a category's average value dragged some participants away

from this value-maximizing bundle. If real consumers place some weight on the average rather

than the best and most desirable purchases, categories with a few stand-out favorites are likely to

draw an outsized wallet share. Deliberate attempts to prioritize and attend to budgets could even

exacerbate this effect, as focusing on what they value may lead consumers to give greater weight

to typical or salient category exemplars.

Budgeting tools. The current work suggests a potential dimension for budgeting tools to

focus on: recouping value at the margin. As budgeting encourages category-level evaluations,

this has potential benefits and costs. As a benefit, it enables consumers to see the whole picture.

But as a cost, they may rely on a holistic value and miss out on value at the margin, as in

experiment 3. As budgeting tools in the fintech space like Mint, Acorns, and Personal Capital

continue to grow in popularity, they have the potential to shape the kinds of financial decisions

consumers make. Such budgeting tools provide ample feedback about spending performance,

relative to allocated levels (e.g., being under or over budget). However, the usefulness of this

performance feedback is necessarily conditioned upon the quality of budget allocations. The

current findings suggest that left to their own devices, consumers will make budget allocations in

accordance with the perceived average value of their budget categories. This may come at the

expense of higher-value expenditures. Therefore, information architects who are interested in

shifting consumption back towards the marginal expenditure might offer feedback about

allocation performance or allocation strategies and encourage consideration of specific expenses.

For example, rather than encouraging allocating to categories that are best-liked or most-

important, one might want to encourage allocating to categories to ensure not missing out on the

best-liked or most-important purchases. Additionally, the strategic organization of budget

categories could be used to attract or discourage allocations to a given category.

Cascading implications. Finally, these findings with respect to budget allocations are

likely to have additional downstream impacts, because these are not outcomes that disappear in

equilibrium. Neither prior work in ensemble perception (Whitney and Yamanashi Leib 2018) nor

our current work on budget allocation finds that these patterns are attenuated with experience;

instead, they can be reinforced or exacerbated, as consumers drift further towards allocations that

equate average values.

Consumers use budgets to guide and manage their spending. While budgets may help

consumers to stay on track in terms of their *level* of spending, budgets may also change the

composition of spending by changing how values are assessed. As budget setting favors

categories with higher average values, budgeting changes how people evaluate options, how they

spend, where they spend, and ultimately, what they consume.

REFERENCES

Amasino, Dianna R., Jack Dolgin, and Scott A. Huettel (2023), "Eyes on the Account Size: Interactions between Attention and Budget in Consumer Choice," *Journal of Economic Psychology*, 97, 102632.

Ariely, Dan (2001), "Seeing Sets: Representation by Statistical Properties," *Psychological Science*, 12(2), 157–62.

Bardolet, David, Craig R. Fox, and Dan Lovallo (2011), "Corporate Capital Allocation: A Behavioral Perspective," *Strategic Management Journal*, 32, 1465–83.

Bartels, Daniel M. and Oleg Urminsky (2015), "To Know and to Care: How Awareness and Valuation of the Future Jointly Shape Consumer Spending," *Journal of Consumer Research*, 41(6), 1469–85.

de Bartolome, Charles A. (1995), "Which Tax Rate Do People Use: Average or Marginal?," *Journal of Public Economics*, 56(1), 79–96.

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

Bettman, James R. and Whan C. Park (1980), "Effects of Prior Knowledge and Experience and Phase of the Choice Process on Consumer Decision Processes: A Protocol Analysis," *Journal of Consumer Research*, 7(3), 234–48.

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.

Choe, Yuna and Christina Kan (2021), "Budget Depreciation: When Budgeting Early Increases Spending," *Journal of Consumer Research*, 47(6), 937–58.

Chong, Sang Chul and Anne Treisman (2003), "Representation of Statistical Properties," *Vision Research*, 43, 393–404.

Davison, Michael and Dianne McCarthy (1988), *The Matching Law: A Research Review*, Routledge.

Du, Rex Y. and Wagner A. Kamakura (2008), "Where Did All That Money Go? Understanding How Consumers Allocate Their Consumption Budget," *Journal of Marketing*, 72, 109–31.

Fernbach, Philip M., Christina Kan, and John G. Lynch (2015), "Squeezed: Coping with Constraint through Efficiency and Prioritization," *Journal of Consumer Research*.

Gathergood, John, Neale Mahoney, Neil Stewart, and Joerg Weber (2019), "How Do Individuals Repay Their Debt? The Balance-Matching Heuristic," *American Economic Review*, 109(3), 844–75.

Gottfries, Nils and Keith Hylton (1987), "Are MIT Students Rational?: Report on a Survey," *Journal of Economic Behavior & Organization*, 8(1), 113–20.

Haberman, Jason and David Whitney (2007), "Rapid Extraction of Mean Emotion and Gender from Sets of Faces," *Current Biology*, 17(17).

——— (2009), "Seeing the Mean: Ensemble Coding for Sets of Faces," *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–34.

Hastings, Justine S. and Jesse M. Shapiro (2013), "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks," *The Quarterly Journal of Economics*, 128(4), 1449–98.

——— (2018), "How Are SNAP Benefits Spent? Evidence from a Retail Panel," *American Economic Review*, 108(12), 3493–3540.

Hayes, Andrew F. (2012), "PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling."

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

Herrnstein, Richard J., George F. Loewenstein, Dražen Prelec, and William Vaughan Jr (1993), "Utility Maximization and Melioration: Internalities in Individual Choice," *Journal of Behavioral Decision Making*, 6(3), 149–85.

Herrnstein, Richard J. and Dražen Prelec (1991), "Melioration: A Theory of Distributed Choice," *Journal of Economic Perspectives*, 5(3), 137–56.

Hildebrand, Diogo, R. Dustin Harding, and Rhonda Hadi (2019), "Culturally Contingent Cravings: How Holistic Thinking Influences Consumer Responses to Food Appeals," *Journal of Consumer Psychology*, 29(1), 39–59.

Hossain, Mehdi Tanzeeb (2018), "How Cognitive Style Influences the Mental Accounting System:Role of Analytic versus Holistic Thinking," *Journal of Consumer Research*.

Howard, Ray Charles, David J. Hardisty, Abigail B. Sussman, and Marcel F. Lukas (2022), "Understanding and Neutralizing the Expense Prediction Bias: The Role of Accessibility, Typicality, and Skewness," *Journal of Marketing Research*, 59(2), 435–52.

Howard, Ray Charles and Amin Shiri (2022), "What Is Average?," *Working Paper*.

Ito, Koichiro (2014), "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," *American Economic Review*, 537–63.

Jia, Miaolei, Xiuping Li, and Aradhna Krishna (2020), "Contraction with Unpacking: When Unpacking Leads to Lower Calorie Budgets," *Journal of Consumer Research*, 46(5), 853–70.

Kahneman, Daniel, Barbara L. Fredrickson, Charles A. Schreiber, and Donald A. Redelmeier (1993), "When More Pain Is Preferred to Less: Adding a Better End," *Psychological Science*, 4(6), 401–5.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263–92.

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

Larrick, Richard P., Richard E. Nisbett, and James N. Morgan (1993), "Who Uses the Cost-Benefit Rules of Choice? Implications for the Normative Status of Microeconomic Theory," *Organizational Behavior and Human Decision Processes*, 56(3), 331–47.

Larson, Jeffrey S. and Ryan Hamilton (2012), "When Budgeting Backfires: How Self-Imposed Price Restraints Can Increase Spending," *Journal of Marketing Research*, 49(2), 218–30.

Le Lec, Fabrice and Benoît Tarroux (2020), "On Attitudes to Choice: Some Experimental Evidence on Choice Aversion," *Journal of the European Economic Association*, 18(5), 2108–34.

Li, Liman Man Wai, Takahiko Masuda, Takeshi Hamamura, and Keiko Ishii (2018), "Culture and Decision Making: Influence of Analytic Versus Holistic Thinking Style on Resource Allocation in a Fort Game," *Journal of Cross-Cultural Psychology*, 49(7), 1066–80.

Liebman, Jeffrey B. and Richard J. Zeckhauser (2004), "Schmeduling," *Working Paper*.

Lukas, Marcel F. and Ray Charles Howard (2023), "The Influence of Budgets on Consumer Spending," *Journal of Consumer Research*, 49(5), 697–720.

Lynch, 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), 108-128.

Masuda, Takahiko and Richard E. Nisbett (2001), "Attending Holistically Versus Analytically: Comparing the Context Sensitivity of Japanese and Americans," *Journal of Personality and Social Psychology*, 81(5), 922–34.

McDowell, Jack J. (2013), "On the Theoretical and Empirical Status of the Matching Law and Matching Theory," *Psychological Bulletin*, 139(5), 1000.

Miller, George A. (1956), "The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological review*, 63(2), 81.

Monga, Alokparna B. and Deborah R. John (2007), "Cultural Differences in Brand Extension Evaluation: The Influence of Analytic versus Holistic Thinking," *Journal of Consumer Research*, 33(4), 529–36.

——— (2008), "When Does Negative Brand Publicity Hurt? The Moderating Influence of Analytic versus Holistic Thinking," *Journal of Consumer Psychology*, 18(4), 320–32.

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.

Mrkva, Kellen and Leaf Van Boven (2017), "Attentional Accounting: Voluntary Spatial Attention Increases Budget Category Prioritization," *Journal of Experimental Psychology: General*, 146(9), 1296–1306.

Nenkov, Gergana Y., Jeffrey J. Inman, and John Hulland (2008), "Considering the Future: The Conceptualization and Measurement of Elaboration on Potential Outcomes," *Journal of Consumer Research*, 35(1), 126–41.

Nisbett, Richard E., Incheol Choi, Kaiping Peng, and Ara Norenzayan (2001), "Culture and Systems of Thought: Holistic versus Analytic Cognition," *Psychological Review*, 108(2), 291–310.

Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance*, 16, 366–87.

Payne, John W., James R. Bettman, and Eric J. Johnson (1988), "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology: Learning Memory and Cognition*, 14(3), 534.

Peetz, Johanna and Roger Buehler (2009), "Is There a Budget Fallacy? The Role of Savings Goals in the Prediction of Personal Spending," *Personality and Social Psychology Bulletin*, 35(12), 1579–91.

Rachlin, Howard and David I. Laibson (1997), *The Matching Law: Papers in Psychology and Economics*, Cambridge, MA: Harvard Univ. Press.

Rees-Jones, Alex and Dmitry Taubinsky (2020), "Measuring 'Schmeduling,'" *The Review of Economic Studies*, 87(5), 2399–2438.

Shin, Jeong-Shik (1985), "Perception of Price When Price Information Is Costly: Evidence from Residential Electricity Demand," *The Review of Economics and Statistics*, 67(4), 591–98.

Smith, Stephanie M. and Stephen A. Spiller (2023), "Consumers Undervalue Multi-Option Alternatives in Two-Stage Choices," *Working Paper*.

Soman, Dilip and Amar Cheema (2011), "Earmarking and Partitioning: Increasing Saving by Low-Income Households," *Journal of Marketing Research*, 48(SPL), S14–22.

Sood, Sanjay, Yuval Rottenstreich, and Lyle Brenner (2004), "On Decisions That Lead to Decisions: Direct and Derived Evaluations of Preference," *Journal of Consumer Research*, 31, 17–25.

Spiller, Stephen A. (2011), "Opportunity Cost Consideration," *Journal of Consumer Research*, 38(4), 595–610.

Spiller, Stephen A. and Dan Ariely (2020), "How Does the Perceived Value of a Medium of Exchange Depend on Its Set of Possible Uses?," *Organizational Behavior and Human Decision Processes*, 161(July), 188–200.

Stilley, Karen M., J. Jeffrey Inman, and Kirk L. Wakefield (2010), "Planning to Make Unplanned Purchases? The Role of In-Store Slack in Budget Deviation," *Journal of Consumer Research*, 37(2), 264–78.

——— (2010), "Spending on the Fly: Mental Budgets, Promotions, and Spending Behavior," *Journal of Marketing*, 74(3), 34–47.

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), 800–814.

Sussman, Abigail B. and Rourke L. O'Brien (2016), "Knowing When to Spend: Unintended Financial Consequences of Earmarking to Encourage Savings," *Journal of Marketing Research*, LIII(October), 790–803.

Thaler, Richard H. (1980), "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization*, 1(1), 39–60.

——— (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4(3), 199–214.

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

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

Ülkümen, Gülden, Manoj Thomas, and Vicki G. Morwitz (2008), "Will I Spend More in 12 Months or a Year? The Effect of Ease of Estimation and Confidence on Budget Estimates," *Journal of Consumer Research*, 35(2), 245–56.

Watamaniuk, Scott N. J. and Andrew Duchon (1992), "The Human Visual System Averages Speed Information," *Vision Res*, 32(5), 931–41.

Wertenbroch, Klaus (1998), "Consumption Self-Control by Rationing Purchase Quantities of Virtue and Vice," *Marketing Science*, 17(4), 317–37.

West, Colin, Gülden Ülkümen, Patrycja Arundel, and Craig R. Fox (2022), "The Choice Architecture of Personal Budgeting Tools: Deconstructing the Effects of Budget Partitioning," *Working Paper*.

Whitney, David and Allison Yamanashi Leib (2018), "Ensemble Perception," *Annual Review of Psychology*, 69, 105–29.

Woiczyk, Thomas K.A. and Gaël Le Mens (2021), "Evaluating Categories From Experience: The Simple Averaging Heuristic," *Journal of Personality and Social Psychology*, 121(4), 747–73.

Yamanashi Leib, Allison, Kelly Chang, Ye Xia, Andy Peng, and David Whitney (2020), "Fleeting Impressions of Economic Value via Summary Statistical Representations," *Journal of experimental psychology. General*, 149(10), 1811–22.

Zelizer, Viviana (1997), *The Social Meaning of Money*, Princeton University Press.

Zhang, C. Yiwei and Abigail B. Sussman (2018), "Perspectives on Mental Accounting: An
Exploration of Budgeting and Investing," *Financial Planning Review*, 1(1–2), e1011.

Zhang, C. Yiwei, Abigail B. Sussman, Nathan Wang-Ly, and Jennifer Lyu (2022), "How
Consumers Budget," *Journal of Economic Behavior & Organization*, 204, 69–88.

Web Appendix

for

Budgeting Increases Reliance on Category-Level Evaluations

Table of Contents

WEB APPENDIX A: SURVEY OF BUDGETING EXPERIENCE

A representative survey by Zhang et al. (2022) suggests consumer budgeting is prevalent. Approximately two out of every three Americans currently use budgets; of those who do not currently budget, 42% have budgeted in the past. Of those who currently budget, 59% use formal budgets. Budgeting is common across income and wealth levels. Consumers typically organize budgets according to categories of spending: The most common labels consumers spontaneously report for their budgets include necessities like rent, mortgage, and insurance, as well as discretionary purchases like dining and entertainment.

We motivate the consumer relevance of our work with a survey of consumers' own budgeting experiences, drawing from and building on Zhang et al. (2022). Using both open-ended and closed-ended survey items, we assess the motivations and strategies for setting, tracking, and following budgets. In particular, we consider how consumers budget for discretionary spending categories. Whereas budgets for necessities are often fixed at specific payment amounts (e.g., rent, recurring bills, debt repayment), budgets for discretionary purchases are more likely to be set based on consumer preferences, in which case the value of budget categories may play an important role.

Method

We surveyed 200 participants from a gender-balanced sample on Amazon Mechanical Turk (AMT) using CloudResearch's approved sample pool (Litman, Robinson and Abberbock 2017). The survey consisted of 4 open-ended questions and 11 closed-ended budgeting questions, many of which included follow-up components. We adapted the basic structure of

Zhang et al. (2022), which began by establishing the participant's personal experience with budgeting. Following their approach, we dropped from our analysis all observations from participants who indicated no current or prior budgeting experience. Though not analyzed, these participants progressed through the survey by imagining the budget they would keep if they were to start budgeting. The complete survey materials including summary statistics (for the analyzed group of participants with current or prior budgeting experience) are available in our ResearchBox.

Key Findings

Budgets are relevant. The first measure (adopted from Zhang et al. 2022) identifies personal budgeting experience. Overwhelmingly, consumers report using budgets to guide their finances. 72% report currently budgeting, 14% report having previously budgeted, and only 14% report never budgeting (Q1). These percentages are comparable to those from the nationally representative sample used by Zhang et al. (2022), who observe rates of 66%, 15%, and 20% for current budgeters, previous budgeters, and never-budgeters, respectively. In our data, those who have budgeted and those who have never budgeted do not differ in gender, age, educational attainment, or income bracket (ps > .27). Following the approach of Zhang et al., we consider only the responses of the 86% of participants who currently or previously budgeted. All subsequent figures use this 86% of respondents with budgeting experience as the denominator, unless specified otherwise.

The widespread use of budgets in our sample reflects a variety of different financial motivations and goals. When asked why they budget (Q5), some participants used budgets to

overcome challenges of self-control (e.g., "*I like to make sure I don't do anything crazy or develop bad spending habits*"; "*I need to budget or I will end up overspending.*") Others were motivated by simplicity (e.g., "*I don't want to worry about money. I want to set aside money into each pool and then spend whatever I have left and not worry about retirement or debt or anything.*") Some articulated goals for spreading consumption across categories (e.g., "*I budget money so that I know how much money I have and I can allocate it to different needs. I can also save money for specific things instead of just having one large lump sum*"*;* "*I budget my money because I like to do things like go to movies, buy clothes, and go to restaurants; but if I don't budget towards these things, then I'll end up spending way too much on these non-necessities, then not have enough towards my bills.*") Regardless of the motivations and goals for budgeting, the act of budgeting creates a categorical structure for evaluating potential expenditures, which may in turn impact consumption.

Budgets are clearly defined and frequently checked. The majority of budgeters (67%) use some type of formal budgets to record and update transactions, compared to the 33% who rely solely on informal budgets (i.e., mental accounts; Thaler 1980, 1985, 1999) to keep track of finances (Q2a). Among those practicing formal budgeting, the most common approaches were pen-and-paper budgeting (37% of respondents) and computer spreadsheets (33% of respondents), followed by budgeting apps (12%) and websites (5%) (Q2b). Consumers regularly monitor their formal budgets, with 57% of those with formal budgets checking at least every few days and 89% checking at least every week (Q3). This high frequency of checking is not random, but rather reflects consulting budgets prior to spending. In our survey, 98% of consumers prefer to consult their budgets prior to making a purchase, compared to only 2% who prefer to check

after making a purchase (Q15). Taken together, these observations suggest budgets are clearly defined, regularly checked, and checking a budget is a precursor to spending. The implication is that budgets will guide spending.

Budgets are consequential for spending across categories. Budgets impact consumption when they are followed. When asked about the importance of following one's budget (1 = "Not very important"; 7 = "Very important"), the modal response was the maximum of 7 (M = 6.18, SD 0.93) (Q14).

To explore the importance of distinct budget categories, we modified a question about the main reasons for budgeting (Zhang et al. 2022: question 5). As an additional potential reason, we added: "to make sure I know how much is available to spend in different categories" (Q6). Critically, a majority (58%) of respondents indicated this as one of the main reasons for budgeting. In fact, of the 10 possible reasons, only two had higher response rates (table A1). We take this as evidence that the multicategorical nature of budgets is an important and appealing aspect of using budgets. In other words: Many consumers are drawn to budgets precisely to guide allocation across different categories. And these allocations are followed. When imagining unexpected budget deviations, 82% of respondents indicated they would rather reduce spending within the overspent category than rebalance their allocations across budgets (Q13). Budgets are sticky and have a direct consequence for how people spend.

TABLE A1

MAIN REASONS FOR BUDGETING

	Response	Frequency	Proportion
1	To make sure I don't spend more than my income	139	0.81
2	**To make sure I know how much is available to spend in different categories**	99	0.58
3	To save for long-term goals	110	0.64
4	To save for short-term goals	85	0.50
5	To avoid debt from predictable overspending	83	0.49
6	To avoid debt from unforeseen expenses	86	0.50
7	To make sure that I can provide for my family	83	0.49
8	To get myself out of debt	51	0.30
9	Other	7	0.04
10	I don't think it is important to budget	1	0.01

NOTE—The "main reasons for budgeting", modified from Zhang et al. (2022): question 5. We introduce and are interested in the response in the second line: "to make sure I know how much is available to spend in different categories". The asterisks are added here for visual emphasis but were not included in the stimuli presented to participants.

These responses suggest consumers use budgets to guide their spending across different categories. How are budget categories structured? Adopting a question from Zhang et al. (2022), about 10% of consumers prefer extremely coarse categories (i.e., "necessities, discretionary") and 20% prefer extremely granular categories (i.e., "rent, utilities, cell phone, internet, car, groceries, dining out, movies, travel, clothing, exercise, healthcare, other"). The remainder of consumers fall somewhere between these two extremes (Q8). While there is substantial heterogeneity across the various levels of responses, the majority (62%) budget at a level of detail that is sufficiently granular to separate Dining Out vs. Entertainment (rows 4-6 of table A2).

Additionally, we asked participants to list their own budget categories using an open-ended format (Q7). The most common self-generated category labels include "*food*", "*rent*", "*utilities*", "*groceries*", "*insurance*", "*gas*", "*car*", and "*entertainment*". Many of these labels refer to fixed-expense categories. Because the budgets for fixed expenses should have little to no variation in allocation or spending over the short-run, we focus on discretionary spending

categories. Therefore, in the following studies, we will consider "Dining" (a label encompassing discretionary elements of "food" and "groceries") and "Entertainment".

TABLE A2

GRANULARITY OF BUDGETS

	Response	Frequency	Proportion	Zhang et al.
1	Necessities, Discretionary	17	0.10	0.12
2	Housing & Transportation, Food, Discretionary, Other	25	0.15	0.18
3	Housing & Transportation, Food, Entertainment, Clothing, Other	23	0.13	0.13
4	Housing, Car, Groceries, Dining Out, Entertainment, Clothing, Other	28	0.16	0.16
5	Rent, Utilities, Cell phone, Groceries, Dining Out, Movies, etc...	43	0.25	0.18
6	Rent, Utilities, Cell phone, Internet, Car, Groceries, etc...	35	0.20	0.23

NOTE—This question adopted from Zhang et al. (2022): question 20. Responses in rows 5 and 6 were truncated in this table for presentation purposes. Participants saw the full granular lists, which are also available in our ResearchBox. We note 62% of respondents use budgets sufficiently granular to distinguish between discretionary categories such as Dining Out and Entertainment. The Zhang et al. proportions are presented in the last column.

Discussion

The survey of budgeting experience explored whether consumers budget, why they do so, and how they set, track, and follow their allocations. The key findings suggest budgets are relevant for most consumers, who formally track and follow their budget allocations. The majority of consumers indicate they use budgets in order to guide their spending across different budget categories. While each consumer uses their own category structure, most have sufficiently granular categories to separate dining and entertainment. Therefore, our paper considers dining and entertainment as two common discretionary budget categories.
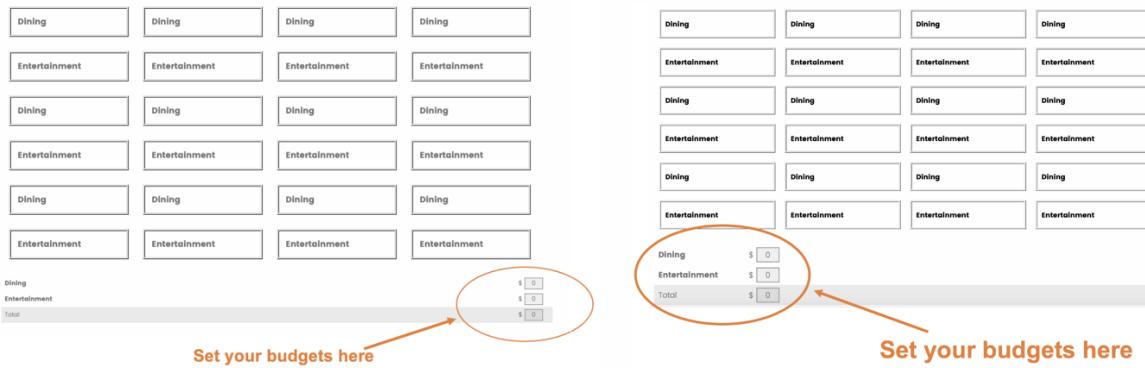
WEB APPENDIX B: SUPPLEMENT TO EXPERIMENT 2

Changes between studies

Version 1 to version 2. From version 1 to version 2, we made the following changes. (1) The constant sum box used to set budgets or show cumulative purchases was moved from right-aligned to left-indented (figure B1). (2) The box widths (and therefore column widths) were reduced slightly to reduce the likelihood participants engaged in horizontal scrolling. (3) To accommodate the width reduction, the option previously labelled "Drinks with friends" was shortened to "Happy hour". (4) To accommodate the width reduction, version 2 displayed items without costs (e.g., "Ice cream") instead of with costs (e.g., "Ice cream $10"). (5) When participants saw the lists of category options to rate (in terms of average category value), they saw the same text that was presented within the mouse-tracked grid. Therefore, in version 1, this included the "$10" reminder with each item; in version 2, it did not. (6) To increase the likelihood participants understood each item cost $10 without including this with each of the 24 options, we modified the instructions and included two additional reminders that all items cost $10. We intended these changes to be surface-level features that would not affect participants' experience or decision. The sole motivation for these changes was to reduce the likelihood of horizontal scrolling (in case participants on some browsers or screen zoom settings could not see the constant sum boxes).

FIGURE B1

COMPARISON OF MOUSE-TRACKED GRIDS IN VERSION 1 AND VERSION 2

Note—Two instruction images depicting the layout of the mouse-tracking grids used in versions 1 and 2. Version 1 (left) had slightly wider columns and the constant sum boxes were right-aligned. Version 2 (right) had slightly narrower columns and the constant sum boxes were left-indented, just beyond the budget category labels.

Version 2 to version 3. We made only one change between versions 2 and 3. We did not change anything in the instructions or the mouse-tracked grid. The sole change was that we incorporated the $10 cost reminder into the lists of options what participants rated category values, thus aligning version 3 with the original presentation in version 1. This is depicted by figure B2.

List of options. We used options very similar to those in experiment 1. These were mostly truncated in length in order to fit within the narrower mouseover boxes. The 12 dining items were: Lunch takeout, Candy bars, Protein shake, A flavored latte, Happy hour, Fast food, Bottle of wine, Appetizers, A meal kit, Ice cream, Morning donuts, and a Veggie snack. The 12 entertainment options were: Movie ticket, Amusement park, Concert, Art museum, History tour, Mini golf, State park, Camping, Meditation class, Ceramics class, Phone game, and a Streaming app.

FIGURE B2

CATEGORY RATING PRESENTATIONS IN VERSION 2 AND VERSION 3

On average, how much value do you get out of this category of purchases?	On average, how much value do you get out of this category of purchases?
Morning donuts	Ice cream ($10)
Ice cream	Candy bars ($10)
Happy hour	Fast food ($10)
Bottle of wine	Veggie snack ($10)
Veggie snack	Bottle of wine ($10)
Protein shake	A flavored latte ($10)
A meal kit	A meal kit ($10)
Appetizers	Lunch takeout ($10)
Lunch takeout	Appetizers ($10)
Candy bars	Happy hour ($10)
Fast food	Protein shake ($10)
A flavored latte	Morning donuts ($10)
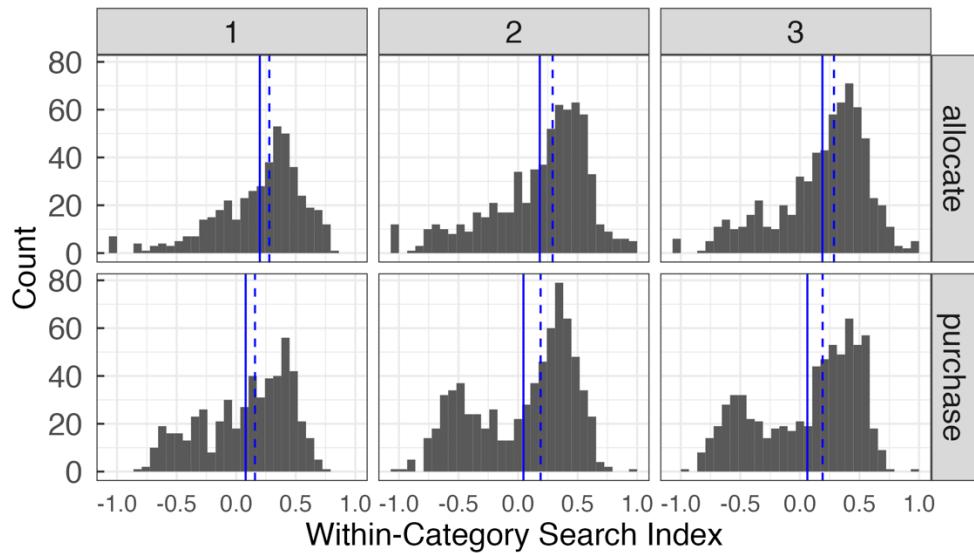
Note—Version 2 (left) did not include the price reminder when assessing the average value of each category. Version 1 and version 3 (right) did include the price reminder.

Results

Within-category search index as a proxy for evaluation mode. Allocating (vs. purchasing) had a consistent effect on the within-category search index across all three study versions. Participants in the allocate condition consistently engaged in more category search, as depicted by figure B3. In each study version, this effect of condition was statistically significant ($ps < .001$; table B1).

FIGURE B3

WITHIN-CATEGORY SEARCH INDEX, BY CONDITION AND EXPERIMENT VERSION

NOTE—Histograms of within-category search index, by condition (rows) and experiment version (columns). Solid blue line represents conditional mean, and dashed blue line represents conditional median.

TABLE B1

SEARCH INDEX REGRESSIONS BY EXPERIMENTAL VERSION

	(1)	(2)	(3)
allocate	0.059***	0.068***	0.063***
	p = 0.00000	p = 0.000	p = 0.00000
Constant	0.139***	0.114***	0.126***
	p = 0.000	p = 0.000	p = 0.000
Observations	971	1,365	1,346
R^2	0.025	0.026	0.023
Adjusted R^2	0.024	0.026	0.023
Residual Std. Error	0.368 (df = 969)	0.417 (df = 1363)	0.407 (df = 1344)

Note: $+$ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

NOTE—Columns correspond to experimental versions 1, 2, and 3, respectively. For all models, the dependent variable is the within-category search index.

Search index-by-dining preference interaction. There is some variability in estimates across the three study versions, as presented in table B2. While the search-index-by-dining preference interaction is significantly positive in versions 1 and 2, it is not significantly different from zero in version 3.

TABLE B2

DINING SHARE REGRESSIONS BY EXPERIMENTAL VERSION

	(1)	(2)	(3)
Search index	−4.759	3.569	−5.729
	p = 0.419	p = 0.379	p = 0.161
Dining preference	3.710***	3.833***	3.211***
	p = 0.000	p = 0.000	p = 0.000
Aggregate liking	−0.682**	−0.448**	−0.358*
	p = 0.002	p = 0.010	p = 0.039
Search index x dining preference	1.397*	1.225**	0.217
	p = 0.027	p = 0.003	p = 0.613
Search index x aggregate liking	0.525	−0.229	0.710+
	p = 0.356	p = 0.567	p = 0.077
Constant	63.299***	60.185***	60.408***
	p = 0.000	p = 0.000	p = 0.000
Observations	971	1,365	1,346
R^2	0.236	0.285	0.193
Adjusted R^2	0.232	0.282	0.190
Residual Std. Error	14.492 (df = 965)	13.521 (df = 1359)	14.280 (df = 1340)

Note: $+$ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

NOTE—Columns correspond to experimental versions 1, 2, and 3, respectively. For all models, the dependent variable is the dining share.

Conditional indirect effects. We consider a separate Process 17 model for each experimental version. We are interested in the conditional indirect effects (the extent to which the mediation pathway of allocation on dining share through the search index depends on the

relative dining preference). For a recap of the conceptual moderated mediation, please see manuscript figure 4. The focal test of this conditional indirect effect is the index of moderated mediation and its bootstrapped 95% confidence interval. These are presented for all three experimental versions in table B3. Versions 1 and 2 are generally similar, as both have 95% confidence intervals that barely include or do not include zero.

TABLE B3

INDEX OF MODERATED MEDIATION, BY EXPERIMENT VERSION

Version 1	Version 2	Version 3
Index of moderated mediation: 0.066 $CI_{95\%}$ = [-0.015, 0.162]	*Index of moderated mediation*: 0.095 $CI_{95\%}$ = [0.030, 0.169]	*Index of moderated mediation*: 0.007 $CI_{95\%}$ = [-0.053, 0.066]

NOTE—Columns correspond to experiment versions 1, 2, and 3, respectively.

Model without mouse-tracked search index. In experiment 1 from the main manuscript, we found a significantly positive allocate-by-dining preference interaction. This is consistent with H2a, which predicts allocators (vs. purchasers) will be more sensitive to a category's average value when deciding upon the dining share of consumption. However, in experiment 2 from the main manuscript, we observe a directional but statistically insignificant interaction. Table B4 presents the individual regressions for each version of experiment 2. There is substantial variation across the estimates of the focal allocate-by-dining preference interaction.

TABLE B4

DINING SHARE REGRESSIONS BY EXPERIMENT VERSION

	(1)	(2)	(3)
Allocate	0.039	4.182*	0.233
	p = 0.986	p = 0.016	p = 0.893
Dining preference	3.825***	4.002***	3.191***
	p = 0.000	p = 0.000	p = 0.000
Aggregate liking	−0.598**	−0.554***	−0.305^{+}
	p = 0.004	p = 0.001	p = 0.071
Allocate x dining preference	0.561*	−0.239	0.219
	p = 0.015	p = 0.170	p = 0.230
Allocate x aggregate liking	−0.097	−0.492**	−0.076
	p = 0.634	p = 0.004	p = 0.652
Constant	62.456***	61.439***	60.000***
	p = 0.000	p = 0.000	p = 0.000
Observations	972	1,369	1,352
R^2	0.242	0.287	0.193
Adjusted R^2	0.238	0.284	0.190
Residual Std. Error	14.442 (df = 966)	13.491 (df = 1363)	14.288 (df = 1346)

Note:	+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

NOTE—Columns correspond to experiment versions 1, 2, and 3, respectively. For all models, the dependent variable is the dining share.

WEB APPENDIX C: SUPPLEMENTAL STUDY ON AVERAGES VS. MARGINS

Did the results from experiments 1 and 2 reflect a sensitivity to a form of value other than *average value*? In particular, is it possible that participants—especially those in the allocate condition—were responding to perceived value at the margin? This supplemental study attempts to rule out that alternative by providing an initial test of the unique role of average value—beyond marginal value—in allocation decisions. We examine this relationship among consumers' own preferences and valuations for dining and entertainment purchases. Specifically, we ask participants to consider the average and marginal value of their own dining and entertainment expenditures. This allows us to test whether a hypothetical budget allocation is influenced only by the value-maximizing purchase (i.e., the marginal good), or whether allocations also depend on the average value of a category. We predict average value of a category will drive a hypothetical budget allocation decision above and beyond the value of the marginal good. Of course, the category's average value and the value of the marginal good may covary. However, if there is a remaining relationship between category average and allocation *after controlling for value at the margin*, then this suggests that consumers are indeed sensitive to average value.

Method

Participants. 606 participants (M_{age} = 41; 49% female) from AMT completed this study[8].

---

[8] We began with 608 complete observations; however, two sets of observations were linked to participants who had incomplete surveys prior to submitting their final survey. We removed these observations, resulting in the final sample size of 606.

Design and stimuli. Participants responded to a series of questions about their own

consumption. These questions focused on discretionary dining and entertainment categories. We

used these categories because they are commonly used by participants in our survey of budgeting

experience (web appendix A) and among consumer researchers (e.g., Cheema and Soman 2006;

Heath and Soll 1996; Sussman and Alter 2012). Discretionary dining purchases were defined as

*"*everything from snacks at a convenience store, to drinks at a coffee shop, to meals at a nice

restaurant. This category should not include necessities, like groceries for the home."

Entertainment purchases were defined as "everything from buying new games to play or

streaming services to watch, to tickets for events, activities, and travel."

Participants first indicated how they would allocate an additional $100 of discretionary

money between their dining and entertainment budgets. Allocations could be any non-negative

amount that summed to $100. Next, participants listed three of their own purchases that were

typical of their dining budget (e.g., "*Healthy snacks, smoothies, and veggie wraps*"; "*Coffee

shops, restaurants, food delivery*"), followed by the purchase they would make if they had an

extra $50 in this budget (e.g., "*I'd go out to a nice sushi dinner with the wife", "I would order a

$50 bottle of wine at dinner.*") This additional $50 purchase is the marginal category purchase: It

is the next-best option that would be consumed with additional funds. This is necessarily a proxy

as the amount allocated to a particular budget was not always precisely $50. The same purchase-

listing process was repeated for the entertainment category.

After identifying typical and marginal purchases, participants then saw and rated their

prior responses on 7-point value scales anchored on 1 = "Not very much" and 7 = "A lot". They

first rated the average category value by considering their own list of typical purchases and

responding to the question: "on average, how much value do you get out of this category of

purchases?" After providing an average category value for both dining and entertainment, participants then rated the value of the marginal purchase in each category as "how much value would you get out of this purchase?" This set of four measures corresponds to average value and the value of the marginal good for both dining and entertainment.

Results

Analysis plan. The amount of money allocated to the dining budget served as our dependent variable. We include the four value measures as predictors[9]. Specifically, these measures are the average value of the dining category, the average value of the entertainment category, the value of the marginal dining purchase, and the value of the marginal entertainment purchase. We expect more allocation to dining when the average dining value is higher and less allocation to dining when the average entertainment value is higher, while controlling for the marginal value of each category.

Analysis. On average, participants allocated \$53.17 of their imagined \$100 to the dining budget (SD = \$21.36). The means of the average and marginal value ratings differed somewhat for dining (M_{avg} = 5.40, SD = 1.38; M_{margin} = 5.59, SD = 1.38; $t(605)$ = 3.51, p < .001) but not for entertainment (M_{avg} = 5.68, SD = 1.33; M_{margin} = 5.76, SD = 1.31; t(605) = 1.59, p = .112).

For our key analysis, we regressed the dining allocation on the average value of dining, the average value of entertainment, the marginal value of dining, and the marginal value of entertainment. As expected, we observed a positive coefficient for the average value of dining (*b*

---

[9] We present this model for ease of explication. We preregistered an equivalent model using sum and difference terms. Both models are mathematically equivalent and lead to the same statistical and substantive inferences.

= 4.27, se = 0.72, $t(601)$ = 5.89, p < .001) and a negative coefficient for the average value of entertainment (b = -4.47, se = .77, $t(601)$ = -5.80, p < .001). The controls for marginal value exhibited the same pattern of directional signs, though neither were significant at conventional levels (b_{dining} = 1.25, se = 0.72, $t(601)$ = 1.75, p = .081; $b_{entertainment}$ = -0.68, se = .79, $t(601)$ = -0.86, p = .392).

Discussion

This supplementary study asks participants to draw upon their own consumer experience to assess average and marginal value for commonly purchased or considered items. We analyze the independent role of a budget category's average value in a hypothetical allocation decision, while controlling for value at the margin. The results suggest budget allocations are indeed sensitive to the unique role of average value. A key limitation is that measurement error (e.g., uncertainty regarding what the marginal purchase would be) may tend to decrease the coefficient on marginal value and increase the coefficient on average value.
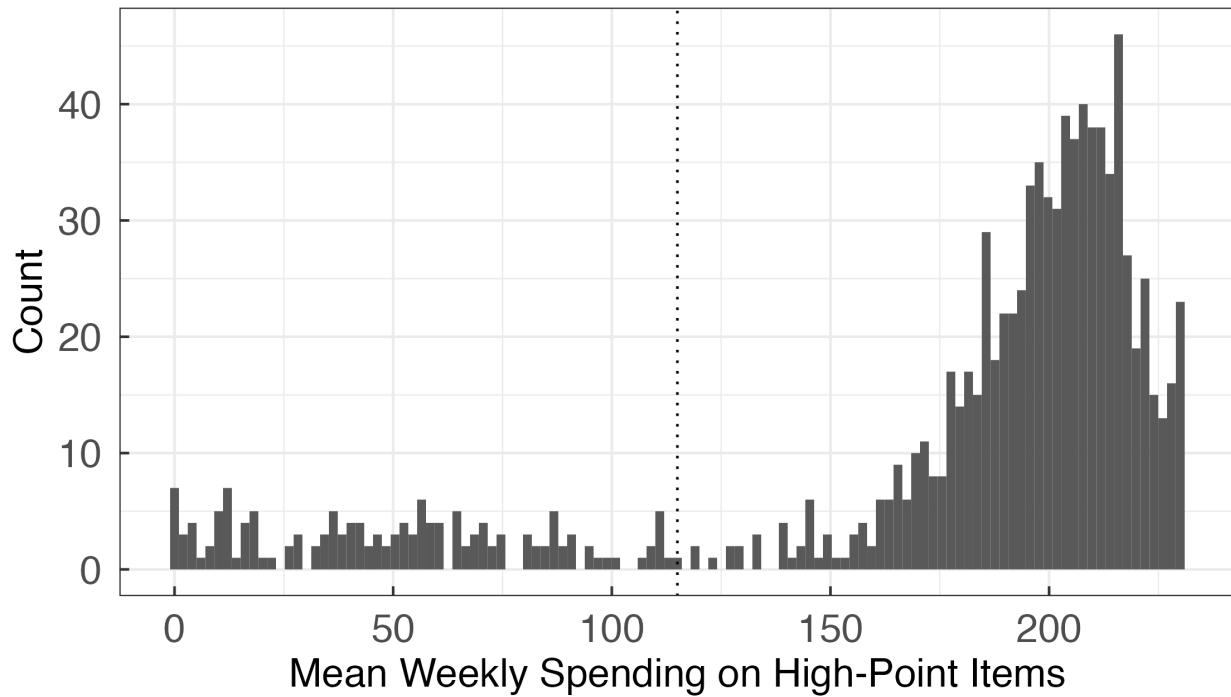
Exclusions and Comprehension Checks

Exclusions. Figure D1 depicts the distribution of spending on high-point purchases used for exclusion in experiment 3. Participants who bought less than half of the available high-point purchases (those to the left of the dashed line) were preregistered to be excluded, as these participants were likely inattentive or misunderstood the game.

FIGURE D1

DISTRIBUTION OF MEAN WEEKLY HIGH-POINT SPENDING, USED FOR EXCLUSION

NOTE— Distribution of spending on high-point items in experiment 3. The mass of the data lay well above 50% (dotted line). Purchasing less than 50% of these high-point items (participants to the left of the dotted line) is outside the range of typical behavior and is taken to indicate inattention or misunderstanding of the task.

Comprehension. Among the included participants, performance on comprehension questions was quite good. Correct response rates were 92%, 97%, 81%, 98%, and 87%, corresponding to the five sequential questions (below). After answering each question, participants were provided feedback about whether their response was correct or incorrect and given an explanation.

- *Q1*: "The goal of this game is to collect as many _____ as possible": 92% correctly identified "points" from a list of three options.

- *Q2a*: "Which of the following is true about the budgets for dining and entertainment purchases?": 97% correctly identified "Budgets may help plan my purchases, though I am not required to follow them" from a list of two options. (This question was only asked to participants in the budget condition.)

Questions Q2b-Q4 were True or False.
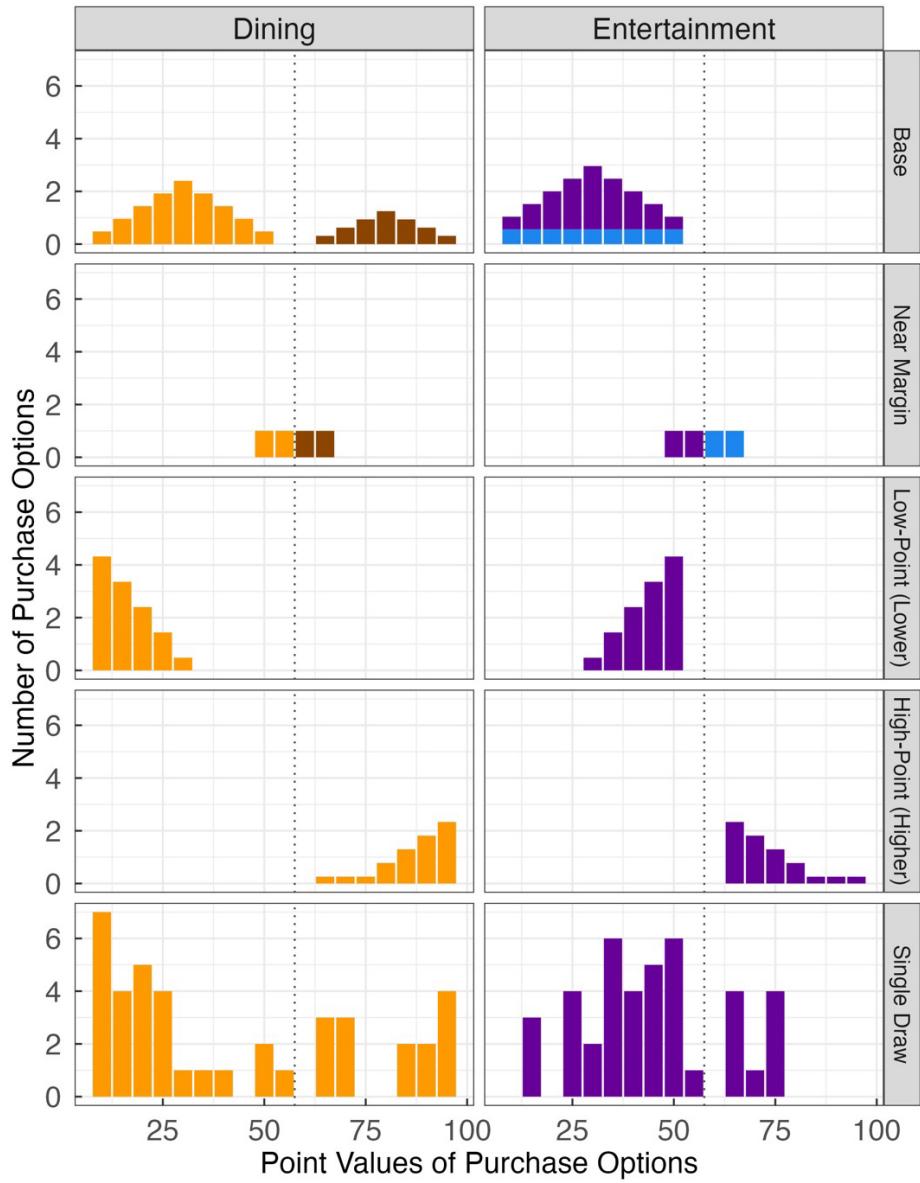
- *Q2b:* "During a given week, you may make 23 or fewer purchases.": 81% correctly identified "True".

- *Q3:* "Any unspent money will carry over to the following week": 98% correctly answered "False".

- *Q4:* "You will have the opportunity to earn a bonus during both the five-week practice round and the five-week game.": 87% correctly answered "False".

Construction of Distributions

Figure D2 depicts the theoretical distributions from which items were drawn in experiment 3. The first row indicates the common portions of the dining and entertainment distributions used in all conditions. The second row indicates the distribution for the two best options from the low-point region of the distribution and the two worst options from the high-point region of the distribution. The third row depicts the low-point region of the distribution when dining is low and entertainment is high; these distributions would be swapped in the condition where dining is high and entertainment is low. The fourth row depicts the high-point region of the distribution when dining is high and entertainment is low; these distributions would be swapped in the other high-value condition. By drawing items from these distributions, there were always exactly 14 dining options worth at least 60 points and there were always exactly 9 entertainment options worth at least 60 points, but the category average values systematically varied by condition. Finally, the bottom row depicts a sample draw from the theoretical distributions in the prior four rows. This single draw is typical of what a participant in that condition may have been presented with.

FIGURE D2
THEORETICAL POINT DISTRIBUTIONS IN EXPERIMENT 3

NOTE—Distributions from which items were drawn in study 3. The dotted vertical line represents the split between the low-point region of the distribution (point values less than 60) and the high-point region of the distribution (point values of 60 or more). Row 1 ensured that possible points did not systematically differ across conditions. 5 dining options were drawn from the brown distribution; 12 dining items were drawn from the orange distribution; 12 entertainment items were drawn from the purple distribution; 5 entertainment options were drawn from the blue distribution. Row 2 ensured deviations of up to 2 items from the value-maximizing bundle would lead to symmetric outcomes. 2 items were drawn from each of the orange, brown, purple, and blue distributions. These options were the best low-value options (either 50 or 55) and the worst high-value options (either 60 or 65) available. Rows 3 and 4 depict the manipulation of the low-value and high-value parts of the distributions. Row 5 depicts a single sample draw a participant may have seen.

Additional Results

For completeness, we provide the full regression output for both preregistered analyses, as well as the exploratory analysis considering H2a, for both dining share and dining difference.

First analysis: sensitivity to average among allocators. We present the regression output for the full preregistered model in table D1. As discussed in the main manuscript, this model regressed the dining share of allocation on both the dining average and the low-point dining average. We include an analysis over the full data (without exclusions) for completeness, as well as using the preregistered difference measure in columns 3 and 4.

TABLE D1
COMPLETE REGRESSION OUTPUT: FIRST ANALYSIS

	Dependent variable:			
	Dining Share		Difference Measure	
	With Exclusions	Without Exclusions	With Exclusions	Without Exclusions
	(1)	(2)	(3)	(4)
Dining average	2.606***	2.055***	11.990***	9.454***
	p = 0.00000	p = 0.0001	p = 0.00000	p = 0.0001
Low-point dining average	0.832+	1.286*	3.829+	5.916*
	p = 0.089	p = 0.012	p = 0.089	p = 0.012
Din. avg. x low din. avg.	0.065	0.012	0.299	0.057
	p = 0.895	p = 0.981	p = 0.895	p = 0.981
Constant	54.983***	55.294***	22.922***	24.354***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	394	478	394	478
R^2	0.075	0.046	0.075	0.046
Adjusted R^2	0.068	0.040	0.068	0.040
Residual Std. Error	9.675 (df = 390)	11.114 (df = 474)	44.503 (df = 390)	51.124 (df = 474)

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001

NOTE—Dependent variable is the dining share of allocation (cols 1-2) and the difference in dining (3-4). Dining average refers to the manipulation of average category values in the high-point region of the distribution (+1 = dining high, -1 = dining low). Low-point dining average refers to the manipulation of average category values in the low-point region of the distribution (+1 = dining high, -1 = dining low). Columns 1 and 3 apply the preregistered exclusions and columns 2 and 4 considers the full data, without exclusions.

Second analysis: budget-by-dining average interaction. The full regression output for our second preregistered analysis is provided by table D2. The dependent variable is the dining share of spending, and the focal preregistered variable is the dining average-by-budget interaction. This analysis does not find any other significant two-way or three-way interactions. Columns 3 and 4 include the difference measure.

TABLE D2
COMPLETE REGRESSION OUTPUT: SECOND ANALYSIS

	Dependent variable:			
	Dining Share		Difference Measure	
	With Exclusions	Without Exclusions	With Exclusions	Without Exclusions
	(1)	(2)	(3)	(4)
Dining average	1.890***	1.783***	8.416***	7.664***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Low-point dining average	1.540***	0.999***	6.943***	5.124***
	p = 0.000	p = 0.0003	p = 0.000	p = 0.00000
Budget	−1.368***	−0.878**	−6.327***	−4.541***
	p = 0.000	p = 0.002	p = 0.000	p = 0.00001
Din. avg. x low din. avg.	−0.068	−0.118	−0.241	−0.815
	p = 0.726	p = 0.661	p = 0.781	p = 0.418
Din. avg. x budget	0.542**	0.449+	2.447**	2.047*
	p = 0.006	p = 0.097	p = 0.005	p = 0.043
Low din. avg. x budget	−0.299	−0.050	−1.326	−0.212
	p = 0.126	p = 0.853	p = 0.126	p = 0.833
Din. x low x budget	−0.114	−0.263	−0.471	−1.015
	p = 0.560	p = 0.331	p = 0.586	p = 0.313
Constant	58.593***	58.150***	38.840***	35.043***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	821	970	821	970
R^2	0.206	0.069	0.212	0.103
Adjusted R^2	0.200	0.063	0.205	0.096
Residual Std. Error	5.579 (df = 813)	8.395 (df = 962)	24.737 (df = 813)	31.308 (df = 962)

Note: + p<0.1; * p<0.05; ** p<0.01; *** p<0.001

NOTE— Dependent variable is the dining share of allocation (cols 1-2) and the difference in dining (3-4). Dining average refers to the manipulation of average category values in the high-point region of the distribution (+1 = dining high, -1 = dining low). Low-point dining average refers to the manipulation of average category values in the low-point region of the distribution (+1 = dining high, -1 = dining low). Budget refers to whether or not participants set a budget allocation prior to spending (+1 = yes, -1 = no). Columns 1 and 3 apply the preregistered exclusions and columns 2 and 4 considers the full data, without exclusions.

Additional analysis: comparing allocations to spending. Though not preregistered, we can conceptually replicate the analyses of experiments 1 and 2, in which the dining share reflects [dining dollars / total allocated dollars] x 100% for budgeters and [dining dollars / total spent dollars] x 100% for non-budgeters. We regress this dining share on the same predictors from the prior model and present the results in table D3. The variable of interest is the dining average-by-budget interaction. Columns 3 and 4 include the difference measure.

TABLE D3
COMPLETE REGRESSION OUTPUT: EXPLORATORY ANALYSIS

	Dependent variable:			
	Dining Share		Difference Measure	
	With Exclusions	Without Exclusions	With Exclusions	Without Exclusions
	(1)	(2)	(3)	(4)
Dining average	1.977***	1.695***	8.979***	7.536***
	p = 0.000	p = 0.00000	p = 0.000	p = 0.000
Low-point dining average	1.336***	1.167***	6.049***	5.626***
	p = 0.00000	p = 0.0002	p = 0.00000	p = 0.00002
Budget	−2.489***	−1.867***	−11.122***	−7.615***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Din. avg. x low din. avg.	0.055	0.078	0.264	0.129
	p = 0.823	p = 0.799	p = 0.815	p = 0.921
Din. avg. x budget	0.629*	0.361	3.010**	1.919
	p = 0.011	p = 0.242	p = 0.008	p = 0.139
Low din. avg. x budget	−0.503*	0.119	−2.220*	0.290
	p = 0.042	p = 0.700	p = 0.050	p = 0.823
Din. x low x budget	0.010	−0.066	0.034	−0.072
	p = 0.969	p = 0.831	p = 0.976	p = 0.956
Constant	57.472***	57.161***	34.045***	31.969***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	821	970	821	970
R^2	0.200	0.080	0.194	0.085
Adjusted R^2	0.193	0.073	0.188	0.079
Residual Std. Error	7.039 (df = 813)	9.573 (df = 962)	32.312 (df = 813)	40.253 (df = 962)

Note: $+$ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

NOTE—Exploratory model to consider H2a. Dependent variable is the dining share (cols 1-2) or dining difference (3-4) of allocation in the budget condition and the dining share of spending in the no-budget condition. All predictors are the same as in the prior specification (i.e., as in table D2).

Budgets predict spending in experiment 3: exploratory analysis. As an exploratory analysis, we regress each individual option's purchase decision (80 per week, for each of 5 weeks) on category, money remaining in the entertainment budget and money remaining in the dining budget, their interactions with category, and a rich set of controls. We include participant fixed effects to account for the fact that some participants routinely underspend on dining. We include week and day fixed effects to account for time trends. We include item-value fixed effects to reduce error. And we include history controls (the number of 95-point dining items seen, the number of 95-point entertainment items seen, the number of 90-point dining items seen, etc.) to account for expectations regarding category-specific remaining items. $10 remaining in the dining budget is associated with a 3.2 percentage point increase in the likelihood of purchasing a dining item but a 2.0 percentage point increase in the likelihood of purchasing an entertainment item; $10 remaining in the entertainment budget is associated with a 2.2 percentage point increase in the likelihood of purchasing a dining item but a 3.2 percentage point increase in the likelihood of purchasing an entertainment item. The difference between the differential effect of dining budget remaining on dining vs. entertainment spending and the differential effect of entertainment budget remaining on dining vs. entertainment spending is statistically significant ($t(65) = 7.78, p < .001$), indicating that funds are not treated as perfectly fungible.

WEB APPENDIX E: ANALYSIS OF ATTRITION

For all three experiments, we report the contingency tables of complete and incomplete responses. We do not observe concerning evidence of differential attrition in experiments 1 or 3. We do observe differential attrition in experiment 2, though this may be partially explained by a mechanical feature of the JavaScript code, as discussed.

Experiment 1

Nearly all participants completed the experiment (99% completion rate). We did not find different rates of attrition across the allocate and purchase categories ($\chi^2(1) = 0.86$, $p = .354$)[10]. The number of complete and incomplete responses is presented in table E1.

TABLE E1
COMPLETE AND INCOMPLETE RESPONSES IN EXPERIMENT 2

	Allocate	Purchase
Complete	399	402
Incomplete	3	7

Experiment 2

As indicated in the main manuscript, there was differential attrition across conditions in experiment 2. Examining the pooled data, about 13% of observations were incomplete in the

---

[10] This applies a Yates' correction to partially address the low frequencies of incomplete responses.

allocate condition, compared to only 5% in the purchase condition ($\chi^2(1) = 84.82$, $p < .001$). We

suspect this may be due in part to a mechanical issue with the JavaScript coding that would have

made it impossible for participants in the allocate (but not purchase) condition to advance

beyond the 6x4 grid page. However, we acknowledge this explanation cannot fully account for

the entire differential attrition between conditions. The contingency table of complete and

incomplete observations for each experimental version is given by table E2. All three versions

had imbalanced completion rates across conditions (separate $\chi^2(1)$ tests indicate all ps $< .001$).

TABLE E2
COMPLETE AND INCOMPLETE RESPONSES IN EXPERIMENT 2

	Version 1		**Version 2**		**Version 3**	
	Allocate	**Purchase**	**Allocate**	**Purchase**	**Allocate**	**Purchase**
Complete	468	535	673	731	679	724
Incomplete	103	38	93	36	87	35

NOTE—Responses separated by experimental version.

Experiment 3

We did not detect meaningfully different attrition rates across conditions in experiment 3.

We consider the entire contingency table across all 8 cells of the 2x2x2 design ($\chi^2(7) = 5.97$, $p =$

.543), as presented by table E3. More importantly, we also consider the completion rates for

budgeters and non-budgeters (collapsing across the dining average and low-point dining average

conditions). Despite potential concerns about differential attrition between budgeters and non-

budgeters—based on the differential attrition between allocators and purchasers in experiment

2—this does not appear to be a problem in experiment 3 ($\chi^2(1) = 1.30$, $p = .253$).

TABLE E3
COMPLETE AND INCOMPLETE RESPONSES IN EXPERIMENT 3

	DDN	DDB	DEN	DEB	EDN	EDB	EEN	EEB
Complete	122	119	125	122	127	116	118	121
Incomplete	12	19	9	10	13	14	15	17

NOTE—Responses separated by experimental condition. A leading "D" (Dxx) refers to the dining average being higher and "E" (Exx) refers to the entertainment average being higher. A middle "D" (xDx) refers to the dining average being higher in the low-point region, and "E" (xEx) refers to entertainment higher in the low-point region. The final "N" or "B" indicates whether a participant was in the No-Budget (xxN) or Budget (xxB) condition.

REFERENCES

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.

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

Litman, L., J. Robinson, and T. Abberbock (2017), "TurkPrime.Com: A Versatile Crowdsourcing Data Acquisition Platform for the Behavioral Sciences," *Behavioral Research Methods*, 49(2), 433–42.

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), 800–814.

Thaler, Richard H. (1980), "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization*, 1(1), 39–60.

——— (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4(3), 199–214.

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

Zhang, C. Yiwei, Abigail B. Sussman, Nathan Wang-Ly, and Jennifer Lyu (2022), "How Consumers Budget," *Journal of Economic Behavior & Organization*, 204, 69–88.