CHASING CONSUMER SURPLUS: AVERAGE VALUE ATTRACTS BUDGET ALLOCATIONS

David Dolifka and Stephen A. Spiller

UCLA Anderson School of Management

2/17/2022

David Dolifka (david.dolifka.phd@anderson.ucla.edu) is a PhD candidate in marketing and Stephen A. Spiller (stephen.spiller@anderson.ucla.edu) is an associate professor of marketing and behavioral decision making, both at UCLA Anderson School of Management. The authors thank members of the UCLA Anderson Behavioral Lab, seminar participants at Duke University, the University of Michigan, and The Ohio State University, and Justin Pomerance for valuable feedback.
Consumers frequently use budgets to manage their spending. Budgets are consequential, as money in budgets is treated as though it is not fungible. What drives consumer budget allocation decisions? To maximize value, budgets ought to be set such that marginal value is equated across budgets. Based on prior research on melioration and the consequences of non-linear price schedules, we propose consumers’ budget allocation decisions are driven by average value in addition to marginal value. This process implies consumers allocate more resources to budgets with higher average value, even at the expense of lower marginal value. The authors present evidence of this pattern in two preregistered experiments. This effect persists even when consumers repeatedly make incentivized budget allocation decisions. These findings reveal specific effects of purchase opportunities on consumer budgets and suggest ways decision aids might improve budgeting processes.

*Keywords*: budgeting; melioration; resource allocation; marginal value
Consumers regularly use budgets to manage their household finances. These budgets are consequential, as they affect what consumers buy and how much consumers spend. To maximize value from consumption, consumers should equate marginal values across budgets. If marginal values are not equated, total value can be increased by forgoing a lower-margin purchase to obtain a higher-margin purchase instead. Yet tracking marginal value is difficult and consumers often neglect or underweight opportunity costs (Thaler 1980; Frederick et al. 2009; Spiller 2011).

Given those difficulties, how do consumers allocate funds across their budgets? Drawing from research on melioration (Herrnstein and Prelec 1991), we propose that consumers are influenced by average consumption value across a set of uses, rather than just marginal consumption value, leading them to chase consumer surplus through their allocation of funds across budgets. In two experiments, we find that holding constant the marginal value, consumers allocate more funds to budgets with higher average values.

We begin by discussing prior research on consumer budgeting. We then discuss convergent findings from psychology and economics regarding people’s tendencies to be sensitive to average values and average prices, even when at the expense of marginal values and prices. After reviewing the prior literature, we present evidence of the core effect in an experiment using a budgeting task where all diagnostic information is available to consumers at the moment of allocation. We then present the results of a second experiment using an incentivized budgeting game. Results from this experiment provide consistent and robust evidence of sensitivity to average values, while ruling out possible alternative explanations regarding differential costs of making mistakes and various forms of uncertainty. We conclude with future directions and potential implications.
BUDGETING

Consumers Budget

We adopt Zhang and Sussman’s (2018) definition of a budget as “a financial plan by which individuals, companies, or institutions allocate present and future funds to various uses such as expenses, savings, investments, and debt repayment” and of the act of budgeting as “the process used to segregate and track the allocation and use of funds against different accounts with implicit or explicit limits.” These definitions highlight that budgeting involves both an allocation stage as well as a spending or usage stage. In this paper, our primary interest is in consumer decisions during the allocation stage. In our second experiment, we also examine follow-through on those decisions in the spending stage to consider welfare implications and to address alternative explanations.

A representative survey by Zhang et al. (2020) suggests consumer budgeting is prevalent. Approximately two out of every three Americans currently use budgets; of those who do not currently budget, 40% have budgeted in the past. 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 fixed expenses like rent, mortgage, and insurance, as well as varying expenses in categories for necessities and discretionary purchases like utilities, gas, food, and entertainment.

Budgets Are Consequential

Budgets have a causal impact on subsequent spending. 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). Budgets affect how consumers respond to price and income shocks (Hastings and Shapiro 2013, 2018) and can lead to underconsumption (Heath and Soll 1996). Using budgets can help consumers reduce consumption of goods they seek to limit due to self-control considerations (Krishnamurthy and Prokopec 2010). Ironically, budgets can also lead to increased spending by reducing focus on minimizing costs, conditional on successfully remaining under the budget limit (Larson and Hamilton 2012), or by changing the reference point while the role of the budget itself depreciates over time (Choe and Kan 2021). Depending on whether a more limited budget (e.g., a happy hour budget; a weekly budget) or more expansive budget (e.g., a food budget; a monthly budget) is more accessible, the perceived costliness of different expenditures will vary, thereby affecting consumption (Morewedge, Holtzman, and Epley 2007).

Allocating Funds

Given its spending implications, it is important to assess the key inputs into the budget allocation decision. One key input is predicted spending: when people believe they will spend more, they tend to allocate more money to that budget (Howard, Hardisty, and Sussman 2019; Peetz and Buehler 2009; Stilley, Inman, and Wakefield 2010a, 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.

Budget allocations are also often intertwined with self-control considerations. Budgets enable 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 tradeoffs they might otherwise avoid, thus reducing dysfunctional behavior (Fernbach, Kan, and Lynch 2015).

Beyond predicted spending and strategic 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 factors regarding how budget categories are grouped, consistent with the broader literature on partition dependence (Bardolet, Fox, and Lovallo 2011; Jia, Li, and Krishna 2020). For example, consumers may allocate more money to movies if they have two budgets devoted to movies and food (where food includes 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 spending predictions, self-control, and incidental factors, a key underlying assumption is that consumption value matters. That is, each of these literatures implicitly or explicitly acknowledge that the more valuable an expenditure category is, the greater the allocation that will be made to its associated budget. But how is a budget allocation determined from the value of its set of uses? Although there is a clear normative model, to our knowledge the question has not been empirically resolved. We seek to address it.
MARGINS VS. AVERAGES

Consider a consumer, Alex, with the opportunity to spend discretionary funds on food and entertainment. Alex quite likes her local restaurant’s Taco Tuesdays but does not enjoy dining out otherwise. Alex also enjoys going to the movies, but not nearly as much as she likes Taco Tuesdays. Given her available funds for discretionary spending, Alex can maximize utility by putting enough money in her monthly dining out budget for Taco Tuesday each week and enough money in her monthly entertainment budget for one movie each week. By doing so, Alex equates the marginal value of her consumption from dining out and from going to the movies. Moving money from food to entertainment would mean giving up tacos in exchange for an extra movie (which would reduce her total utility) and moving money from entertainment to food would mean giving up a movie in exchange for a non-taco meal out (which would also reduce her total utility).

Yet, prior work on opportunity cost neglect suggests Alex may not consider the marginal value (Thaler 1985; Frederick et al. 2009; Read, Olivola, and Hardisty, 2007; Magen, Dweck, and Gross 2008; cf. Spiller 2011). If she neglects the marginal value, what might she use instead? Below, we argue she may use the average value. If Alex is sensitive to the average value of her purchases, she may chase the consumer surplus she gets from Taco Tuesdays and allocate extra funds to her dining out budget. That is, because her average consumption value from Taco Tuesdays is greater than her average value from going to the movies, Alex might allocate more money to her dining out budget. This allocation would be a mistake, because those funds cannot be used on more Taco Tuesdays and instead could only be used on dining out that is less valuable than a movie.
In many contexts, average value and marginal value may tend to covary. Yet, it is important to understand the extent to which they independently contribute to budget allocation decisions. Below, we describe research suggesting why consumers may indeed follow average consumption value even at the expense of marginal value.

**Maximization vs. Melioration**

In situations of distributed choice (i.e., bundles of consumption decisions that result from numerous individual choices), people allocate choices in accordance with the matching law. That is, people choose alternatives in proportion to the benefits they receive from those choices, such that long-run average returns across alternatives are equated (Davison and McCarthy 1988; Herrnstein 1990; McDowell 2013; Rachlin and Laibson 1997).

This long-run behavioral tendency can result from two simple propositions (Herrnstein and Prelec 1991; Herrnstein et al. 1993). First, people track the average value per unit they obtain through consumption of a particular option (*value accounting*). Second, on a given consumption occasion, people choose the alternative with the higher average value per unit (*melioration*). Together, these principles lead to long-run behavior consistent with the matching law, as people choose the higher-value option until either it is no longer the higher-value option, at which point they reach an equilibrium, or else it is the only option chosen.

If we apply these principles to budgeting, they imply that budgets with higher average consumption value will attract funds, even if at the expense of marginal value. Considering Alex’s choices described above, value accounting will lead Alex to recognize greater per-unit consumption value from her dining out budget than from her entertainment budget. Melioration will lead her to allocate more funds to her dining out budget. As a result, Alex’s average consumption value from dining out will decrease, due to the lower value of a non-taco meal. This
process of melioration will lead to a convergence of average values, while creating a wedge (and utility loss) in marginal values.

Consistent with prior treatments of value accounting (Herrnstein and Prelec 1991), we use averaging in an inclusive sense to capture shifts in the value distribution that lead to differences in the attributed per-unit value. This may include shifts in the mode, extreme values, or most recent values. Holding all else constant, any such change will result in a change in the average; however, it is possible these shifts may not all have the same influence. As a proxy, we collectively refer to such a shift in the distribution as a shift in the average and leave further distinctions for future research.

*Use of Averages in Economic Behavior*

Though equating marginal value is a core tenet of the normative approach to allocating scarce resources, people and firms often act as though they rely upon average values rather than marginal values. Firms reliably fail to act in accordance with marginal analysis when setting prices (Faulhaber and Baumo 1988; Altomonte, Barattieri, and Basu 2015), and consumers vary in the extent to which they attend to marginal values when making decisions (Larrick, Nisbett, and Morgan, 1993; Larrick, Morgan, and Nisbett 1990). In individual choice opportunities, consumers are frequently insensitive to the degree to which marginal utility diminishes as quantities increase unless they are explicitly reminded (Li and Hsee 2021).

The matching law and melioration processes described above have been connected to the use of simplifying heuristics in the face of complex non-linear pricing policies in a number of different situations, sometimes collectively referred to as schmeduling (Liebman and Zeckhauser 2004). Such schmeduling is observed for judgments and decisions in the face of tax schedules (de Bartolome 1995; Rees-Jones and Taubinsky 2020), purchase decisions in the face of price
schedules (Gottfries and Hylton 1987; Ito 2014; Shin 1985) and credit card repayment decisions across multiple accounts (Gathergood et al. 2019). As two concrete examples, consumers’ energy consumption is sensitive to changes in average price when marginal price is held constant (Ito 2014; Shin 1985), and their tax expectations imply use of “ironing,” or considering their average tax rates rather than their marginal tax rates (Rees-Jones and Taubinsky 2020).

Consumers are sensitive to averages in a variety of choice contexts in which they should not be. 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 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 2021).

**Hypotheses and Experiment Overview**

In short, the sensitivity to average rewards as established in the context of research on the matching law and melioration has consequences in a variety of domains of economic behavior. Such robustness of a relatively simple rule supports the plausibility of sensitivity to averages in the context of consumer budget allocations. To our knowledge, prior research has not considered the influence of melioration in budget allocation processes prior to making distributed choices. Given the core principles underlying value accounting and melioration, and consistent findings from the economics literature, our key hypotheses are:

**H1:** Consumers allocate more funds to budgets with greater average value, holding marginal value constant.
H2: Sensitivity to average value (holding marginal value constant) can reduce expected consumer surplus.¹

In experiment 1, we test H1 by experimentally varying average value separately from optimal marginal value and observing consumers’ budget allocations. In experiment 2, we test both H1 and H2 in an incentivized budget allocation game in which we hold optimal marginal value constant, manipulate average value, and observe both budget allocations and spending decisions.

EXPERIMENT 1

In experiment 1, we consider consumer budget allocation across multiple vendors, where all options and preferences are known at the time of allocation, and participants only make allocation decisions (not subsequent spending decisions). We expect to find that, holding constant the number of supramarginal options (i.e., options with values greater than or equal to the marginal good and therefore belonging in the most-valuable consumption bundle), as the average value of options increases, the budget allocation will increase as well, even though that implies it necessarily comes at the expense of total value.

Method

Participants. 101 participants recruited from Amazon Mechanical Turk (AMT) completed experiment 1. One participant’s data were unusable due to a technical issue, resulting in an analyzable sample of 100.

¹ Whether such sensitivity to the average may be adaptive in a broader sense is beyond the scope of this investigation.
Design and stimuli. Each participant made 10 budget allocation decisions in a single experimental session. Each of those 10 budget allocation decisions involved lunch vouchers for a simulated two-week pay period of 10 weekdays each.

On each simulated weekday in each two-week period, one of two vendors, chosen at random, offered a made-to-order lunch with a daily chef. Vendor 1 (“Chef4U”) was staffed by chefs Andre, Betty, Chad, Dale, Emily, and Forest. Vendor 2 (“LetsEat!”) was staffed by chefs Alice, Bernard, Carter, Debbie, Earl, and Frida. Participants read that they preferred chefs whose names began with A to those whose names began with B, those whose names began with B to those whose names began with C, and so on. For each day, vendors and chef names were drawn randomly with replacement; the only constraint was that there could not be a tie between the 7th and 8th most-preferred chefs in a two-week pay period to ensure there was a single value-maximizing allocation, given the budget constraint of seven vouchers discussed below.

Participants imagined their employer provided them with seven lunch vouchers for each two-week period, and they needed to specify how many vouchers they would allocate to each of the two vendors. All 10 upcoming vendors and chefs’ names were visible as participants made their allocation decisions; see figure 1 for an example of an allocation decision. Budgets across the two vendors were constrained such that they had to sum to exactly seven. As part of the instructions and prior to beginning their allocation decisions, participants had to complete a set of comprehension questions. Preregistration, materials, data, and code are available at

https://researchbox.org/353&PEER_REVIEW_passcode=MIJYNO.
**Results**

*Analysis plan.* There were 10 allocation decisions for each of 100 participants with usable data, for 1000 total observations. For every allocation decision, we calculated three values: (a) the number of supramarginal chefs for vendor 1 (i.e., the number of days that vendor 1 chefs were among the seven most-valuable chefs for that two-week period, and therefore deserved a voucher; for the example shown in figure 1, this is 2: Betty and Dale); (b) the number of submarginal chefs for vendor 1 (i.e., the number of days that vendor 1 chefs were among the three least-valuable chefs for that two-week period, and therefore did not deserve a voucher; for
the example shown in figure 1, this is also 2: Emily twice); and (c) the difference between the average value of vendor 1 chefs and the average value of vendor 2 chefs for that two-week period. To make these calculations, we treat the highest value chefs (Andre, Alice) as 6, the next-highest value chefs (Betty, Bernard) as 5, etc., and calculate the mean across all chefs associated with each vendor separately for each two-week period. The two vendor means are compared to calculate the difference in average between vendors; in figure 1, this is -.83. This implicitly treats the ordinal preference ranking as though it has interval properties. For this analysis, to the extent that the ordinal ranking does not have interval properties, our measure will be a correlated imperfect proxy of the true average, leading to a conservative test. In two cases out of 1000, only one vendor was available for all 10 days; in those cases, we could not calculate a difference in average value, so we treated those two observations as missing.

**Analyses.** We analyzed vouchers allocated to vendor 1 as a function of number of supramarginal options, number of submarginal options, and difference in average value, with errors clustered by participant. If participants allocated in such a way as to maximize total value, we would expect to observe a coefficient of 1 on the number of supramarginal options, a coefficient of 0 on the number of submarginal options, and a coefficient of 0 on the difference in average value. This is not what we observe. Instead, controlling for the number of supramarginal

---

2 We preregistered that we would use the number of total chefs rather than the number of submarginal chefs. This change only affects the coefficient on the number of supramarginal chefs, thereby making it more interpretable. Critically, it affects neither the coefficient estimate nor the test of the coefficient on the average value difference.

3 This difference in average chef values is calculated as the average of Vendor 1 ("Chef4U") chefs minus the average of Vendor 2 ("LetsEat!") chefs. In figure 1, this corresponds to the average of Dale, Betty, Emily, and Emily \((3 + 5 + 2 + 2) / 4\), minus the average of Bernard, Frida, Bernard, Bernard, Debbie, and Carter \((5 + 1 + 5 + 5 + 3 + 4) / 6\).

4 We preregistered that we would use a mixed model and, if we encountered convergence problems, use clustered errors. The mixed model analysis, with or without adjustments to address the singular fit, leads to similar estimates and identical inferences. All analyses with clustered errors use the `lm_robust` function from the `estimatr` package (Blair et al. 2021); all degrees of freedom for such analyses are estimated.
options and for the number of submarginal options (and therefore the number of total options), increasing the difference in average value by 1 point (e.g., Andre vs. Bernard rather than Betty vs. Bernard) increases the budget allocation by 0.14 vouchers (se = .05, t(83) = 2.68, p = .009)\(^5\).

The coefficients on the number of supramarginal and submarginal options were less important for our key hypothesis but indicated that participants were attending to the task. Controlling for the other variables, as the number of supramarginal options increases by 1, the number of vouchers allocated to vendor 1 increases by .56 (se = .05, t(83) = 11.10, p < .001). Controlling for the other variables, as the number of submarginal options increases by 1, even though it ought to be irrelevant, the number of vouchers allocated to vendor 1 increases by .18 (se = .06, t(85) = 2.84, p = .006). The coefficient on the number of submarginal options is significantly smaller than that on the number of supramarginal options (t(84) = 4.55, p < .001)\(^6\).

**Discussion**

Experiment 1 indicates that consumers’ budget allocation decisions are driven not only by the marginal value of potential goods (as indicated by a significant coefficient on the number of supramarginal options), but also by the average value of those goods.

Though not tested in experiment 1, we further propose that the sensitivity of consumers’ budget allocations to average values can conflict with maximization (H2). In experiment 2, we seek to conceptually replicate the key effect of distribution on allocation (H1) from experiment 1, provide a direct test of the downstream consequences for value (H2), and address alternative explanations for this finding.

\(^5\) One might worry that this could be due to misspecified functional form in the other predictors (number of supramarginal and submarginal options). Allowing for non-linearities in the other predictors by using a full set of dummy variables rather than their numeric values as continuous variables led to identical inferences regarding the role of average value (b = 0.17, se = 0.05, t(83) = 3.08, p = .003).

\(^6\) We test the difference in coefficients by reestimating the same model, parameterized using the sum and difference of two focal variables rather than the raw variables themselves.
EXPERIMENT 2

In experiment 2, participants play a budgeting game in which they repeatedly allocate tokens across two budgets and then spend those budgets on purchases worth different numbers of points. We examine both distributional implications for allocation and allocation implications for earnings, while addressing alternative explanations that pose threats to internal validity.

Whereas in experiment 1, the number of supramarginal and submarginal options for each spending category varied across trials and participants, in experiment 2 we hold the number of supramarginal and submarginal options for each spending category constant. Whereas in experiment 1, the difference in average value varied randomly across trials and participants, in experiment 2, we systematically manipulate the supramarginal and submarginal distributions through condition assignment.

Method

Participants. 396 participants recruited from AMT completed experiment 2.

Design. Participants played a budgeting game in which they repeatedly allocated a pool of 23 tokens between two budgets, spendable on blue tiles and red tiles, respectively; different tiles were worth different numbers of points. The point-maximizing allocation was always 9 tokens to blue, 14 tokens to red for all participants. More points led to a higher end-of-game bonus: Participants earned $0.01 for every 10 points over 1650 they earned each week. The game was structured as occurring over a sequence of simulated weeks, but the entire game took place in a single experimental session lasting on average approximately 20 minutes. Participants set

---

7 There were originally 403 complete observations. In six cases, the completed observation followed an incomplete observation with the same identifier, indicating the participant was not a naïve participant, so we excluded those observations from analyses. In a seventh case, there were two complete observations with the same identifier. The second began the survey after the first had finished, so we exclude the second observation.
their own allocations for five practice weeks (which we do not analyze) and the subsequent five game weeks (which we do). In a sixth and final game week, rather than allocating their own budgets, participants were assigned the point-maximizing allocation. Each week, participants spent their budgets by selecting which tiles they wanted to purchase out of 80 total opportunities (16 per day for 5 days). The basic structure of a game week is depicted in figure 2, and we describe the game in the following section. Participants’ allocation decisions allow us to test H1. Their spending decisions allow us to test H2 and rule out alternative explanations.

Participants were randomly assigned to one of four conditions in a 2 (supramarginal distribution: blue high, blue low) x 2 (submarginal distribution: blue high, blue low) design. In the blue high supramarginal condition, blue tiles worth 60 points or more had a higher average value than red tiles worth 60 points or more (realized means across participants: 85 vs. 75); in the blue low supramarginal condition, the reverse held (71 vs. 84). Similarly, in the blue high submarginal condition, blue tiles worth fewer than 60 points had a higher average value than red tiles worth fewer than 60 points (35 vs. 22); in the blue low submarginal condition, the reverse held (23 vs. 37). Preregistration, materials, data, and code are available at https://researchbox.org/353&PEER_REVIEW_passcode=MIJYNO.

*Gameplay.* The full game consisted of instructions with comprehension checks, five practice weeks, five game weeks with self-set budgets, and a sixth game week with a pre-set budget. Each week consists of a budget allocation stage and a spending stage. As in experiment 1, our primary focus for H1 is on the budget allocation stage. We implement and analyze the spending stage to (a) test H2, (b) incentivize budgeting decisions, and (c) rule out alternative explanations. These two stages are depicted in figure 2.
Figure 2. Budgeting game stages. Each of five practice weeks and the first five game weeks followed these steps. First, participants allocated 23 tokens across a blue budget and a red budget. Next, participants made purchase decisions each day for each of five days. Tiles were randomized across days, and participants faced each day’s screen in sequence. Optimal purchases conditional on budgets were shown in bolder colors. In the sixth and final game week, participants did not make a choice during the allocation stage, but rather were assigned an optimal allocation of 9 tokens to blue, 14 tokens to red.

In the budget allocation stage, participants had 23 tokens which they allocated across two budgets: blue and red. The optimal allocation, not disclosed to participants, was 9 blue, 14 red. Our primary outcome for analysis is the average proportion of tokens allocated to the blue budget.
In the spending stage, participants used tokens in their red and blue budgets to buy red and blue tiles. Each tile cost one token from the corresponding budget and was worth a certain number of points indicated on the tile, ranging from 5 to 100. Each simulated day from Monday to Friday each week, participants encountered a 4 × 4 grid of 16 tiles as depicted in figure 2. Participants could purchase as many tiles as they had tokens in the corresponding budget. After making decisions for one day, participants were shown their purchased tiles and then continued to the next day’s grid. Participants were not permitted to revisit previous decisions. Unused tokens 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 six incentivized weeks (the first five of which included allocation decisions) with total incentives averaging approximately 20% of overall compensation. Realized bonuses ranged from $0 to $1.41, with a median of $0.84, and were paid on top of a fixed $3.25\(^8\) participation payment.

*Tile distributions.* The tile distributions were structured such that average value was systematically varied between conditions while marginal value (conditional on optimal allocation) was held constant. An omniscient player would follow two simple rules: (a) each week during the allocation phase, always allocate 9 tokens to the blue budget and 14 tokens to the red budget; and (b) during the spending phase, buy every tile worth at least 60 points and no tile worth less than 60 points. Participants were not told either of these rules.

Tile values were always multiples of 5. A single draw of tiles was used for each participant, such that each participant repeatedly faced the same set of 80 tiles every week throughout the game, but different participants faced different sets. Participants were informed the same set of tiles would be used repeatedly. Tiles were drawn according to carefully

---

\(^8\) The participation payment was $3.00 for the first 21 participants. It was subsequently adjusted in line with realized typical experiment duration.
structured principles. These principles ensured: (a) the optimal allocation and spending rules described above were the same across conditions; (b) if participants spent optimally, deviating from the optimal allocation by up to 2 tokens (e.g., any allocation to blue of between 7 and 11 tokens) would not lead to systematically different outcomes across conditions; and (c) potential bonuses would not systematically differ across conditions. The specific distributions used to ensure these principles held are described in detail in the Appendix and the average supramarginal and submarginal values by condition are given above.

**Decision aid.** Our primary focus is on the budget allocation decision. To reduce uncertainty around spending previously budgeted tokens as a potential confound, we sought to make spending primarily a matter of executing previously determined decisions rather than actively making new decisions. Thus, we provided a decision aid. Specifically, we made the color of tiles that ought to be purchased (conditional on budget allocation) more intense than the color of tiles that ought not to be purchased; see figure 3. Furthermore, if participants attempted to deviate from these recommendations, they were given an alert and an opportunity to adjust their purchases prior to submitting their choices. The decision aid was conditional on budget allocations, such that participants who faced the same distribution of tiles but set different budgets would see different recommended tiles\(^9\). If participants attempted to overspend either of their budgets, they were forced to decrease their spending from that budget prior to finalizing their choices for that day to ensure they did not exceed their budget.

---

\(^9\) The decision aid was not conditional on prior spending. As a result, a deviation from the recommendations early in the week would mean that the remaining recommendations were no longer optimal. If participants deviated from the recommendation, on subsequent days within the same week they were shown a note alerting them to the fact that the decision aid may no longer be accurate given prior deviations.
Summary. In summary, participants repeatedly allocated tokens to two budgets and then spent tokens from those budgets on tiles to earn points. The distributions of tiles were structured such that the average values systematically varied across participants, but optimal allocations were held constant. Participants knowingly faced the same weekly distribution of tiles for the entire session of practice and incentive weeks. The decision aid visualized the optimal tiles, conditional upon budget allocation, thus reducing the complexity of the spending decision.

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 tiles recommended by the decision aid. Of 396 participants, 11 purchased fewer than 50% of the recommended tiles across all 6 game weeks and were thus excluded, leaving a final sample of 385.

Average allocation. We first test H1 by focusing on the average allocation to blue budgets as a function of condition. Note that unlike in experiment 1, the number of supramarginal and submarginal options was held constant, and we consider the effect of average
value by using condition assignment. We calculated average proportion of tokens allocated to blue budget across the first five game weeks (i.e., excluding the five practice weeks and the one pre-set allocation week) and regressed it on supramarginal distribution (+1 = blue high, -1 = blue low), submarginal distribution (+1 = blue high, -1 = blue low), and their interaction. Participants in the blue high supramarginal condition allocated more tiles to blue (M = .470, SD = .078) than those in the blue low supramarginal condition (M = .432, SD = .077; b = .019, se = .004, t(381) = 4.73, p < .001, Cohen’s d = .48). There was no effect of submarginal condition (M_{BlueHigh} = .449, SD = .069; M_{BlueLow} = .453, SD = .089; b = -.002, se = .004, t(381) = -.43, p > .6, d = .04) nor was there a significant interaction (b = -.007, se = .004, t(381) = -1.80, p = .072, d = .18).¹⁰ Figure 4 depicts the distributions across conditions.

**Costly deviations.** Exploratory analyses indicated that points earned were lower when participants set budgets further from the optimal allocation. Using participant-week as the unit of observation, we regressed points earned in each of the six game weeks on the deviation of that budget from the optimal allocation, including clustered errors and participant and week fixed effects. By including participant fixed effects, we account for individual differences in attentiveness, as one might expect that individuals who are less attentive would both tend to set more deviant budgets and earn fewer points conditional on those budgets. This analysis indicated that for every additional token participants deviated from the optimal budget, participants earned

¹⁰ Though the interaction was marginally significant, we hesitate to interpret it for two reasons. First, it was an ordinal interaction, and the simple effect of supramarginal condition was significant within each submarginal condition (blue high: b = .026, se = .006, t(381) = 4.64, p < .001; blue low: b = .012, se = .006, t(381) = 2.06, p = .040). Second, it appears to be driven by extreme outliers. When those two extreme outliers are removed (one of which put all tokens towards blue, one of which put no tokens towards blue), the effect of supramarginal condition remains (b = .016, se = .004, t(379) = 4.57, p < .001, d = .47), but the interaction was no longer even marginally significant (b = -.005, se = .004, t(379) = -1.31, p = .191, d = .13).
30 fewer points ($b = -30.49$, $se = 2.15$, $t(79) = -14.15$, $p < .001$)\(^{11}\). This provides initial evidence for H2.

Figure 4. Experiment 2 budget allocations by condition. Solid lines represent condition means; dashed lines represent condition medians. Dotted lines represent optimal allocations.

Chosen allocation vs. pre-set allocation. The analysis above is useful to characterize how points are lost as a function of misallocating funds, but it relies on participant fixed effects to account for endogeneity concerns. The experiment allows a separate preregistered test of H2, by comparing points earned in week 6, when allocations were preset to be optimal, with points earned in week 5. Participants earned more points when they faced an exogenously-determined allocation pre-set to be optimal in week 6 than when they chose their own allocation in week 5,

\(^{11}\) This estimate is affected by non-linearities, but the results generally held when modeling deviations as a set of dummy variables rather than imposing a linear relationship. A one-token deviation did not significantly differ ($b = -2.98$, $se = 2.45$, $t(170) = -1.22$, $p = .224$), but a two-token deviation did significantly differ ($b = -14.73$, $se = 3.72$, $t(219) = -3.96$, $p < .001$), as did deviations of three or greater ($|b| > 40$, $|t| > 3.9$).
for an average difference of 23.6 points (se = 4.2, t(381) = 5.56, p < .001; 66% of participants earned more points in week 6, 26% earned the same number of points, and just 8% earned fewer points). This was not affected by the manipulations (ps > .2). Because points increased with optimal allocation, we find no evidence that non-normative spending strategies would rationally lead to the selection of non-optimal allocations: better allocations indeed led to more points.

A possible concern with the preceding analysis is that it may simply represent a time trend across weeks unrelated to the preset allocation. Indeed, there was also a trend from week 4 to week 5 in points earned (an average difference of 13.0 points, se = 4.4, t(381) = 2.92, p = .004). This raises the concern that perhaps the preregistered analysis merely captures a continued trend due to experience rather than an effect of allocation. Though nearly twice as large in magnitude, the additional points earned in week 6 vs. week 5 was not statistically significantly larger than those earned in week 5 vs. week 4 (difference in differences of 10.7 points, se = 7.0, t(381) = 1.52, p = .130). However, the non-significant difference in differences appears to be driven by the noisiest of the participants who still passed the exclusion criteria. Using any of a variety of stricter criteria (though not preregistered), the difference in differences is significant. For example, if one only includes participants who always chose the recommended tiles in the first three weeks of the game (for which performance does not enter into this analysis), the average difference in differences is 24.3 (se = 7.3, t(291) = 3.32, p = .001). Participants earned more points when budgets were pre-set to the optimal allocation, and exploratory analyses indicate that (at least for reasonably attentive participants) this is not due to an existing trend.

Costs of misallocation that vary by condition. One might be concerned that the effect of supramarginal condition could be entirely due to incentives conditional on imperfect play. That is, allocating all tokens to blue is a less-costly mistake when in the blue high supramarginal
condition than when in the blue low supramarginal condition, so participants might avoid such an allocation more when in the blue low supramarginal condition than when in the blue high supramarginal condition. Corresponding logic holds for allocating all tokens to red.

Such logic holds for gross misallocation, but not for moderate misallocation, because the distribution around the margin (±2) did not systematically vary across conditions (see Appendix). For attentive participants who make the recommended purchases, blue allocations of any amount between 7 and 11 are not susceptible to this critique, because the cost of misallocation does not differ across conditions.

We test this account in exploratory analyses using participant-week observations from weeks 1 through 5 (we use weeks 1 through 5 because we subsequently consider allocation decisions). First, we construct a set of six dummy variables for each person-week reflecting whether the blue budget was 6 or less, 7, 8, 10, 11, or 12 or more, leaving a blue budget of 9 as the excluded reference level. Then, we regressed points earned on that set of dummy variables and their interactions with supramarginal condition, including fixed effects for participant and week, with errors clustered at the participant level. Given the distribution of tiles, relative to allocating 9 tokens to blue budgets, allocating 12 or more tokens to blue budgets was associated with earning more points per week when in the blue high supramarginal condition (-62 vs. -103; \( t(168) = 2.64, p = .009 \)), whereas allocating 6 or fewer tokens to blue budgets was associated with earning marginally fewer points per week when in the blue high supramarginal condition (-158 vs. -90; \( t(82) = -1.68, p = .097 \)). In other words, consistent with the concern above, there were differential costs to gross misallocations across conditions. Importantly, and consistent with the way the experiment was designed and tiles were selected, none of the other interactions were significant (\( ps > .16 \)), indicating there is no evidence that moderate misallocations were
differentially costly across conditions. (These estimates are rather noisy. All conclusions hold when limiting analyses to participants who always followed recommended purchases.) These patterns are depicted in figure 5.

![Figure 5](image_url)

*Figure 5. Adjusted points earned per week as a function of budget allocation. Given different participants faced different numbers of total potential points, point totals are adjusted by participant such that all participants have a maximum of 1820 (the median), conditional on perfect play. Panel A represents all participant-weeks. Panel B limits data to participants who always followed the decision aid’s recommendations. Notice, particularly in panel B, missed points are symmetric between conditions within [7, 11] and asymmetric outside that range.*

Given strong conceptual reasons based on stimulus design that were consistent with the empirical evidence, we repeat our main analyses on the subsets of person-weeks that should not be susceptible to this concern. Specifically, using only person-weeks for which the budget allocated to blue was between 7 and 11 inclusive (representing 68% of total observations), we regressed budget on supramarginal distribution, submarginal distribution, and their interaction, all contrast-coded, including week fixed effects with errors clustered at the participant level.
Again we observe a significant difference in supramarginal distribution \( (b = 0.144, \text{se} = 0.052, t(303) = 2.76, p = .006) \). We observe a similar effect if we limit the blue budget to between 8 and 10 inclusive (representing 44% of total observations; \( b = 0.103, \text{se} = 0.035, t(195) = 2.96, p = .003 \)). Though these internal analyses involve subsetting based on the dependent variable, they provide further corroborating evidence that the observed sensitivity to average is not due to costs of misallocation that differ by condition.

**Experience.** Finally, one might ask whether experience attenuated this effect. There was no evidence of this. In exploratory analyses, we examined the linear trend in allocation to blue across the first five game weeks (which themselves followed five practice weeks) as a function of condition. The intercept was negative \( (t(381) = -3.96, p < .001) \), indicating that participants allocated less to blue on average across weeks, but there was no significant effect of supramarginal condition \( (t(381) = 0.68, p > .4) \), submarginal condition \( (t(381) = 0.14, p > .8) \), nor the interaction \( (t(381) = -1.68, p = .093) \) on the linear trend. In other words, there is no evidence the effect of the manipulation attenuated with experience.

**Discussion**

In experiment 2, all participants faced the same number of supramarginal and submarginal options in each category. Yet, the average supramarginal value and average submarginal value was systematically varied across conditions. As a result, participants allocated more tokens to the blue budget when it had a higher supramarginal average than when it had a lower supramarginal average; submarginal average had no effect. Thus, experiment 2 provides further support for H1, revealing that consumer budget allocations are sensitive to average value, holding marginal value constant. This was driven by the supramarginal distribution rather than the submarginal distribution. In experiment 2, we have sufficient precision to meaningfully
compare the coefficients; the coefficient on above-threshold distribution is statistically significantly larger than the coefficient on below-threshold distribution \( (t(381) = 3.69, p < .001) \). Experiment 2 also provides support for H2: misallocation was costly, suggesting sensitivity to averages can come at the expense of sensitivity to margins. We hasten to add that in many circumstances, average value and marginal value likely covary.

In order to reduce the role of uncertainty, as discussed below, the decision aid depicted tiles that ought to be purchased differently than those that ought not be purchased. As a result, recommended tiles were likely more salient than non-recommended tiles. This could conceivably contribute to the relative effect of supramarginal vs. submarginal distribution, though results from experiment 1 were consistent without such a recommendation tool influencing relative salience. Importantly, it is not clear how mere salience could explain sensitivity to average value rather than marginal value. In addition to the decision aid, purchased tiles were displayed to participants immediately following their purchase. While this may reinforce the salience induced by the very act of purchase, this is also intrinsic to everyday purchase behavior: we live with our purchases, not the goods we leave on the shelves.

**GENERAL DISCUSSION**

Consumers’ budget allocations matter because they affect spending. Budgets ought to be set to equate marginal value across budgets. Yet, in many consumption situations, average value

---

12 A parallel exploratory analysis of experiment 1 results was consistent. When difference in overall average value was split into difference in supramarginal average value and difference in submarginal average value, difference in supramarginal average value was a marginally significant predictor \( (b = .10, se = .05, t(77) = 1.95, p = .055) \) and difference in submarginal average value was not \( (b = .03, se = .06, t(68) = .49, p > .6) \). However, experiment 1 was not well-powered to test the difference between these coefficients, and the difference was not significant.
and marginal value covary, and prior work on distributed choice indicates average value is a powerful driver of choice. The current work drives a wedge between marginal and average value to examine the independent contribution of average value to budget allocation decisions. In two experiments, we find consistent evidence that consumers are sensitive to average value, not just marginal value, when allocating funds across budgets. This holds even when all alternatives are salient at the time of allocation (experiment 1) and when participants follow through on their spending decisions in an incentive-compatible budgeting task (experiment 2). This sensitivity is driven by the value of to-be purchased options rather than the value of all potential purchases (experiment 2). When budget allocations were pre-set to the point-maximizing allocation, performance increased, including for maximally-attentive participants, suggesting misallocation can be costly (experiment 2).

*Three Sources of Uncertainty*

An important possible counter-explanation to our proposed process that consumers are sensitive to *average* value is that they are instead sensitive to *expected* marginal value. In other words, optimal hedging in the face of uncertainty could lead consumers to allocate more funds to the budget with a higher average value in a rational attempt to maximize expected value. As we are interested in sensitivity to average value per se, not due to its implications for uncertainty, it is worth discussing why this cannot account for the current results.

We consider three interrelated sources of uncertainty. First, consumers may not know which purchase opportunities they will encounter. Using our opening example, Alex may know there will be Taco Nights on an unpredictable schedule, so some months there may be two taco nights and other months six taco nights. Such uncertainty regarding *opportunity sets* implies that the value of the marginal purchase may not be knowable even if the distribution from which
opportunities are drawn is known; this will affect consumers’ budget allocations and could favor over-allocating to budgets with higher average value. This does not account for our results. In experiment 1, the entire opportunity set was known at the time of the budget allocation. In experiment 2, the opportunity set remained constant through all practice and game weeks, providing an opportunity for learning.

Second, conditional on setting a budget, consumers may have uncertainty regarding optimal spending rules: having set her budget, Alex may not know whether she ought to buy a pizza. In experiment 1, this was directly resolvable through known preferences across chefs. In experiment 2, the decision aid directly provided the optimal spending rule.

Third, conditional on a budget allocation and a known spending plan, consumers may be uncertain regarding their rule implementation: Alex may intend to hold out for a final Taco Night but make a mistake and opt for pizza instead. In experiment 1, it is conceivable (though in our opinion not likely) that sophisticated participants who both (a) considered their downstream choice behavior and (b) considered the possibility of mis-choosing may have accounted for uncertainty regarding rule implementation. In experiment 2, however, all allocation results hold when only considering the 69% of participants who followed the optimal spending rule on all 115 occasions throughout the five game weeks in which they set their own allocation.

Thus, whereas sensitivity (vs. insensitivity) to averages to address uncertainty may increase or sustain expected surplus, sensitivity to averages due to melioration can decrease expected surplus. While we do not argue that consumers will never be sensitive to averages to resolve uncertainty, this is not the sole driver of such sensitivity.
Learning and Uncertainty

Recent critiques of research on melioration have pointed out that melioration can be rationalized as maximizing behavior given reasonable prior beliefs and lack of information about unexplored states of the world (Sims et al. 2013). Our paradigm differs from the traditional melioration paradigm in at least two key ways. First, payoffs do not depend on the sequence of decisions. Second, payoffs are known with certainty at the time a spending decision is made. Especially given the cases considered in experiment 1 (in which all options are known at the time of allocation) and experiment 2 (in which the same distribution of tiles was used repeatedly), this critique does not apply with equal force to the current investigation.

Participants had the opportunity to learn. Although in experiment 1 they received no feedback, participants in experiment 2 learned their precise point earnings each week. Furthermore, these participants exhibited no apparent decrease in effect size during the five allocations, which followed five practice weeks. This indicates insufficient learning was not a driving factor, even with arguably more-precise feedback than consumers receive in the wild. Of course, we cannot rule out that perhaps with more experience, the effect would attenuate, but there is simply no evidence suggesting that in our data.

Future Directions

While helping to address the key question of how consumer budget allocations align with consumption value, this research leaves open a number of remaining questions. In our aim to cleanly manipulate average value while holding marginal value constant, it was important to precisely identify the values (or as in experiment 1, at least the preference orderings) of individual choice options. Were we to instead measure the idiosyncratic value of options across consumers with heterogeneous preferences, even a small amount of measurement error could
substantially limit our ability to cleanly isolate and identify the effect of average value separately from the effect of marginal value. We recognize that the present stimuli represent a stylized simplification of true consumption value. When making budgets for their daily expenses, an additional point of difficulty will typically be the fact that the purchase-to-value mapping is not as precise as we have identified. Such effects may add to the effect of average we observe here but are unlikely to diminish it. For example, we observe the effect in experiment 1 where all we know are preference rankings, not preference strengths.

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, their long-run selves may seek to constrain spending opportunities available to their short-run selves 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 marginal value and average value of each.

In addition to the important inter-group differences, we also observe considerable intra-group differences in budget allocation, suggesting the presence of meaningful degrees of heterogeneity in allocation decisions. 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 exhibit such effects, as planners are more likely to consider their opportunity costs (Spiller 2011; Bartels and Urminsky 2015; Fernbach, Kan, and Lynch 2015). In contrast, thinking holistically rather than analytically (Choi, Koo, and Choi 2007) may lead consumers to focus on the set as a whole rather than on the marginal purchase.

Finally, we set prices identically across products within and between categories (e.g., one tile costs one token). This makes the task tractable for researchers and reduces the number of relationships that need to be learned for participants, providing a conservative test. But in consumers’ natural consumption environments, prices vary. What is the consequence of such variance? When average consumption value influences allocation, is this average value assessed across products, across dollars, across bundles, or across some other metric? This remains an important question neither addressed nor addressable with our current data.

Implications

*Budgeting patterns.* The current findings indicate consumers may allocate too much to categories from which they derive the greatest per-unit value, all else equal. If consumers place some weight on the average rather than the marginal purchase, categories with a few stand-out favorites are likely to draw an outsized share of wallet. 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, rather than the marginal purchases which are unlikely to be typical of category value. Similarly, 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 affect the value of the marginal
purchase and the reallocation comes at the expense of the marginal purchase in a different category.

*Budgeting tools.* The current work suggests a potential dimension for budgeting tools to focus on: recouping value at the margin. 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. The current findings suggest that with respect to budgeting, greater consideration given to identifying the marginal purchase may help consumers to reallocate in helpful ways. 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.

*Cascading implications.* Finally, these findings with respect to budget allocations are likely to have additional downstream impacts. First, these are not outcomes that disappear in equilibrium. Neither prior work on melioration nor our current work on budgeting 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. Second, because these patterns affect the first stage of the budget process (allocation), they have implications for the second stage (spending) as well, as we find with evidence supporting H2.

Even when the effects on quantitative performance may be modest, the qualitative effects on consumer experience may be large: The more important impact may be *differential* spending rather than *suboptimal* spending. In experiment 2, a blue point is worth the same as a red point, and in experiment 1, Betty cooks as well as Bernard. In our daily lives, we may not be significantly less satisfied with an extra meal out rather than an extra movie, but the experience is meaningfully distinct.
In chasing consumer surplus, consumers’ budgets are systematically distorted by non-normative factors. These factors influence how much they spend in different categories, and through that, how much value they derive and the kinds of goods they purchase. Consumers’ tendency to chase consumer surplus thereby has the potential to meaningfully shape downstream consumption behavior in multiple ways.
REFERENCES


APPENDIX

Figure A1 depicts the specific distributions from which tiles were drawn in experiment 2. The first row indicates the common portions of the blue and red tile distributions used in all conditions. The second row indicates the distribution for the lowest two supramarginal tiles and highest two submarginal tiles, again set to be common to all conditions. The third row depicts the submarginal distribution when blue is high and red is low; when blue is low and red is high, these distributions were swapped. The fourth row depicts the supramarginal distribution when blue is low and red is high; when blue is high and red is low, these distributions were swapped. By drawing tiles from these distributions, there were always exactly 9 blue tiles worth at least 60 points and there were always exactly 14 red tiles worth at least 60 points, but the average values systematically varied with condition. Finally, the bottom row depicts a sample draw from the blue low supramarginal, blue high submarginal condition, typical of what a participant in that condition may have seen.

If an individual draw would lead to a maximum earnable point total of less than 1750 or greater than 1900, the draw was discarded and drawn again. This ensured that participants had a chance at a bonus of at least $0.60 and no more than $1.50.
Figure A1. Distributions from which tiles were drawn in experiment 2. Row 1 ensured that possible points did not differ across conditions. 5 blue tokens were drawn from the light blue distribution; 12 blue tokens were drawn from the dark blue distribution; 12 red tokens were drawn from the light red distribution; 5 red tokens were drawn from the dark red distribution. Row 2 ensured deviations of up to 2 tokens from optimal would lead to symmetric outcomes. 2 tokens were drawn from each of the light blue, dark blue, light red, and dark red distributions. These tiles were the highest submarginal (either 50 or 55) and lowest supramarginal (either 60 or 65) tiles available. Rows 3 and 4 depict the manipulation of the submarginal and supramarginal distributions. Row 5 depicts a single sample draw a participant may have seen.