Title

Essays on Consumer Decision Making and Financial Constraint

Author

Dolifka, David Andrew

Publication Date

2025

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Consumer Decision Making and Financial Constraint

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy in Management

by

David Andrew Dolifka

2025

ABSTRACT OF THE DISSERTATION

Essays on Consumer Decision Making and Financial Constraint

by

David Andrew Dolifka

Doctor of Philosophy in Management

University of California, Los Angeles, 2025

Professor Stephen A. Spiller, Chair

Consumers must constrain their spending to ensure they have ample resources for the future. This dissertation explores how consumers perceive, manage, and respond to their financial constraint. Chapter 1 explores consumers' subjective sense of feeling like they have a small or large amount of money. This chapter focuses on the relationship between how money is earned and how it is perceived (and, ultimately, spent). The key findings are that work wellbeing increases the perceived size of a consumer's income. Consequently, consumers with high work wellbeing experience less financial constraint and engage in more discretionary consumption. Chapter 2 asks how consumers decide exactly how much they can spend, based on different presentations of their own financial information. A descriptive survey suggests consumers overwhelmingly think about their constraint in terms of two different financial metrics: either income or balance. We present evidence that consumers do not fully account for the informational differences between income (a flow) and balance (a stock). As a result, attention to either financial metric influences consumers' assessments of current constraint, with the potential

to influence spending—and savings—over time. Chapter 3 investigates how consumers establish their own constraint in the context of budgeting. Our core proposition is that the psychology of allocating (i.e., establishing a constraint) is distinct from the psychology of purchasing. Specifically, allocating disproportionately engages evaluations at the category level, making budget allocations especially sensitive to the perceived average value of a category. This can be suboptimal, as budget constraints based on average value can lead to consumption outside of the set implied by the principle of marginal analysis. Taken together, these chapters shed light on the psychology of financial constraint and its broader implications for everyday consumer decisions.

The dissertation of David Andrew Dolifka is approved.

Catherine Mogilner Holmes

Hal E. Hershfield

Franklin Peter Shaddy

Stephen A. Spiller, Committee Chair

University of California, Los Angeles

2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Through moments of struggle and doubt, you have been a source of support, inspiration, and commiseration.

Through moments of excitement and discovery, you have been the ones discussing—or perhaps debating—ideas and pushing me to grow in how I think about research and the world.

And most importantly, through all the other moments—the ones that made six years fly by—you have been my family, my mentors, my dearest friends, and the people who have come closest to understanding me. I cherish these ordinary moments with you the most, whether they were listening to your comments during a seminar; being [un]productive together in the office; realizing the Zoom meeting is over time, again; sharing a meal, or a drink; inexplicably accumulating ping pong balls; exercising; playing spikeball; or probably beating you in shuffleboard. (kidding)

I am grateful for these moments and this community.

Too many are deserving of this gratitude, and none are deserving of inadvertent omission, so let me simply say to all of you: *Thank You*.

Chapter 1 is a manuscript currently under review at the *Journal of Consumer Research.* David Andrew Dolifka is the sole author of this manuscript and is responsible for all aspects of the research.

Chapter 2 is a manuscript being prepared for initial submission. This paper is joint work with Stephanie Smith and Stephen A. Spiller. David Andrew Dolifka wrote the manuscript and conducted all analyses. David Andrew Dolifka designed and collected data for the descriptive survey and study 1. All three authors contributed to the design and data collection of studies 2 and 3.

Chapter 3 is a manuscript being revised for the *Journal of Consumer Research.* David Andrew Dolifka and Stephen A. Spiller both contributed to the study design, data collection, data analysis, and manuscript preparation.

CURRICULUM VITAE

David Andrew Dolifka

Employment

Instructor, Eccles School of Business at the University of Utah 2024 - Present

Education

Ph.D.	University of California, Los Angeles Management (Marketing Area)	2025 *expected*
B.A.	Middlebury College, *Summa cum Laude* Economics, *Honors*	2013

Research Interests

Consumer Financial Decision Making, Economic Psychology, Mental Accounting, Income Inequality

Publications (* *Denotes equal authorship*)

Dolifka, David*, Katherine L. Christensen*, and Franklin Shaddy (2024), "Highlighting Opportunities (Versus Outcomes) Increases Support for Economic Redistribution" *Social Psychological and Personality Science.*

Carpenter, Jeffrey*, and **David Dolifka*** (2017), "Exploitation Aversion: When Financial Incentives Fail to Motivate Agents," *Journal of Economic Psychology*, 61, 213-224.

Working Papers

Dolifka, David, and Stephen A. Spiller. "Budgeting Increases Reliance on Category-Level Evaluations," *Preparing for 3rd round review at Journal of Consumer Research.*

Dolifka, David. "Happy Workers are 'Rich' Consumers: Work Wellbeing Increases Discretionary Spending," *Under 2nd round review at Journal of Consumer Research*.

Dolifka, David. "Conducting Experimental Research on Shopify," *Working paper available* at https://ssrn.com/abstract=4451011.

Dolifka, David, Stephanie M. Smith, and Stephen A. Spiller. "Consumer Responses to Income Versus Balance Information," *Preparing for initial submission at Journal of Marketing Research*.

Dolifka, David, and Stephen A. Spiller. "Experiencing Undeserved Financial Gains Early vs. Late: The Moderating Role of Morality," *Early working paper.*

Selected Works in Progress

Dolifka, David. "Outsourced Discounts: The Two-Fold Appeal of Non-Seller Rebates," *Data Collection.*

Rude, Eitan*, **David Dolifka***, and Stephen A. Spiller. "Perceived Portfolio Performance and Consumer Spending," *Data collection.*

Wang, Roy, **David Dolifka**, and Stephen A. Spiller. "New Money or More of the Same? Tracking the Meaning of Money as it Grows," *Data collection.*

Teaching Experience

MKTG 4020 (Undergraduate capstone course): Instructor (2024 – Present)
Eccles Business School at the University of Utah

- Instructor rating: 5/5

MGMT 411 (MBA core marketing): Teaching Assistant, nine quarters (2019 – 2023)

- Teaching assistant rating: 5/5

Invited Guest Lecturer: "Conducting Online Experiments in the Social Sciences"
UCLA Graduate School of Education, PhD statistical inference class (2019-2022)

INTRODUCTION

Few dimensions in life carry as much consequence as one's personal finances. While human beings represent far more than the sum of their material wealth, it is undeniable that having access to more financial resources provides people with more options in life. And while wealth does not necessarily imply happiness (Kahneman and Deaton 2010; c.f., Killingsworth 2021; Killingsworth, Kahneman and Mellers 2023), it does correspond to many critically important aspects of our lives. On average, consumers' finances relate to where they choose to live (Brown-Saracino 2017), the schools they attend (Bowles and Gintis 2002), the life partners they choose (Oppenheimer 2000), their retirement age (Modigliani 1966), the food they eat (Drewnowski and Darmon 2005), and their physical (Bloom and Canning 2000) and mental health (Ridley et al. 2020). Therefore, it is important to improve our understanding of how consumers make financial decisions.

At the core of the consumer's financial decision making is their financial constraint. Conceptually, this constraint is the amount of money that is perceived as available to be used. Both subjective (e.g., feelings of wealth) and objective (e.g., financial metrics) aspects of the consumer's finances can influence this constraint, which in turn guides decisions involving spending and saving (e.g., Gasiorowska 2014; Hamilton et al. 2019; Netemeyer et al. 2018; Tully, Hershfield and Meyvis 2015; Tully and Sharma 2022; Zauberman and Lynch 2005).

Seeking to understand how consumers perceive constraint somewhat resembles trying to hear their inner financial dialogue around the question: "Can I afford this?" (e.g., the 'planner' and 'doer' from Shefrin and Thaler 1988). The broad goal of my dissertation is to move closer towards hearing—and hopefully understanding—how the consumer's inner voice responds to this question. Such research is important because this voice persists from moment to moment, day to day. Therefore, the motivation for this research is to improve our ability to predict what this inner voice will say, thus lending insights to consumer decisions across varied financial contexts.

The three chapters of this dissertation generally align with three different aspects of consumer financial constraint: (1) How consumers subjectively feel about their resources; (2) how they use and respond to different metrics pertaining to their resources; and (3) how they establish their own constraint through budgeting.

Chapter 1 seeks to better understand consumers' subjective sense of whether they feel rich or poor. This inquiry begins by asking whether the source of financial resources—the consumer's work—influences how they perceive the size and adequacy of those resources. The chapter focuses on how consumers evaluate their own resource adequacy through comparison (Tully and Sharma 2022) and proposes that consumers will compare their income against the income they feel they deserve (based on their work wellbeing; e.g., see Rosen 1986). A set of experiments and an analysis of secondary data suggests consumers who experience high work wellbeing feel—and spend—as though they are less financially constrained.

Chapter 2 explores how consumers determine exactly how much money they can spend. This chapter begins with the acknowledgement that consumers *must* limit their spending to avoid overconsuming in the present. Therefore, we consider how consumers identify their level of constraint. The manuscript focuses on two candidate measures for assessing and determining constraint, based on the consumer's objective finances: Income and balance. Drawing upon work on stock-flow reasoning (Cronin, Gonzalez and Sterman 2009; Spiller, Reinholtz and Maglio 2020; Sweeney and Sterman 2000) and debt aversion (Prelec and Loewenstein 1998; Soster, Gershoff and Bearden 2014; Wertenbroch, Soman and Nunes 2001), we suggest consumers will be likely to think about constraint in terms of whichever metric is provided. But because income is a flow and balance is a stock, thinking about constraint in terms of income versus balance can lead to differences in spending in the current moment, as well as differences in accumulation outcomes over time.

Chapter 3 investigates how perceived value guides constraint. Specifically, we consider the case of budgeting, where consumers establish their own constraint to guide subsequent consumption (Zhang et al. 2022). The paper suggests that consumers use a different mode of evaluation when allocating funds to budgets, compared to when they later use those funds for spending. As a result, budget allocations tend to disproportionately reflect the average value of a set of options, which may serve as a suboptimal guide for subsequent consumption. This chapter therefore considers how consumers set their own level of constraint based on different dimensions of value.

Together, these three chapters explore financial decision making with a particular focus on how consumers perceive, respond to, and manage their financial constraint.

REFERENCES

Bloom, David E. and David Canning (2000), "The Health and Wealth of Nations," *Science*, 287(5456), 1207–9.

Bowles, Samuel and Herbert Gintis (2002), "Schooling in Capitalist America Revisited," *Sociology of Education*, 75(1), 1-18.

Brown-Saracino, Japonica (2017), "Explicating Divided Approaches to Gentrification and Growing Income Inequality," *Annual Review of Sociology*, 43(1), 515–39.

Cronin, Matthew A., Cleotilde Gonzalez, and John D. Sterman (2009), "Why Don't Well-Educated Adults Understand Accumulation? A Challenge to Researchers, Educators, and Citizens," *Organizational Behavior and Human Decision Processes*, 108(1), 116–30.

Drewnowski, Adam and Nicole Darmon (2005), "Food Choices and Diet Costs: An Economic Approach," *Journal of Nutrition*, 135(4), 900–904.

Gasiorowska, Agata (2014), "The Relationship between Objective and Subjective Wealth Is Moderated by Financial Control and Mediated by Money Anxiety," *Journal of Economic Psychology*, 43, 64–74.

Hamilton, Rebecca W., Chiraag Mittal, Anuj Shah, Debora V. Thompson, and Vladas Griskevicius (2019), "How Financial Constraints Influence Consumer Behavior: An Integrative Framework," *Journal of Consumer Psychology*, 29(2), 285–305.

Kahneman, Daniel and Angus Deaton (2010), "High Income Improves Evaluation of Life but Not Emotional Well-Being," *Proceedings of the National Academy of Sciences*, 107(38), 16489–93.

Killingsworth, Matthew A. (2021), “Experienced Well-Being Rises with Income, Even above $75,000 per Year,” *Proceedings of the National Academy of Sciences*, 118(4).

Killingsworth, Matthew A., Daniel Kahneman, and Barbara Mellers (2023), “Income and Emotional Well-Being: A Conflict Resolved,” *Proceedings of the National Academy of Sciences*, 120(10).

Modigliani, Franco (1966), “The Life Cycle Hypothesis of Saving, the Demand for Wealth and the Supply of Capital,” *Social Research*, 33, 160–217.

Netemeyer, Richard G., Dee Warmath, Daniel Fernandes, and John G. Lynch (2018), “How Am I Doing? Perceived Financial Well-Being, Its Potential Antecedents, and Its Relation to Overall Well-Being,” *Journal of Consumer Research*, 45(1), 68–89.

Oppenheimer, Valerie K. (2000), “The Continuing Importance of Men’s Economic Position in Marriage Formation,” in *The Ties That Bind: Perspectives on Marriage and Cohabitation*, 283–301.

Prelec, Drazen and George Loewenstein (1998), “The Red and the Black: Mental Accounting of Savings and Debt,” *Marketing Science*, 17(1), 4–28.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel (2020), “Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms,” *Science*, 370(6522).

Rosen, Sherwin (1986), “The Theory of Equalizing Differences,” *Handbook of Labor Economics*, 1(C), 641–92.

Shefrin, Hersh M. and Richard H. Thaler (1988), “The Behavioral Life Cycle Hypothesis,” *Economic Inquiry*, 26(4), 609–43.

Soster, Robin L., Andrew D. Gershoff, and William O. Bearden (2014), "The Bottom Dollar Effect: The Influence of Spending to Zero on Pain of Payment and Satisfaction," *Journal of Consumer Research*, 41(3), 656–77.

Spiller, Stephen A., Nicholas Reinholtz, and Sam J. Maglio (2020), "Judgments Based on Stocks and Flows: Different Presentations of the Same Data Can Lead to Opposing Inferences," *Management Science*, 66(5), 2213–31.

Sweeney, Linda Booth and John D. Sterman (2000), "Bathtub Dynamics: Initial Results of a Systems Thinking Inventory," *System Dynamics Review*, 16(4), 249–86.

Tully, Stephanie M., Hal E. Hershfield, and Tom Meyvis (2015), "Seeking Lasting Enjoyment with Limited Money: Financial Constraints Increase Preference for Material Goods over Experiences," *Journal of Consumer Research*, 42(1), 59–75.

Tully, Stephanie M. and Eesha Sharma (2022), "Consumer Wealth," *Consumer Psychology Review*, 5(1), 125–43.

Wertenbroch, Klaus, Dilip Soman, and Joseph Nunes (2001), *Debt Aversion as Self-Control: Consumer Self-Management of Liquidity Constraints*, Fontainebleau, France: INSEAD.

Zauberman, Gal and John G. Lynch (2005), "Resource Slack and Propensity to Discount Delayed Investments of Time versus Money," *Journal of Experimental Psychology*, 134(1), 23.

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

CHAPTER 1: HAPPY WORKERS ARE 'RICH' CONSUMERS: WORK WELLBEING INCREASES DISCRETIONARY SPENDING

ABSTRACT

Work provides consumers with income, which can be evaluated objectively (the dollar amount) or subjectively (the perceived size). The current research considers how work wellbeing affects consumer spending. A set of controlled experiments involving real and imagined work, a longitudinal study observing spending over time, and an analysis of data from the New York Federal Reserve Bank's Survey of Consumer Expectations suggest work wellbeing increases discretionary consumption. This effect occurs because work wellbeing makes consumers feel like they are earning a larger income, compared to what they require (i.e., their reservation wage). Perceiving more income subsequently leads to more discretionary spending. These findings cannot be easily explained by differences in objective income, future income expectations, mood, affect, or general wellbeing.

Keywords: work wellbeing, income, subjective financial assessments, consumer spending

Work is the bedrock of American society and culture. A person's job is their livelihood, providing the means to satisfy basic needs and pursue greater goals. Much of life is spent at work. The average person spends about half of their waking hours on the job during the workweek (Bureau of Labor Statistics 2024). According to some estimates, this amounts to approximately 30% of a lifetime spent working (Naber 2017). People believe what they do reflects who they are: 55% of Americans report getting a sense of identity from their job (Riffkin 2014). Work is among the most defining aspects of a person's life, so naturally, researchers and practitioners have taken a keen interest in work wellbeing. For example, the *World Wellbeing Movement* recently launched a stream of related research initiatives (e.g., Cunningham et al. 2024), and *Indeed*—the world's largest job recruitment platform—recently began measuring work wellbeing and publishing its "Global Work Wellbeing Report" (Indeed 2024). Beyond the noble goal of understanding wellbeing at work for its own sake, some research considers the downstream consequences of work wellbeing (e.g., employee productivity, firm performance, etc.) (Bellet, De Neve and Ward 2023; De Neve, Kaats and Ward 2023). One apparent gap in this literature involves how work wellbeing impacts the consumer, which is the focus of the present research.

The paper is organized as follows. The first section defines and discusses work wellbeing. The second section draws on labor economic theory to connect work wellbeing to consumers' financial perceptions. The basic argument goes as follows: A consumer's given level of income feels large when they experience more work wellbeing because this intrinsic reward makes them willing to work for less money, making actual income feel large by comparison. Consequently, when consumers feel like they earn a larger income, they should make more discretionary purchases. The third section provides support for these proposed relationships through a set of experiments involving real and imagined work, a longitudinal survey observing natural variation in work wellbeing, and an analysis of secondary data from the Survey of Consumer Expectations. The paper concludes with a discussion of key findings, contributions, implications, and future directions.

WORK WELLBEING

What is Work and How is Work Wellbeing Evaluated?

Work is the perceived set of inputs (generally forms of labor) exchanged for income. The nature and composition of work varies across jobs, and work consists of one or more tasks. Some workers—such as a gig worker hired to paint a bedroom—may earn income for a single task. Other workers may earn income for labor spread across multiple tasks within a single job. For example, a professor's work requires labor spread across different tasks (e.g., conducting research, teaching, mentoring, service, etc.). Individual task-level evaluations are combined to form global evaluations of work (Taber and Alliger 1995). Therefore, task-level evaluations predict the overall evaluation of work, regardless of whether work consists of a single or multiple tasks (Spector 1997).

As every worker personally understands, the nature and quality of work can vary. A multitude of scales and constructs seek to measure workers' assessments of their job (Aziri 2011; Judge, Hulin and Dalal 2012), and these tools typically focus on specific facets of the work experience (Spector 2022). The present research is interested in consumers' overall subjective impression of their work. Ultimately, this reflects the extent to which people experience wellbeing directly through their work. Work wellbeing draws from decades of research on subjective wellbeing (Diener et al. 1999; Diener, Scollon and Lucas 2009; Diener and Ryan 2009; Kahneman and Krueger 2006; Lyubomirsky 2001) and is a simple, subjective reflection on a person's own job. Work wellbeing is distinct from subjective wellbeing in its domain specificity. Meta analysis suggests the correlation between job and life satisfaction is about 0.36 (Bowling, Eschleman and Wang 2010), highlighting the importance of understanding wellbeing across life's many facets (Diener et al. 2009; Kahneman and Krueger 2006; Spector 2022). Consistent with this approach, the current paper considers how work wellbeing impacts consumer judgments, decisions, and behavior.

The structure of work wellbeing resembles the structure of subjective wellbeing, with three key dimensions: evaluative, affective, and eudaimonic (De Neve and Ward 2023).

The evaluative aspect reflects a high-level, cognitive assessment of work. This encompasses a constellation of attitudes regarding the nature of tasks, organizational structure, flexibility, fringe benefits (e.g., childcare, healthcare), compensation, etc. (De Neve and Ward 2023). Such overall assessments of work resemble measures of "job satisfaction" (Judge et al. 2012; Spector 1985, 1997, 2022; Taber and Alliger 1995; Weiss 2002), which is the degree to which people like their work, in all that it entails (Spector 2022). Notably, this evaluative aspect of work wellbeing tends to be relatively stable from day-to-day but may drift over time (De Neve and Ward 2023). In summary, the evaluative aspect is the extent to which a person likes their job, and this should be relatively unchanging in the short term.

The affective component is the degree of felt and experienced enjoyment derived directly through work. This component depends on both characteristics of the job and the person. For example, the job may involve enjoyable or unenjoyable tasks; or the group of assigned coworkers could be pleasant or unpleasant to be around. These are characteristics of the job. Yet, different people may find enjoyment through different tasks (Bhattacharjee and Mogilner 2014; Mogilner, Kamvar and Aaker 2011). Some may prefer low-effort jobs, while others could experience the same job as boring or aversive. Some may find spontaneity fun (e.g., Oh and Pham 2022), though others may prefer sticking to a strict schedule. Some may prefer a quiet day at the office, whereas other people may prefer a lively workplace full of social connection. To the extent that job characteristics can change from day to day (e.g., changing demands, allocation of labor, work environments, etc.), the affective component of work wellbeing is likely to fluctuate over time (De Neve and Ward 2023).

The eudaimonia dimension reflects meaning and purpose from work (De Neve and Ward 2023). This form of wellbeing may arise from the awareness that individual contributions serve some greater good, such as a nurse who improves the health of others, or an accountant employed by a green energy company (Cassar and Meier 2018). The eudaimonia aspect of work wellbeing extends beyond organizational mission. Workers may experience a sense of meaning simply through being able to track their own progress, have a sense of autonomy in their work, or feeling like their contributions are valued (Ariely, Kamenica and Prelec 2008; Dik, Byrne and Steger 2013; De Neve and Ward 2023; Ryan and Deci 2001).

In summary, work wellbeing resembles subjective wellbeing in structure. Taken together, the evaluative, affective, and eudaimonic dimensions capture a person's current, subjective evaluation of their job. As recommended in De Neve and Ward (2023) and consistent with the operationalization by Indeed's "Global Work Wellbeing Report," work wellbeing is measured as the average of those three dimensions. Work wellbeing correlates with—but is meaningfully distinct from—general wellbeing, as people compartmentalize and separately evaluate different areas of their lives (Bowling et al. 2010; De Neve and Ward 2023). The remainder of this paper explores whether work wellbeing might impact consumer judgments, decisions, and behaviors.

SUBJECTIVE INCOME AND WORK WELLBEING

Beyond their objective level of wealth, consumers subjectively assess their financial state. Though researchers may use different terminology and measures to capture subjective financial assessments (see Tully and Sharma 2022), such measures generally consider the perceived size or adequacy of money (Ahn, Ateca-Amestoy and Ugidos 2014; Clark and Oswald 1996; Dias, Sharma and Fitzsimons 2022; De La Rosa and Tully 2022; Mishra, Mishra and Nayakankuppam 2006; Morewedge, Holtzman and Epley 2007; Paley, Tully and Sharma 2019; Spiller 2011; Tully, Hershfield and Meyvis 2015; Wertenbroch, Soman and Chattopadhyay 2007; Zauberman and Lynch 2005). These assessments are made across a variety of wealth components, including income (Clark and Oswald 1996; Gasiorowska 2014; Morewedge et al. 2007; Tully et al. 2015). Subjective financial assessments are important predictors of consumer behavior, even when controlling for objective finances (Fernbach, Kan and Lynch 2015; Gasiorowska 2014; Hamilton et al. 2019; Karlsson et al. 2005; Morewedge et al. 2007; Paley et al. 2019; Tully et al. 2015; Tully and Sharma 2022). Therefore, this research considers how work impacts subjective income assessments, which in turn affects spending.

Income Comparisons and Subjective Income

Income is an objective amount that carries a subjective meaning. For example, $1,000 of weekly income is an objective amount of money, but this amount only gains subjective meaning through interpretation and comparison (Tully and Sharma 2022). This $1,000 might therefore be interpreted by the consumption it represents (Zauberman and Lynch 2005), how it stacks up against co-workers (Ahn et al. 2014; Clark and Oswald 1996; Sharma and Alter 2012), or whether it is more or less than prior paychecks (Loewenstein and Prelec 1993). I propose another such comparison: a consumer's own reservation wage, which is the amount of income they require for their work (Friedman 1957). This is a natural point of comparison for subjectively interpreting an amount of income because of the high degree of similarity between earned and required income (Evers, Imas and Kang 2022; Goldstone 1994; Tversky 1977). The key proposition is that consumers will interpret their $1,000 of income—in part—by comparing that to the amount of income they feel like they require. Subjective income reflects the consumer's assessment of whether their paycheck feels like a lot or a little, compared to what they require (the reservation wage). Just as consumers feel wealthy when their income exceeds a coworker's, I expect consumers to feel wealthy based on the degree to which their income exceeds their own requirements. Therefore, subjective income should grow as reservation wages fall. And what affects reservation wages?

Work wellbeing decreases reservation wages, according to labor economic theories of compensating wage differentials. The logic is that workers trade off the intrinsic and extrinsic rewards of work (Rosen 1986). When work is more intrinsically rewarding, people demand fewer extrinsic rewards (Ryan and Deci 2000). Therefore, extant theory suggests an increase in work wellbeing leads to a decrease in reservation wage; and a decrease in work wellbeing leads to an increase in reservation wage. Empirical and experimental research uncovers this pattern across different work domains, including both job seekers as well as currently employed people (Ariely et al. 2008; Astebro et al. 2014; Burbano 2016; Chen et al. 2019; Leete 2001; Stern 2004).

These findings are typically used to predict whether people will work. For example, the reservation wages of Uber drivers predict the labor supply (Chen et al. 2019). This and related research (Chen et al. 2020) clarify (i) workers have reservation wages; (ii) workers have an awareness of their reservation wages, as reflected by their behavior; and (iii) reservation wages are dynamic and change over time. For these reasons, the reservation wage is an excellent point of comparison—not just for deciding whether to work—but for interpreting income received through work. As an example, extant research suggests an Uber driver might think "I would not earn enough to justify driving tonight, given what I feel I require." I propose that same driver might look back on their work and think "I did not earn enough tonight, given what I feel I require." This subjective sense of not earning enough money is subjective income. In both cases, the sufficiency of income is assessed through comparison to a person's reservation wage.

In summary, people work to earn income, and I propose their work wellbeing will influence how they subjectively perceive the adequacy of that income. Specifically, work wellbeing increases subjective income because experiencing work wellbeing drives down reservation wages. When consumers attempt to make sense of "how much is $1,000 of income?" or "how do I feel about my $200 earnings from driving Uber today?", they compare these objective amounts of income to the income they feel they require. An employee who loves their job and would work for very little feels like $1,000 is a lot of money. An Uber driver who enjoyed a day of pleasant passengers and meaningful conversation understands that $200 is a large income for an intrinsically rewarding string of rides.

H1: Work wellbeing increases consumers' subjective income assessments.

H2: Reservation wages mediate the effect of work wellbeing on subjective income assessments.

When consumers have more objective income, they tend to spend more (Friedman 1957). Above and beyond objective measures, consumers are sensitive to their subjective finances (Dias et al. 2022; Fernbach et al. 2015; Gasiorowska 2014; Kahneman and Krueger 2006; Karlsson et al. 2005; De La Rosa and Tully 2022; Mishra et al. 2006; Morewedge et al. 2007; Paley et al. 2019; Shah, Shafir and Mullainathan 2015; Spiller 2011; Tully et al. 2015; Tully and Sharma 2022; Zauberman and Lynch 2005). Consumers who perceive more subjective income also spend more, particularly for discretionary categories in which consumers have the ability to regulate their level of spending (De La Rosa and Tully 2022; Morewedge et al. 2007). Therefore, if work wellbeing increases subjective income, work wellbeing should ultimately increase consumers' discretionary spending.

H3: Work wellbeing increases consumers' discretionary spending.

H4: Subjective income assessments mediate the effect of work wellbeing on discretionary spending.

OVERVIEW OF STUDIES

The paper presents results from six studies, including four preregistered experiments, a longitudinal spending survey, and an analysis of secondary data from the New York Federal Reserve Bank. Data, code, and materials are available at https://researchbox.org/3968&PEER_REVIEW_passcode=EFOMGE.

Study 1 (N = 798) tests the relationships between work wellbeing, reservation wages, and subjective income by manipulating the work wellbeing of an imagined software testing job. The study design holds constant expectations for future income, such that any findings cannot be explained by expectations regarding future promotions, employment, or job loss. Study 2 (N = 500) manipulates the real work of people hired to complete a brief image processing task and replicates the key effect from study 1. This study rules out an alternative explanation that work difficulty—as opposed to work wellbeing—is driving the effect. Studies 3 (N = 399) and 4 (N = 1201) test the potential downstream effect on consumer discretionary spending. Study 3 uses a mediation design to explore the proposed process that work wellbeing increases subjective income, which in turn increases discretionary spending. Additionally, study 3 measures and controls for positive and negative affect to address an alternative explanation that differences in participants' mood or affect are driving the results (i.e., affective spillover). Study 4 further clarifies the proposed process through moderation, in which wellbeing associated with work has different effects than general wellbeing (associated with non-work). Study 5 (N = 1442) is a two-wave, longitudinal survey that examines changes in consumers' real work wellbeing and discretionary spending over time. Finally, study 6 analyzes data from a newly constructed dataset that merges observations from three surveys administered by the New York Federal Reserve

Bank. Results are presented in the form of an interactive specification table, and readers can easily test and observe the results from 835302 model specifications using the provided tool: https://specificationtable.shinyapps.io/shiny/. A summary of study measures and key results is given by table 1.1.

TABLE 1.1: SUMMARY OF STUDY MEASURES AND KEY RESULTS

		Key Measures		
Study	***N***	**Subjective Income**	**Spending**	**Other finding(s)**
1	798	Size of $55,000 income [0-100 slider measure] (H1: ***)		Reservation wage mediates the effect between work wellbeing and subjective income (H2: ***) *Alternative addressed: anticipated future income does not explain the effect*
2	500	Size of task compensation [1-7 scale] (H1: ***)		*Alternative addressed: work challenge does not explain the effect*
3	399	Size of income from own job with imagined changes [1-7 scale] (H1: ***)	Purchase likelihood for a previously considered discretionary product [1-7 scale] (H3: ***)	Subjective income mediates the effect of work wellbeing on purchase likelihood (H4: ***) *Alternative addressed: affective spillover. Measured positive and negative affect controlled in model*
4	1201		Purchase likelihood for a previously considered discretionary product [1-7 scale] (H3: ***)	*Alternative addressed: general wellbeing. Work wellbeing and non-work wellbeing operate differently.*
5	1442		Self-rated discretionary spending in the last week [1-7 scale] (H3: **)	Cross-sectional analysis: work wellbeing associated with more discretionary spending (H3: *)
6	6775		835302 model specifications are reported through an interactive specification table; 100% of the theoretically preferred models find a significant association between work and discretionary spending	

NOTE— +$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$. Measurement details are given in brackets and significance tests to corresponding hypotheses are given in parentheses.

STUDY 1: WORK WELLBEING INCREASES SUBJECTIVE INCOME

Does work wellbeing increase subjective income? (H1) Can this be explained by higher work wellbeing producing systematically lower reservation wages? (H2) To address these questions, study 1 experimentally manipulates imagined work enjoyment of a prospective job. The study design holds both the objective income and employment horizon constant, allowing for a direct test of how workplace wellbeing impacts consumers' financial perceptions.

Method

A total of 799 complete responses were collected through Prolific. Prior to the manipulation, all participants were asked to imagine they needed to find a new job for exactly one year and that several jobs offered one-year contracts. The fixed duration of employment was emphasized by the following statement: "Keep in mind, this is a one-year contract: you will be required to remain in the job for a full year and you will not be permitted to remain beyond a year." The purpose of the one-year contract was to ensure income perceptions could not be reasonably explained by participants' expectations about how long they would remain in the job (e.g., remaining in a high-wellbeing job longer, therefore earning more over the lifespan, therefore rationally perceiving more income).

FIGURE 1.1: JOB POSTINGS (STUDY 1)

Daily Times

Weather: P.31 $2.70

EMPLOYMENT

Position: Video Game Tester

Overview: We are expanding and currently seeking to hire multiple temporary Video Game Testers (1-year contract, non-renewable).

Responsibilities: The primary responsibility of Video Game Testers is to play our games at various stages of development. Testers spend their time exploring our games, leveling-up characters, inventing game strategies, and becoming experts with game mechanics. Choose from Action, Role-Playing, Puzzle, Sports, and Strategy genres, available for both mobile and desktop platforms. Provide feedback to the development team.

Qualifications: No prior experience is required for this position. Just be prepared to spend every day playing video games.

Position Type: 40 hours per week. One-year contract.

Salary: Inquire directly.

Daily Times

Weather: P.31 $2.70

EMPLOYMENT

Position: Accounting Software Tester

Overview: We are expanding and currently seeking to hire multiple temporary Accounting Software Testers (1-year contract, non-renewable).

Responsibilities: The primary responsibility of Accounting Software Testers is to use our products at various stages of development. Testers spend their time examining spreadsheets, populating sample account information, verifying calculations, and repeatedly testing our software portfolio. Perform repeated daily tests on tax preparation, payroll, business accounting, and auditing software, on both mobile and desktop platforms. Provide feedback to the development team.

Qualifications: No prior experience is required for this position. Just be prepared to spend every day testing accounting software.

Position Type: 40 hours per week. One-year contract.

Salary: Inquire directly.

NOTE—A side-by-side comparison of the one-year software tester job descriptions used in the high work wellbeing (left) and low work wellbeing (right) conditions in study 1.

Participants were then randomly assigned to read one of two newspaper job postings for a growing company in need of temporary software testers. Both jobs offered full employment for exactly one year (40 hours per week) and required no prior experience. Both jobs had an unlisted salary, asking applicants to "inquire directly." The key manipulation was whether the software tester job was for video games (high work wellbeing) or accounting software (low work wellbeing). The full job postings are presented in figure 1.1.

After reading the job description, participants indicated their reservation wage for the prospective job. Reservation wages were carefully elicited through a two-part question. The first question asked participants to indicate the "amount you would request for working in this job with a binding one-year contract." This was the "amount you want and think the employer might say 'yes' to." This does not reflect the reservation wage but rather what an applicant may try to negotiate for, either due to strategic motivations or market-based salary expectations. Subsequently, the second question directly measured the reservation wage as the private amount "(not conveyed to the potential employer)" that served as the "minimum salary you would accept to work in this job with a binding one-year contract. (This is the amount you need in order to say 'yes' to working in this job)." Responses to both the public salary request and the reservation wage were collected on the same screen in $2,500 salary intervals (1 = minimum: "$15,000/year"; 30 = maximum: "More than $85,000/year").

On a new page, participants were hypothetically offered the job with a $55,000 annual salary. Subjective income was measured by indicating how small or large the $55,000 salary felt using a slider anchored on 0 = "feels very small" and 100 = "feels very large." Finally, work wellbeing was captured by agreement with three items corresponding to the evaluative, affective, and eudaimonic dimensions. Participants indicated the extent to which they agreed or disagreed with the following statements: "I would like the things I do in this job", "the job seems enjoyable", and "the work seems rewarding."

Results

One participant's Prolific ID was associated with two observations. To preserve naïveté, this participant's second observation was dropped. This resulted in a final sample of 798 observations (M_{age} = 38.7 years, 49% female). Attrition is examined across all studies (appendix F). Differential attrition is not a concern in the current data ($\chi^2(1) = 1.36$, $p = .381$).

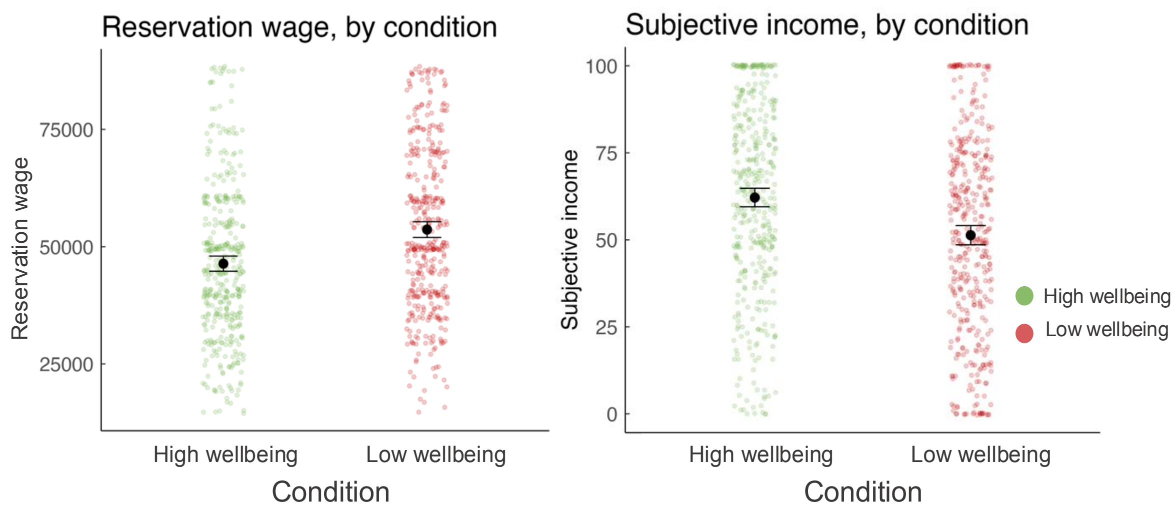
The manipulation had the intended effect on work wellbeing ratings. Anticipated work wellbeing (the average of the three measures; $\alpha = 0.93$) was indeed greater in the high wellbeing condition ($M = 5.70$, $SD = 1.38$) than in the low wellbeing condition ($M = 4.00$, $SD = 1.76$; $t(795)$[1] $= 15.11$, $p < .001$).

On average, participants demanded $49,987 for the job (measured in annual salary bins ranging from 1-30; $M = 15.00$, $SD = 6.88$). These reservation wages differed across conditions. Participants who considered the video game testing job demanded an average of $46,372 ($SD$ = $16,362), which was less than those who considered the accounting software testing job (M = $53,640, SD = $17,260; $t(796) = 6.10$, $p < .001$, Cohen's $d = 0.43$). This finding is consistent with predictions on compensating wage differentials (Rosen 1986). Of particular interest is how participants perceive the size of the $55,000 income. Consistent with H1, the $55,000 felt larger in the high wellbeing condition ($M = 62.15$, $SD = 27.22$), compared to the low wellbeing condition ($M = 51.32$, $SD = 28.04$; $t(796) = 5.54$, $p < .001$, Cohen's $d = 0.39$) (figure 1.2).

A mediation analysis tests the relationship between work wellbeing, reservation wage, and subjective income. The distribution of the indirect effect was estimated through bootstrapping with 5,000 resamples. As expected by H2, the effect of work wellbeing (0 = low wellbeing condition; 1 = high wellbeing condition) on subjective income was mediated by reservation wages, as indicated by the 95% confidence interval around the indirect effect that excludes 0 ($ab = 8.02$, $CI_{95\%} = [5.41, 10.77]$). The direct effect of work wellbeing on subjective income was attenuated and remained marginally significant ($b = 2.75$, $se = 1.47$, $t(795) = 1.86$, $p = .063$).

[1] A careful reader might notice there is one fewer degree of freedom than would be expected given $N = 798$ and two parameters being estimated (the intercept and the effect of condition). This is explained by one participant having 'NA' values for job ratings, thus leading to 795 degrees of freedom, rather than the expected 796.

FIGURE 1.2: RESERVATION WAGES AND SUBJECTIVE INCOME (STUDY 1)



NOTE—Means (black dots), 95% confidence intervals (black bars) and raw data (plotted points) for reservation wages (*left*) and subjective income (*right*).

Discussion

The present study illustrates a straightforward but novel link between working and subjective financial perceptions: Reservation wage and subjective income move in opposite directions (holding constant objective income). This is because in order to interpret the magnitude of actual income, it must be compared against *something*. A consumer's own reservation wage serves as a semantically similar comparand. When reservation wages are low, income feels large (because earning $55,000 when you only require $40,000 highlights a relatively large surplus of income); however, when reservation wages are high, income feels small (because earning $55,000 when you require $52,000 highlights only a small difference). Indeed, work wellbeing increased subjective income (H1), and this effect was explained through decreased reservation wages (H2). These relationships arise when (i) objective income is held constant across conditions (ii) for a job with a binding one-year contract that cannot be shortened or extended. Therefore, these results cannot be explained by differences in (i) objective income or (ii) expectations of future earnings. This foreshadows the results of the secondary data from the New York Federal Reserve Bank (study 6), in which the key relationships are unaffected by variables such as promotion likelihood, expected wage changes, or expected retirement age.

Though illustrative, study 1 has several limitations. First, the sequential elicitation of the reservation wage followed by subjective income may have unnaturally encouraged the comparison of objective income and reservation wages when forming subjective income assessments. If this were the case, the effect of condition on subjective income would be an artifact of the experimental design. However, this seems unlikely: An otherwise identical follow-up study excludes the reservation wage and salary request elicitations and finds a similar effect of the work wellbeing manipulation on subjective income ($b = 11.12$, $se = 2.40$, $t(394) = 4.64$, $p < .001$, Cohen's $d = 0.47$; appendix A). A second limitation is that the current design manipulates only *imagined* work wellbeing, rather than *experienced* work wellbeing. A third, related limitation is that the study focuses on prospective, future work. Though illustrative, impressions of future work prospects represent only a narrow application of the proposed theory. Work wellbeing should also impact how consumers who are *already* employed perceive the adequacy of their income. These limitations are directly addressed in the subsequent study.

STUDY 2: EXPERIENCED WORK WELLBEING INCREASES SUBJECTIVE INCOME

Do actual workers feel differently about the size of their income, depending on the work wellbeing they *experience*? To answer this question, workers were hired on Prolific to perform image processing tasks manipulated to be high or low in work wellbeing. After working for a set amount of time, participants provided feedback (to the employer) regarding how they felt about this size of their earnings, given the work they experienced.

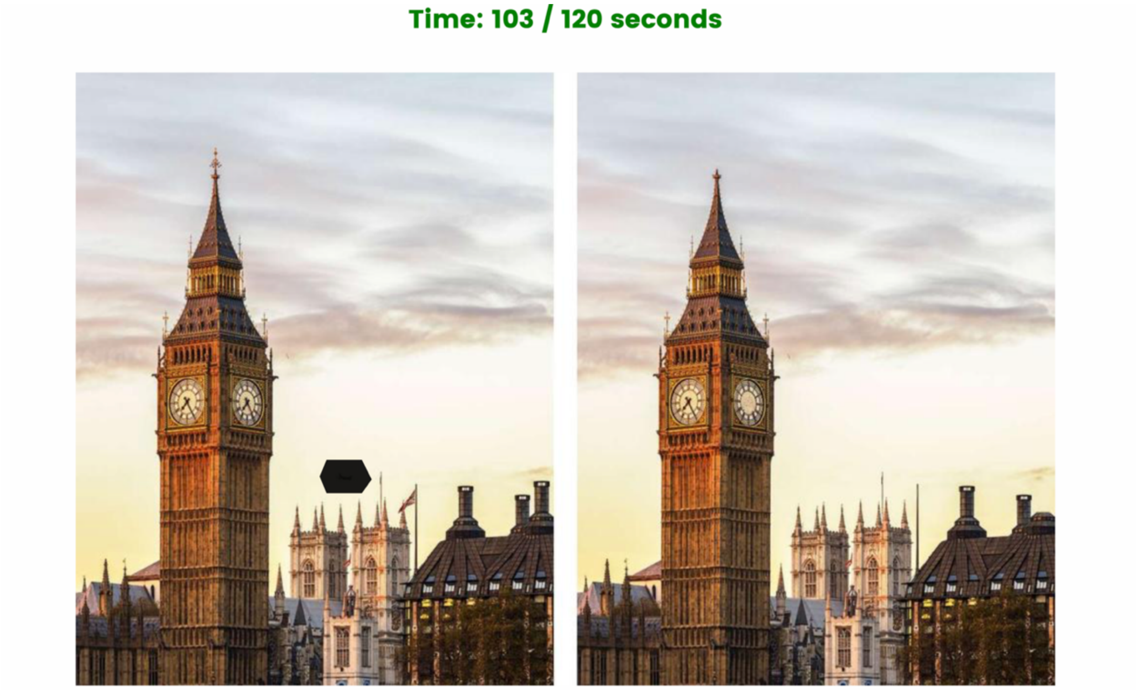
Method

A total of 500 participants were recruited through Prolific as part of a preregistered interim analysis design[2] (André and Reinholtz 2024). The study was intended to resemble hiring for a specific work task (image processing) as opposed to academic research. Participants saw a post titled "Hiring people for quick image processing," and subsequently encountered a set of instructions without an informed consent (as approved through the institutional review board to maintain the appearance of real work). The instructions stated the job was to "process some images that can be used for CAPTCHAs and other verification programs" for two minutes.

Workers were randomly assigned to either the high or low work wellbeing condition. In the high wellbeing condition, participants saw sets of two slightly modified images with slight discrepancies between them (see figure 1.3). Workers were instructed to "find as many [differences] as you can by clicking the LEFT image anywhere you think the two [images] are different." These workers received clear instructions (working until the time elapsed), received performance feedback (the number of correct solutions), while working on an engaging task, all of which contribute to positive work evaluations (Spector 2022).

[2] The preregistration identified $N = 500$ as the interim analysis size, at which point the p-value was below the adjusted $\alpha = .029$; therefore, data collection terminated at the interim sample size.

FIGURE 1.3: HIGH WELLBEING, SPOT-THE-DIFFERENCE TASK (STUDY 2)



NOTE—Example of an image set used in the high work wellbeing condition. Participants searched for differences between the two images—there are six in total—and marked them by clicking the image on the left, as illustrated by the one black region in the left photo.

Workers in the low wellbeing condition also examined images and clicked on focal regions of CAPTCHA images (rather than side-by-side images). However, this task was meticulously and carefully programmed to be glitchy. Regarding the evaluative component of work wellbeing, the task was designed to be boring, redundant, and visually straining. The task itself—repeatedly processing CAPTCHA images—was expected to be boring and redundant. The CAPTCHA images were edited to be blurry, and participants received no feedback about their work progress. Regarding the affective component, the program was designed to frustrate workers by randomly de-selecting the CAPTCHA squares after being clicked, occasionally ignoring button clicks, and displaying a flickering banner that read “Please work quickly and diligently to complete this task." Regarding the eudaimonic component, the study adapts the manipulation from (Ariely et al. 2008), in which the worker's finished product—the completed CAPTCHA—is occasionally not saved, forcing participants to re-do the exact same work. A video demonstration is available at https://youtu.be/eP9ZH04X-8E. Some participants reached out to share their observations of these glitches, such as this respondent who wrote "I was trying to work quickly and diligently as asked but the screen kept flashing, and the images were blurry on the desktop. I would gladly do this full-time if it worked correctly."

The program automatically advanced after two minutes and informed workers "your work has been approved for a payment of $0.60." Regarding the compensation, participants responded to the question "Does your $0.60 payment feel like a small amount or a large amount of money for this task?" on a 1-7 scale anchored on 1 = "A very small amount" and 7 = "A very large amount." Subsequently and on a new page, participants then encountered a set of questions more representative of an academic study. These questions involved the three-item measure for work wellbeing (manipulation check); a measure of work challenge (confound check); and age, gender, and income demographic variables. These questions were measured on the same response scales as in study 1.

Results

All 500 participants were included for the analysis (M_{age} = 38; 53% female). Participants reported a higher level of work wellbeing (α = 0.92) in the spot-the-difference condition (M = 5.88, SD = 1.12) compared to the glitchy CAPTCHA condition (M = 3.96, SD = 1.94; $t(498)$ = 13.85, $p < .001$). There was the expected effect of the work wellbeing manipulation on subjective income, as the \$0.60 income felt larger in the spot-the-difference condition (M = 3.60, SD = 1.19) than the glitchy CAPTCHA condition (M = 3.13, SD = 1.31; $b_{\text{condition}}$ = 0.473, se = 0.112, $t(498) = 4.22$, $p < .001$, Cohen's d = 0.38). Examining all observations independent of condition assignment, respondents who were higher in work wellbeing also rated the size of their income as larger (b = 0.221, se = 0.030, $t(498) = 7.42$, $p < .001$).

Notably, perceived work challenge was significantly higher in the high wellbeing condition (M = 5.07, SD = 1.60) than in the low wellbeing condition (M = 3.90, SD = 1.93; t = 7.38, $p < .001$), ruling out the alternative that work challenge, difficulty, or effort is a confound in the experimental design. As expected, the effect of condition on subjective income is robust to controlling for work challenge ($b_{\text{condition}}$ = 0.464 se = 0.118, $t(497) = 3.91$, $p < .001$).

Discussion

Study 2 hires workers (under the guise of real image processing work) to personally experience work wellbeing. In both experimental conditions, participants spend the same amount of time working. In the high wellbeing condition—which was also rated as the more challenging condition—workers perceived their actual income as larger, compared to the low wellbeing condition. These results suggest the present theory about work wellbeing and subjective income is not limited to assessing future, prospective work. Instead, the proposed effect on subjective income appears to hold for ongoing and completed work. (In other words: This is not a theory limited to job seekers, but rather a theory that generalizes to consumers who are actively employed.) An additional benefit of study 2 is that it moves the findings from the realm of hypothetical to experiential. Whereas study 1 asked participants to imagine an unfamiliar job and then rate a hypothetical salary offer, study 2 directly manipulates the task and then asks workers how they feel about their actual compensation.

STUDY 3: MEDIATION THROUGH SUBJECTIVE INCOME

The first two studies provide experimental support for the proposition that work wellbeing increases subjective income (H1). This occurs for both imagined and experienced work (and income) and cannot be explained by differences in objective finances (held constant), time working (held constant), or perceived work difficulty (measured and controlled for).

The next set of studies move beyond income perception to explore the potential for downstream consequences on consumer decisions. Specifically, these studies consider whether work wellbeing affects discretionary purchases (H3). Furthermore, these studies seek to clarify the proposed process—that work wellbeing affects subjective income perceptions, which in turn affects purchase decisions—in the face of additional alternative explanations. In particular, these studies use both a mediation (study 3) and moderation approach (study 4) to isolate the role of income perceptions as the mechanism underlying any relationship between work wellbeing and consumer spending (H4).

Method

A total of 401 participants from Prolific participated in this experiment. Prior to any manipulation, all participants identified a discretionary purchase as "something you have often considered purchasing but have yet to buy (this should cost $100 or less)." The purpose of this question was to identify a relevant and plausible expenditure for a subsequent measure of purchase likelihood. Next and on a separate page, all participants identified their current or previous full-time job. Participants were then randomly assigned to either the high or low work wellbeing condition.

Participants considered their own job for the manipulation of work wellbeing. Specifically, participants were instructed to imagine their job changing to focus more on the aspects, tasks, and parts of work they either liked (high work wellbeing) or disliked (low work wellbeing). Importantly, the instructions emphasized that salary, title, and hours would remain unchanged. The exact wording is reproduced below.

Now, please think about the very [best / worst] parts of this job. Consider the aspects / tasks of this job you most [liked / disliked], and imagine they are your whole job. Specifically, please imagine your entire full-time job is comprised of your most [liked / disliked] parts of work. Everything else (e.g., salary, title, hours) remains unchanged. The only difference in this job is that you [get to / have to] spend all your time on the [enjoyable / unenjoyable] aspects of work.

Following these imagined work changes, participants indicated if they would be more or less likely to purchase the previously considered expenditure on a 1-7 purchase likelihood scale (1 = “much less likely”, 7 = “much more likely”). They then rated whether the income from their job would feel small or large (1 = “very small”, 7 = “very large”).

The study included additional measures to address possible alternative explanations. Positive and negative affect were measured separately, immediately after the subjective income rating. Positive affect was measured as the level of agreement (1 = “strongly disagree” to 7 = “strongly agree”) with the statement: “I am currently experiencing positive emotions.” The parallel format was used to measure negative affect. Subsequently, participants indicated the objective income they expected to earn through their job, given the imagined changes (measured in $10,000 income bins), to ensure this was not inadvertently affected by the manipulation. As a check on the manipulation, participants indicated whether they anticipated high or low work wellbeing on a 1-7 scale (1 = "not at all enjoyable" to 7 = "very enjoyable.")

Results

Two participants had incomplete responses, leaving an analyzable sample of 399 observations (M_{age} = 38, 59% female).

Manipulation and Confound Checks. The manipulation check identified the expected effect of condition on anticipated work wellbeing ($M_{\text{high wellbeing}} = 5.28$, $SD_{\text{high wellbeing}} = 1.56$, $M_{\text{low wellbeing}} = 3.38$, $SD_{\text{low wellbeing}} = 2.01$; $t(397) = 10.57$, $p < .001$). Somewhat surprisingly[3], there was a marginal effect of condition on objective income expectations ($M_{\text{high wellbeing}} = 6.26$, $SD_{\text{high wellbeing}} = 3.09$, $M_{\text{low wellbeing}} = 5.68$, $SD_{\text{low wellbeing}} = 2.88$; $t(397) = 1.94$, $p = .053$); however, this appears to reflect the random assignment of *actual* income (see appendix B), as opposed to an effect of the manipulation. Controlling for objective income expectations in every subsequent model does not substantively or statistically impact the interpretation of the results.

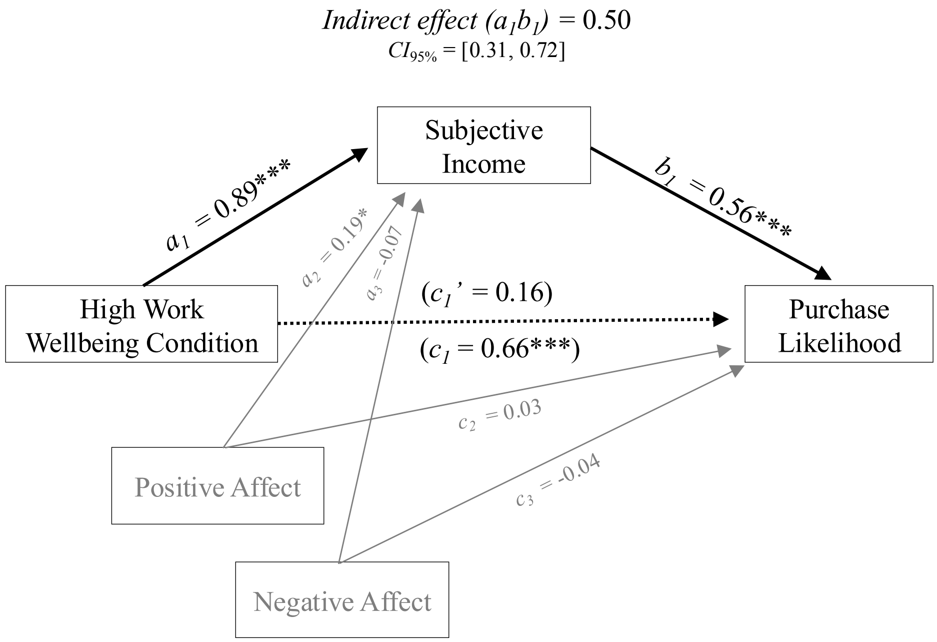
The work wellbeing manipulation also impacted participants' positive and negative affect. There was more positive affect following the high wellbeing manipulation ($M = 5.07$, $SD = 1.33$) than the low wellbeing manipulation ($M = 4.21$, $SD = 1.57$; $t(397) = 5.94$, $p < .001$).

[3] This is surprising because participants were instructed that their actual income would not change and there was no effect of condition on objective income in a preregistered replication of this study ($p = .585$; see appendix B).

Similarly, there was less negative affect following the high wellbeing manipulation ($M = 2.82$, $SD = 1.52$) than the low wellbeing manipulation ($M = 3.73$, $SD = 1.67$; $t(397) = -5.68$, $p < .001$). Given this evidence that imagining job changes impacted participants' affect, a concern is whether participants' moods and feelings can explain the results. Therefore, all analyses are considered with and without controls for positive and negative affect, and objective income expectations.

Primary Analyses. Consistent with H3, there was an effect of condition on purchase likelihood. Participants in the high work wellbeing condition indicated they were more likely to make their previously identified discretionary purchase ($M_{\text{high wellbeing}} = 5.22$, $SD_{\text{high wellbeing}} = 1.61$) than those in the low wellbeing condition ($M_{\text{low wellbeing}} = 4.36$, $SD_{\text{low wellbeing}} = 1.92$; $b = 0.86$, $se = 0.18$, $t(397) = 4.86$, $p < .001$, Cohen's $d = 0.49$). This result remains significant with controls for positive affect, negative affect, and objective income ($b = 0.66$, $se = 0.18$, $t(394) = 3.62$, $p < .001$, Cohen's $d = 0.35$). Furthermore, there was the expected effect on subjective income assessments, such that a person's own income was rated to be larger in the high wellbeing condition ($M_{\text{high wellbeing}} = 4.60$, $SD_{\text{high wellbeing}} = 1.44$, $M_{\text{low wellbeing}} = 3.40$, $SD_{\text{low wellbeing}} = 1.81$; $b = 1.20$, $se = 0.16$, $t(397) = 7.32$, $p < .001$, Cohen's $d = 0.73$). As before, this result is robust to the inclusion of the additional control variables ($b = 0.89$, $se = 0.16$, $t(394) = 5.54$, $p < .001$, Cohen's $d = 0.53$) (H1).

FIGURE 1.4: MEDIATION PATH DIAGRAM (STUDY 3)



NOTE—The preregistered mediation pathways with path estimates, significance levels, and the bootstrapped indirect effect.

Mediation Analysis. According to H4, work wellbeing impacts purchasing through subjective income. I conduct a mediation analysis for the model regressing purchase likelihood on the work wellbeing condition, while controlling for positive and negative affect, with subjective income as the proposed mediator. A bootstrapped 95% confidence interval was estimated around the indirect effect (5,000 resamples). Results are consistent with indirect-only mediation, with a significant indirect effect through subjective income assessments (ab = 0.50, $CI_{95\%}$ = [0.31, 0.72]) and no remaining direct effect of work wellbeing (c_1' = 0.16, p = .336). The pattern of results (figure 1.4) is consistent regardless of whether affect or objective income expectations are included or excluded in the model.

Discussion

Study 3 advances the research in several ways. First, it provides causal evidence for the proposed effect of work wellbeing on consumer spending decisions (H3). The target of these spending decisions was not random but rather a discretionary product that was personally relevant to each participant, thus enhancing the realism of the exercise. Second, the study also examines the effect of work wellbeing on subjective income perceptions for a person's actual income. Whereas study 1 asked participants to consider a $55,000 income for a prospective job and study 2 asked participants to assess their $0.60 compensation, study 3 focuses participants on the actual amount of money they already earn in a year. The consistent effect of work wellbeing across these three instances of income (hypothetical salary, real small wages, actual annual income) is reassuring. Third, the study addresses an alternative explanation that had been untested in the prior experiments. Specifically, could affective spillover (from an affectively rich manipulation) impact subsequent ratings of purchase likelihood and subjective income? While the manipulation of work wellbeing *did* have an effect on both participants' positive and negative affect, the model controls for measured affect and the focal effect persists. These affect-related findings are consistent with a replication of the current study using the PANAS short form scale rather than single measures of positive and negative affect (appendix B).

STUDY 4: MODERATION BY DOMAIN

Whereas study 3 explored process by mediation, study 4 explores process by moderation. Recall, work wellbeing should increase discretionary spending because consumers who experience greater work wellbeing will feel like they are earning more income. Therefore, another way of testing the proposed process is to consider whether a wellbeing manipulation operates different for work versus non-work. The theoretical rationale is that wellbeing at work should affect subjective income, while wellbeing in some other domain (e.g., leisure) should not. Study 4 tests this by manipulating wellbeing and also whether the domain is work or non-work.

Method

1203 participants on Prolific were randomly assigned across a 2 (Wellbeing: high, low) x 3 (Domain: work, volunteer, leisure) experimental design. Volunteering was selected as an activity resembling work, except for the key distinction that work earns an income and volunteering does not. Leisure—specifically, watching a favorite TV show—was selected as the most comparable time use to work. In 2023, the average American spent 5.15 hours each day on "leisure and sports" and 5.02 hours each day working (American Time Use Survey 2024).

Within the work domain, the method exactly matched study 3 through the collection of the dependent variable (change in purchase likelihood). For the other domains (volunteering, leisure), the method differed only in the guided imagination exercise. Those assigned to the volunteer domain identified a current or past volunteer experience. They then read through the manipulation text that was nearly identical to that in the work situation, except words like "work" and "job" were replaced with "experience" and "volunteer" (appendix C). Those assigned to the leisure domain imagined watching a TV show they knew well[4] and then encountered the following manipulation.

[4] It is perhaps fitting that the most-identified show was The Office: a *workplace* comedy.

> Now, please think about the most [likable and/or most enjoyable / irritating and/or least enjoyable] character in the show. Imagine this character gets much more screen time in the TV show. The whole series changes to focus more on the character aspects you find [likable and most enjoyable / irritating and least enjoyable]. The general theme of the TV show remains unchanged, but the character you are imagining plays an increasingly large role. Whatever made you [really enjoy / not enjoy] this character before is now more prominent. Watching this program now reminds you of all of the things you [truly enjoy / do not enjoy] about this TV show.

Upon completing the guided imagination manipulation, participants were instructed to imagine this would be their experience every time they [go to work / volunteer / watch this TV show]. Participants then indicated their purchase likelihood for the previously considered item using the same 1-7 measure as in the prior study. Wellbeing was measured with the same three items from studies 1 & 2, generalized to accommodate the three different domains ("I would like doing this," "the experience seems enjoyable," and "the experience seems rewarding").

Results

Two participant identifiers were linked to prior observations in this study. Removing non-naïve observations resulted in a final sample of 1201 participants (M_{age} = 38.4, 40% female).

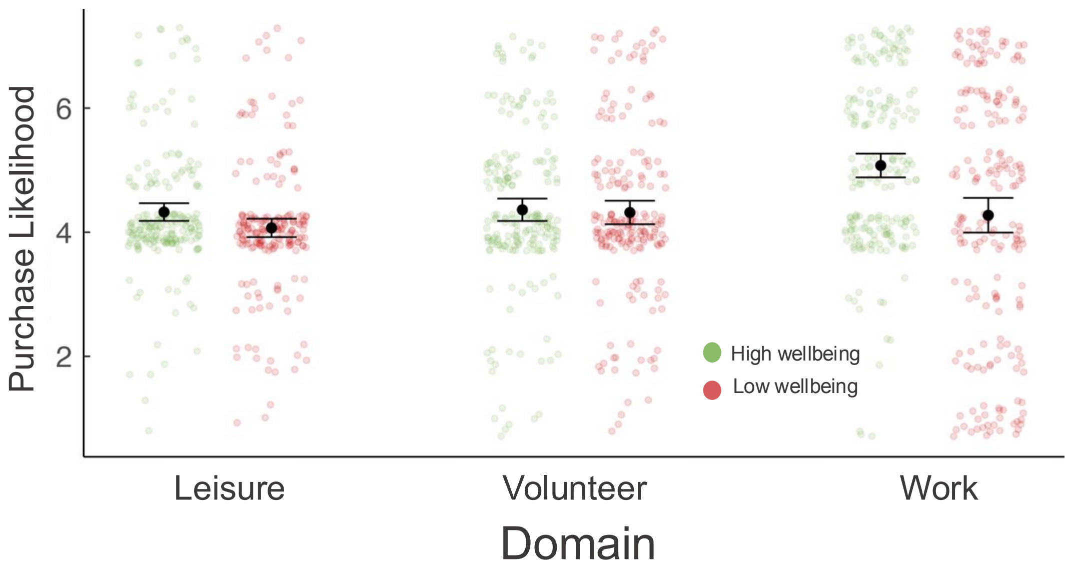
Manipulation Check. The three wellbeing items (liking the situation, enjoying the experience, feelings of reward) were very highly interrelated (α = 0.97) and collapsed into a single measure. There was a main effect of condition on anticipated wellbeing ratings ($M_{\text{high wellbeing}}$ = 5.73, $SD_{\text{high wellbeing}}$ = 1.24, $M_{\text{low wellbeing}}$ = 2.78, $SD_{\text{low wellbeing}}$ = 1.73; $t(1199) = 34.06$, $p < .001$). The simple effect of the wellbeing manipulation on anticipated wellbeing differed across domains, with a larger simple effect for work (*M*s = 5.76 vs. 2.28, Cohen's d = 2.28) than either of the non-work situations (Volunteer: *M*s = 5.85 vs. 3.12, Cohen's $d = 1.90$; Leisure: *M*s = 5.59 vs. 2.90, Cohen's $d = 1.79$), as indicated by tests of the wellbeing condition x non-work condition interactions (both *p*s < .001). However, the preregistered interactions discussed in the following section are robust to controlling for the measures of anticipated wellbeing (appendix C).

Preregistered Analysis Plan. To assess whether the effect of wellbeing on purchase likelihood is unique to *work* wellbeing—as opposed to non-work wellbeing—is a test of interactions. The wellbeing condition was contrast-coded as +1 for high and -1 for low. Two dummy variables indicated assignment to the volunteer or leisure conditions (with work as the reference group). Purchase likelihood was regressed on the contrast-coded wellbeing variable, the two dummy variables, and the two-way interactions. This model simultaneously tests the simple effect of wellbeing in the work domain (a replication of study 3) and the degree to which this effect differs in the non-work domains, as indicated by the interactions.

Purchase Likelihood. Replicating prior results, there was a simple effect of wellbeing in the work condition ($b = 0.401$, $se = 0.069$, $t(1195) = 5.77$, $p < .001$, Cohen's $d = 0.47$). There was no effect of wellbeing in the volunteer condition ($b = .022$, $se = .070$, $t(1195) = 0.31$, $p = .756$, Cohen's $d = 0.03$) and a marginal effect in the leisure condition ($b = 0.128$, $se = 0.069$, $t(1195) = 1.84$, $p = .066$, Cohen's $d = 0.24$). The focal results are the two interactions, which describe the extent to which the simple effect in the work condition differed from the simple effects in the non-work conditions. Each interaction is significant and negative, indicating the effect of high versus low wellbeing is eliminated (volunteer) or attenuated (leisure) in the non-work conditions

($b_{\text{volunteer x wellbeing}}$ = -0.379, *se* = 0.098, *t*(1195) = -3.40, *p* < .001, Cohen's *f* = .08; $b_{\text{leisure x wellbeing}}$ = -0.273, *se* = 0.098, *t*(1195) = -2.78, *p* = .005, Cohen's *f* = .08). The full regression results are presented in appendix table 1.C1 and the raw data with means are plotted in figure 1.5. As previously discussed, these interactions are robust to controlling for differences in measured wellbeing (both *p*s < .05). Robustness checks are documented in appendix C.

FIGURE 1.5: PURCHASE LIKELIHOOD RESULTS, BY CONDITION (STUDY 4)



NOTE— Means (black dots), 95% confidence intervals (black bars) and raw data.

Discussion

Study 4 further clarifies the proposed model (work wellbeing increases subjective income, which in turn leads to more discretionary spending) by examining how wellbeing operates differently across work versus non-work domains. Manipulating the imagined wellbeing in domains of volunteering or leisure had a significantly smaller effect on purchase likelihood than manipulating work wellbeing. This is exactly what should be expected, because *why should* the wellbeing associated with volunteering or leisure impact spending? It should not, according to the current theory, because neither volunteering nor leisure involve income. Study 4 clarifies that work wellbeing is indeed distinct from more general wellbeing, and work wellbeing engages a unique process that causes consumers to feel richer, thus encouraging purchasing. Taken together, studies 3 and 4 provide consistent evidence for this process through both mediation and moderation, while simultaneously addressing and ruling out the concern that the observed effects are due solely to affective spillover from the manipulation, or differences in general wellbeing more broadly. While these studies provide support for the proposed process in contained experimental settings, the question remains whether this relationship exists beyond the lab. Studies 5 and 6 consider consumers' actual spending data.

STUDY 5: LONGITUDINAL SURVEY OF DISCRETIONARY SPENDING

Thus far, the set of experiments has examined how work wellbeing increases subjective income, and in turn, the hypothetical purchase likelihood of a discretionary good. The experiments benefit from designs using real work and real income (study 2) as well as those focusing on a person's own work and own income (studies 3-4). However, the question remains whether these findings extend beyond small-scale, controlled experiments. In other words: Does the relationship between work wellbeing and discretionary spending generalize to actual consumers? Study 5 takes a first step towards answering this question by analyzing consumers' discretionary spending over time, using a longitudinal survey design.

Method

A total of 2002 participants were recruited to take part in a longitudinal study, though participants were unaware there would be multiple rounds of data collection. The initial recruitment (on Prolific) was advertised as a "Quick survey *only for people employed outside of Prolific*." As an additional check on employment status, participants indicated their primary source of income as one of the following: "Working at a job (outside of Prolific)," "Working on Prolific," "Retirement / pension," "Government benefits or family assistance," "Currently no primary source of income," or "None of the above." This question was ultimately used to screen which participants would be invited back for the second wave of the survey. As preregistered, only participants who had a primary income from "Working at a job (outside of Prolific)" were eligible for the second wave, two weeks later.

In both waves of the survey, participants were asked to recall and assess their discretionary spending over "the last week (7 days)." Specifically, participants were informed they should not include "recurring expenses (e.g., bills, rent, etc.), nor should you include necessities (e.g., groceries, gas, essential household products, etc.)." For this recall portion, participants were asked to write down "only purchases you consider to be discretionary from the last 7 days (e.g., dining out, grabbing drinks, new clothes or accessories, entertainment, beauty care, tools & home improvement, electronics, games, furniture, gifts, etc.)." Subsequently, participants' open-ended responses were piped back to them on the next page so they could (1) "estimate the total cost of these purchases (in dollars)" and then (2) rate the extent to which this represents "more spending or less spending than your average 7-day period," anchored on -3 = "much less discretionary spending" and 3 = "much more discretionary spending." As noted in the preregistration, this spending rating is the key variable of interest (the open-ended recall and cost estimation exercises were intended to facilitate recall and evaluation).

After completing the recall and assessment of discretionary spending, participants completed demographic questions, all contained together on the same page. In addition to age, gender, and income questions, participants also indicated "in the last week (7 days), to what extent have you been enjoying or not enjoying your work (at your primary source of income)?" on a 7-point scale anchored on -3 = "work has been very unenjoyable" and 3 = "work has been very enjoyable." This simple measure of enjoyment (the affective dimension of work wellbeing) is appropriate to use in isolation, because affect is the dimension of work that is most likely to vary in the short run (De Neve and Ward 2023). Because the evaluative and eudaimonic aspects of work wellbeing are more stable, these should remain relatively unchanged during a short two-week period, making short-run fluctuations in work wellbeing most consistent with changes in work affect. A final measure of voting preference was included because the 2024 presidential election was set to take place between the first and second waves of the survey[5].

After two weeks, the second wave was launched to all eligible participants, reflecting the 1965/2002 respondents who claimed to earn their primary income from a job outside of Prolific. The survey remained open for exactly three days, during which 1507 complete responses were collected.

Results

Considering the entire sample, there was the expected positive correlation between recent work wellbeing and discretionary spending (b = 0.043, se = 0.020, t(2000) = 2.17, p = .030; this

[5] Voting preference did not meaningfully related to either the dependent or independent variable and was therefore not included in the model for analysis.

is robust to controlling for income, age, gender). However, the key question is whether changes in work wellbeing correspond with changes in discretionary spending, over time. Therefore, the main analyses focus on only the 1507 sets of observation pairs, reflecting participants who completed both waves of the longitudinal study. The preregistered plan removed 64 observation pairs due to reported changes in the primary source of income. One participant identifier was linked with two responses during the second wave, so the redundant observation was dropped. This resulted in a final analyzable sample of 1442 participants (M_{age} = 41, 56% female).

During both waves (two weeks apart), participants provided measures of both work wellbeing (specifically, work enjoyment) and discretionary spending. These measures were used to create the following within-person difference scores: $WORK\_DIFF_{i} = (WORK\_WELLBEING_{i, t=2} - WORK\_WELLBEING_{i, t=1})$ and $SPEND\_DIFF_{i} = (SPENDING_{i, t=2} - SPENDING_{i, t=1})$. I also constructed a set of dummy-coded day variables to account for any possible effects of time (either day-of-week, or days between waves) of the second-wave response. All first-wave responses occurred on the same day. *SPEND_DIFF* was regressed on *WORK_DIFF* and a set of day-of-week dummy variables.

Robust regression results indicate a significant positive relationship between within-person changes in work wellbeing and discretionary spending. Specifically, a 1-unit increase in work wellbeing was associated with a 0.095-unit increase in discretionary spending ($t(1437) = 2.62$, $p = .009$). It should be noted that use of a robust regression—an estimation technique that is more robust to the presence of prediction outliers—is a departure from the preregistration, which did not specify the exact estimation technique. Furthermore, the same analysis using an OLS regression did not show a significant association between changes in work wellbeing and changes in discretionary spending ($b = 0.042$, $se = 0.037$, $t(1437) = 1.31$, $p = .258$). This discrepancy between the two estimation techniques reflects the presence of a small number of highly influential prediction outliers, to which OLS is extremely sensitive. This potential issue likely arises because of the use of within-person difference scores for relatively stable constructs, leading to narrow distributions with long tails. For example, 78% of all respondents had a $WORK\_DIFF_i$ score in the interval [-1, 1], though the range of the data is [-6, 6]. Therefore, observations far from the mean of the data carry exaggerated influence over the estimation. As an illustration, consider the following: There were 7 (out of 1442) participants who deviated 6 units in work wellbeing (2 were -6; 5 were +6). Including these 7 observations with an OLS regression produces a slope of 0.042 ($p = .258$), as previously reported. Running the same OLS regression *without* these 7 observations *doubles* the slope to 0.080 ($p = .037$). This example is merely intended to illustrate how the magnitude of the relationship between 1442 paired observations is remarkably reliant upon a small number of extreme responses.

A detailed explanation and visualization of the data, the outliers (quantified by DFFITS, DFBETA, Cook's Distance, and Studentized Deleted Residuals), alternative specifications, and additional estimators are provided in appendix D. The appendix follows the guide from Aguinis, Gottfredson and Joo (2013) and shows consistent and convergent results that *any* of the recommended techniques (respecification, removal of outliers using any of the metrics used, alternative estimation techniques) coincide with the inferences of the robust regression. Based on the convergent results and inferences documented in appendix D, it appears most accurate and representative of the data as a whole to report robust regression results. Of course, this is left to the reader to decide.

Discussion

Study 5 asks consumers to look back on the prior week and reflect upon their work and spending. The study does not attempt to manipulate work wellbeing but rather capitalizes upon natural variation in work wellbeing over time. And these changes in work wellbeing correspond with changes in discretionary spending. These results underscore an important aspect of this research that may not be apparent in the experimental studies: The proposed links are not theoretically—nor empirically—constrained to individual differences (or job differences). Study 5 offers initial evidence that this is also a within-person phenomenon, suggesting consumers may modulate their spending over time in accordance with how they perceive the adequacy of their income in that moment. This finding has implications regarding the *timing* of consumption appeals, suggesting marketers may find success in identifying the moments when customers perceive their income to be largest. This is further considered in the general discussion.

STUDY 6: SURVEY OF CONSUMER EXPECTATIONS

Beyond participants recruited from online subject pools, is there evidence of the proposed relationship between work wellbeing and discretionary spending on a broader scale? This is explored through a nationally representative survey from the New York Federal Reserve Bank's Survey of Consumer Expectations (SCE) (Federal Reserve Bank of New York 2024). The unique appeal of the Survey of Consumer Expectations is that it is a rare instance of a large, national dataset that includes questions pertaining to both *work* and *consumption*.

Survey Background Information. The publicly available data include observations from 2014-2024, during which heads of households from a nationally representative sample completed one or more different online surveys. A total of ten surveys covering different financial and economic topics were periodically administered, and each survey may have been completed more than once between 2014-2024. This analysis considers and merges observations from three relevant datasets: the SCE Labor Market Survey (for measures of work wellbeing), the SCE Household Spending Survey (for discretionary spending), and the SCE General Survey (for additional control variables). Merging data across the *labor* (31671 responses), *spending* (27870 responses), and *general* (165336 responses) surveys is possible using a unique household identifier. The labor and spending surveys were never completed in the same month, though the intake SCE General Survey may have coincided with either the labor or spending survey.

Method

Dataset Construction. A total of 11907 households appeared in both the *labor* and *spending* surveys after 2014[6], which is the most basic requirement for any form of analysis. Of these, 6775 households had the necessary (non-missing) data in both the *labor* and *spending* surveys. Nearly all of the data loss from 11907 to 6775 households is attributable to individuals who were not employed and therefore were not presented with the work-related questions.

The main challenge with the SCE data is that the *labor*, *spending*, and *general* surveys were administered at different times, often quite far apart. This is unfortunate, as an ideal dataset

[6] The relevant spending data was not measured prior to 2015

would measure work, spending, and a host of relevant control variables at the same moment in time. The best approach to this challenge is to identify observations from the three surveys that are as near to one another as possible. The rationale is that (1) work wellbeing in February and March of the same year are likely similar, especially on the evaluative and eudaimonic dimensions[7] (De Neve and Ward 2023), and (2) discretionary spending in February and March of the same year are likely similar; therefore (3) work wellbeing in February should be related to discretionary spending in March (and vice versa). In addition to minimizing the time between work and spending measures, it is also ideal to minimize the time to the intake (first response) of the general survey, which contains several important demographic, psychographic, and financial variables that could plausibly change over time (e.g., marital status, feelings of financial independence, and home ownership could all change). Therefore, in cases when households had multiple responses for either labor or spending in 2015-2024, it was desirable to consider their first response (as this is guaranteed to minimize the maximum time between surveys, given the intake version of the General Survey occurred first).

A total of 4774/6775 households had the minimum-possible one-month difference between labor and spending survey responses (and the mean distance between the furthest-apart of the labor, spending, and general surveys was minimized to 3.4 months). A total of 1655 households had 3-month gaps between labor and spending surveys. The remaining 346 households had substantially longer gaps and are not analyzed. I separately analyze the households with measures taken the minimum one month apart ($N = 4774$) and the next-best

[7] Study 5 suggests the affective component of work wellbeing is also relatively stable over a two-week span, as 77% of the participants in the longitudinal survey differed in work affect by 1 unit or less (on a 7-point scale).

group of households with measurements taken three months apart (N = 1655) and report all results in the included specification table.

Measures of Work Wellbeing. Work was evaluated through five questions, with the complete question wordings and response scales provided in table 1.2 (and appendix E).

TABLE 1.2: WORK-RELATED QUESTION WORDING & MEASUREMENT (STUDY 6)

Order	Label	Question wording	Measurement	Work wellbeing
Q1	***Pay satisfaction***	How satisfied would you say you are with your level of compensation at your [current/main] job?	1-5 scale (normalized to 7-point)	
Q2	***Non-pay satisfaction***	And how satisfied would you say you are with other aspects of the job, such as benefits, maternity/paternity leaves, flexibility in work hours, etc?	1-5 scale (normalized to 7-point)	Affective dimension
Q3	***Skill fit***	On a scale from 1 to 7, how well do you think this job fits your experience and skills?	1-7 scale	Eudaimonic dimension
Q4	***Promotion likelihood***	On a scale from 1 to 7, how would you rate the opportunities for a promotion or other career progression with your current employer, over the next three years?	1-7 scale	
Q5	***Overall job satisfaction***	Taking everything into consideration, how satisfied would you say you are, overall, in your [current/main] job?	1-5 scale (normalized to 7-point)	Evaluative dimension

The five items correspond to pay satisfaction, non-pay satisfaction, skill fit, promotion likelihood, and comprehensive job satisfaction. Which of these measures should be considered a component of work wellbeing? The evaluative component of work wellbeing *is* comprehensive job satisfaction (Q5) (De Neve and Ward 2023), so this measure is recommended for inclusion. The affective component of work wellbeing reflects day-to-day experiences—and not impressions of pay or promotion—which resembles non-pay satisfaction (Q2). The eudaimonic component of work wellbeing is the extent to which a person finds their work meaningful and rewarding; this most closely resembles whether the job matches an individual's experience and skills (Q3). Therefore, a recommendation is to consider the measures of non-pay satisfaction

(Q2), skill fit (Q3), and job satisfaction (Q5) as reflecting a worker's wellbeing. The reader may easily specify alternative combinations of these five metrics—including treating pay satisfaction and promotion likelihood instead as control variables—and these results are readily available through the online specification table.

Measures of Discretionary Spending. Spending was measured as the proportion of spending across nine different categories: housing, utilities, food, clothing, transportation, medical care, recreation, education, and other/gifts/miscellaneous (exact wordings from original survey are reproduced in appendix E). Prior work on how consumers categorize expenditure types suggests discretionary expenses would include spending on recreation, other/gifts/miscellaneous, and clothing (Zhang et al. 2022). To corroborate this, 100 participants on Prolific (N = 100) read the original SCE category descriptions and rated each expense category on a 1-7 scaled anchored on 1 = "very fixed" to 7 = "very discretionary." As expected, "recreation," (M = 6.35) "other/gifts/miscellaneous," (M = 5.56) and "clothing" (M = 5.53) were rated as the most discretionary expenditure categories. Food was rated as next most discretionary (M = 4.66), and all other expenditure categories had ratings below 3.25. Therefore, it is recommended to consider discretionary spending as the total proportion of spending on recreation, other/gifts/miscellaneous, and clothing. Combining these three discretionary categories—as opposed to viewing one in isolation—is desirable for reducing measurement error. Specifically, expenditures that are clearly discretionary but could plausibly fit in multiple buckets could be lost using a single category approach (e.g., Cheema and Soman 2006). For example, a new ski jacket could be considered recreation (skiing) or clothing. The advised construction of discretionary spending as the total proportion across all three categories reduces this concern. The reader may consider or test alternate constructions of the discretionary spending variable through the specification table.

Control Variables. Inclusion of the SCE general survey generates a comprehensive list of relevant demographic, psychographic, and financial control variables. The included variables are presented in table 1.3. Many of these (those without "D" or "B" in table 1.3) could be independently added or removed from each specification. A set of 10 demographic variables (marked with a "D" in table 1.3) and 9 employer benefits variables ("B") were added or removed from models as a set. This kept the number of models tractable (under 1 million, compared to over 25 billion).

TABLE 1.3: SET OF CONTROL VARIABLES (STUDY 6)

Black	Dummy variable (D)	Education	1-9 scale (D)
White	Dummy variable (D)	Married	Dummy variable
Asian	Dummy variable (D)	Personal health	1-5 scale
Pacific Islander	Dummy variable (D)	Home owner	Dummy variable
Native American	Dummy variable (D)	Financial independence	1-5 scale
Age 18-31	Dummy variable (D)		
Age 32-50	Dummy variable (D)	Are any of the following offered through work?	
Age 51 and over	Dummy variable (D)	*Pension*	Dummy variable (B)
Male	Dummy variable (D)	*Employer retirement contribution matching*	Dummy variable (B)
Expected retirement age	Years	*Health insurance provided*	Dummy variable (B)
Expected wage change next year	Percent	*Dental insurance provided*	Dummy variable (B)
Self employed	Dummy variable	*HSA available*	Dummy variable (B)
Pay satisfaction	1-5 scale (normalized to a 7-point)	*Subsidized housing*	Dummy variable (B)
Promotion likelihood	1-7 scale	*Life insurance*	Dummy variable (B)
Income	Annual, in dollars	*Commuter benefits*	Dummy variable (B)
Logged income	Log (1 + Income)	*Other*	Dummy variable (B)
Income volatility	1-4 scale		

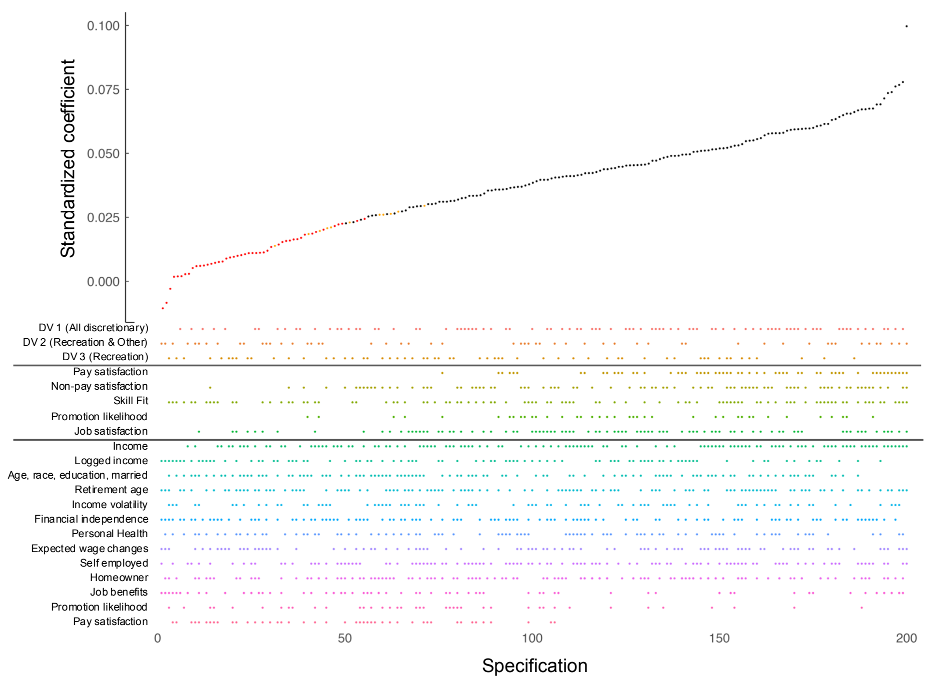
NOTE—Available control variables. Demographic variables indicated with a "D" (race, age, education) were included or excluded as a set. Employer benefits indicated with a "B" (pension, retirement contribution matching, health insurance, dental, HSA, subsidized housing, life insurance, commuter benefits, other) were included or excluded as a set.

Analysis. Given the many possible ways of constructing work wellbeing and the proportion of discretionary spending—as well as the many combinations of control variables—I will analyze the entire set of plausible model specifications. Each model will regress the proportion of discretionary spending (constructed as either spending on recreation; or recreation + other; or recreation + other + clothing) on the work measure (all possible combinations of the 5 variables). The dependent variable will be standardized to allow for comparisons of coefficient estimates across models.

Results

All 835302 specifications can be examined using a simple online application to de/select variables to include in the model. The online specification table is available at: https://specificationtable.shinyapps.io/shiny/.

FIGURE 1.6: SPECIFICATION CURVE (STUDY 6)



NOTE—The curve (top) and specification plot (bottom) present 200 regression models, ordered by the coefficient on the work variable. These models were sampled across the entire space of 417651 models in which work and spending were measured only one month apart.

Each dot in the curve of the upper panel represents the estimated coefficient from the model regressing discretionary spending on a measure of work wellbeing and some set of control variables. Black dots are significant below $p < .05$, orange dots are significant below $p < .10$, red dots are *NS*.

The lower panel presents the model specifications corresponding to each coefficient estimate in the top panel. This includes discretionary spending construction (first 3 rows), the work wellbeing construction (next 5 rows), and the set of control variables (next 13 rows). The purpose of this panel is to visualize whether any variables cluster around models with high or low coefficient estimates (e.g., the last row variable—pay satisfaction as a control—appears frequently when the estimate is small and very infrequently when the estimate is large).

Specification Curve. To help visualize the pattern of results, a specification curve (Simonsohn, Simmons and Nelson 2020) presents 200 randomly sampled model specifications, 20 from each decile ranked by the coefficient estimates on the work variable of interest.

Specification Table. The online specification table presents the results from 835302 different model specifications. Of these, 57% produced a significantly positive ($p < .05$) coefficient estimate on work wellbeing, though this statistic is not independently diagnostic (the frequency of significant models may be inflated or depressed by the presence of unideal models). Decomposing the specifications, we observe a difference in whether work and spending were measured close in time (71% significant using observations one month apart) or further in time (42% significant using observations three months apart). This discrepancy is reassuring, as the work-consumption relationship *should* be stronger when work and spending are measured nearer in time (and weaker when measured further in time).

Probing the models using data one-month apart, the composite measure of discretionary spending (recreation, other, clothing) produces the highest proportion of significant coefficients on the focal work wellbeing variable (76%). Reassuringly, this number decreases as the specification of "discretionary spending" becomes narrower ({recreation, other} = 72% significant; {recreation} = 66% significant), which is to be expected given the higher likelihood of measurement error when using fewer categories.

Using the preferred construction of work wellbeing (job satisfaction as the evaluative dimension, non-pay satisfaction as the affective dimension, and skill fit as the eudaimonic dimension) produces 90% significant specifications (97% are below $p < .10$). Examining these specifications, there is one control variable that makes a clear difference in the proportion of specifications with a significant work coefficient: pay satisfaction. When controlling for pay satisfaction, 61% of the specifications have $p < .05$ and 90% have $p < .10$. When not controlling for pay satisfaction, 100% of the 12283 specifications have $p < .05$ (and 99.65% are significant at levels below $p < .001$).

Discussion

Study 6 constructs a new dataset by merging three different surveys from the New York Federal Reserve Bank's Survey of Consumer Expectations. The question is whether work wellbeing is associated with higher levels of discretionary spending. Results are presented in the form of 835302 regression outputs, from which the reader can navigate models using different constructions of discretionary spending, work wellbeing, control variables, and time lags between measurements. The pattern of results passes numerous sanity checks: Models perform much better when using work-spending data only one month apart, as opposed to three months apart; broader constructions of discretionary spending perform better than narrower constructions; preferred constructions of work wellbeing perform better than alternative specifications using the work variables. Using the preferred specification generates a majority of significant coefficient estimates (90% are $p < .05$), with one important caveat.

The presence of pay satisfaction as a control is the only variable with any meaningful impact on the work wellbeing coefficient. When using pay satisfaction as a control, only 61% of these models have $p < .05$ and only 90% have $p < .10$ on the work coefficient. When pay satisfaction is not used as a control—irrespective of whether it is used in the construction of work wellbeing or omitted entirely—100% of the 12283 models are significant ($p < .05$), with 99.65% significant below $p < .001$. Upon reflection, this is not surprising, as controlling for pay satisfaction conceptually resembles controlling for the mediator (subjective income). To the extent that being satisfied with one's pay level resembles perceiving one's pay as small or large, this is exactly the pattern of results that should be expected. The relationship between subjective income and pay satisfaction is further discussed in the general discussion.

The method of study 6 is also useful in providing the reader the opportunity to consider alternative specifications they feel are more interesting or deserving. Several of the available control variables pertain to alternate accounts already discussed in the set of experiments. For example, variables on promotion likelihood and expected wage change relate to the alternative that work wellbeing influences spending only through rational expectations of future earnings. Retirement age addresses a similar explanation that consumers who derive high work wellbeing may work longer, earning more (objective) income over the lifespan. Good health, not being middle-aged, income, being married, having stable housing, and enjoying a sense of independence are all established correlates of subjective wellbeing more broadly (Diener et al. 1999; Diener and Ryan 2009), helping to clarify the distinct effect of work wellbeing, rather than general wellbeing. Controlling for these variables allows the reader to understand the change in the distribution of the work wellbeing coefficients, advancing not only the current research but enabling deeper exploration as well.

GENERAL DISCUSSION

The worker and the consumer are often treated as separate entities, siloed within separate literatures and areas of study. However, these roles are inherently interconnected, and overlooking their relationship may limit our understanding of both the worker and the consumer. The current research builds on this interconnection and considers the effect of work wellbeing on consumer judgments, decisions, and behavior. Results suggest work wellbeing increases subjective income (H1; studies 1-3), because higher work wellbeing drives down reservation wages, making a given level of objective income feel subjectively larger in comparison (H2; study 1). If work wellbeing makes consumers feel as though they earn a larger income, then work wellbeing should increase discretionary spending (H3, H4). This is indeed the pattern of results. Through controlled experiments involving hypothetical spending (studies 3-4), work wellbeing increases the likelihood of discretionary purchases. These results cannot be explained by mood, affective spillover, or expectations around general wellbeing. In a longitudinal survey capitalizing on natural variation in work wellbeing over time, consumers report higher discretionary spending after experiencing improvements in work wellbeing (study 5). A cross-sectional analysis using a representative sample from the Survey of Consumer Expectations corroborates these findings in a rich dataset full of individual-level control variables. Taken together, these findings suggest that happy workers feel—and spend—as though they are "rich" consumers.

Contributions

Subjective Financial Perceptions. The present research relates most closely to the consumer literature on subjective financial perceptions. Whereas prior research has considered how *outward* comparisons (e.g., social comparisons) affect relative income assessments (Ahn et al. 2014; Gasiorowska 2014; Haisley et al. 2008; Sharma and Alter 2012; Tully and Sharma 2022), this research highlights the role of *inward* comparisons against reservation wages. The link between work wellbeing and subjective income is especially important to the broader literature on subjective financial perceptions because income plays such a pivotal role in the consumer's finances (Gasiorowska 2014). This research is important for consumer researchers at a time characterized by major shifts in employment norms, expectations, and experiences. Notably, the COVID-19 pandemic ushered in a new era of work-from-home employment standards (Aksoy et al. 2022) and a surge in gig economy participation (Statista 2024). To the extent these labor market shifts have an impact on work wellbeing, there may be spending ripple effects that are important to predict and understand.

Retail Therapy. An established literature on compensatory consumption and retail therapy (Atalay and Meloy 2011; Cryder et al. 2008; Lee and Böttger 2017; Lerner, Small and Loewenstein 2004; Rick, Pereira and Burson 2014) identifies that negative emotional experiences tend to increase spending, which may seem inconsistent with the current findings.

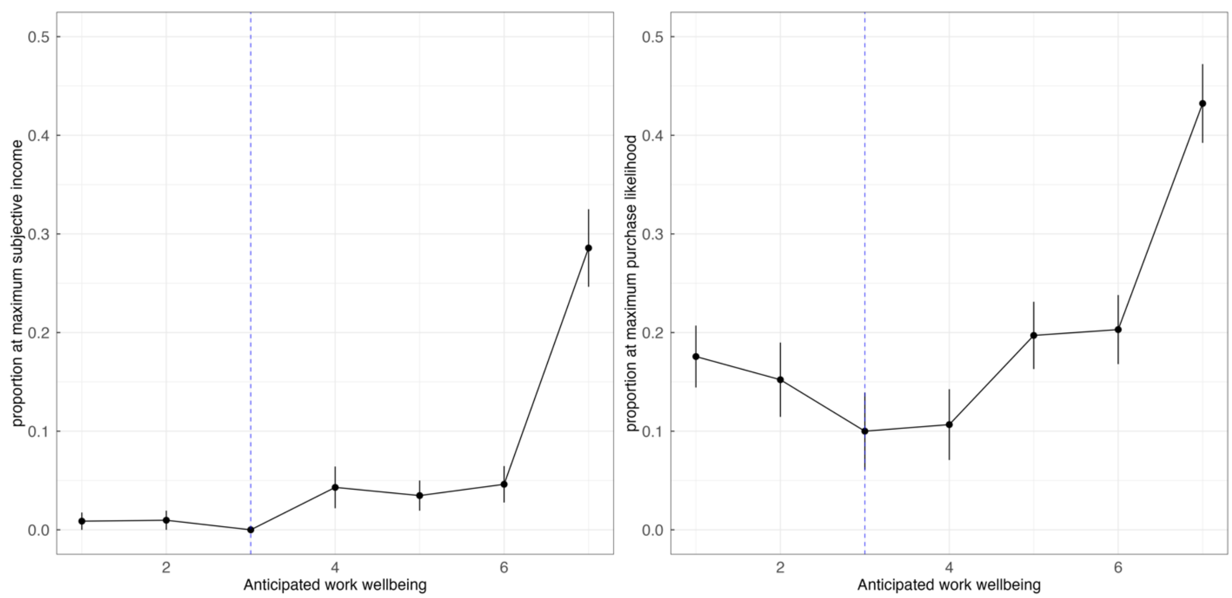
However, a deeper look underscores the compatibility with—and contributions to—the retail therapy literature.

First, the current research clarifies that domain specificity matters. While retail therapy predicts consumers attempt to regulate negative emotion through increased consumption (Atalay and Meloy 2011), retail therapy does not clarify the role of the emotional domain. The present research suggests the work domain as a potential boundary condition. When work wellbeing is low, consumers may indeed desire consumption for mood repair but may feel too impoverished to make an expenditure.

Second, retail therapy makes predictions along the negatively valenced end of the scale but is nonpredictive for positively valenced experiences. On the contrary, the present theory suggests consumers will perceive their income differently across all levels of work wellbeing. Therefore, the extent to which retail therapy versus the current theory is applicable depends in part on whether employees are generally at low or high levels of wellbeing. At low levels of wellbeing, both theories may apply; however, at high levels of wellbeing, only the current research makes a clear prediction.

Third, retail therapy and the current research both operate on spending; however, they operate through vastly different mechanisms. Retail therapy makes no prediction on subjective income. One would then expect to see a monotonic relationship between work wellbeing and subjective income (where retail therapy plays no role) and a non-monotonic relationship between work wellbeing and spending (if retail therapy encourages spending among low-wellbeing workers). An exploratory analysis of the combined data from studies 3 (including the supplement to study 3 using the PANAS scale; see appendix B) and 4 (work domain only) finds such a pattern. Figure 1.7 depicts the proportion of participants responding at the *maximum* level for subjective income (left) and purchase likelihood (right). The proportion perceiving income as very large appears to decrease monotonically as work wellbeing falls. This is to be expected. Consider how this potentially differs from the purchase likelihood plot on the right. Here, the proportion indicating they are most likely to make the discretionary purchase falls with work wellbeing, as expected. But for participants anticipating extremely low wellbeing, the pattern appears to plateau or even reverse, which is the prediction of retail therapy. Additional exploratory evidence for retail therapy is discussed in appendix C.

FIGURE 1.7: PROPORTION OF RESPONSES AT SCALE MAXIMUM IN STUDIES 3-4



NOTE—Data from studies 3, the supplement, and study 4 (work domain only); all collapsed across experimental conditions. The x-axis is anticipated work wellbeing, following the study 3-4 manipulation of imagined job changes. The left figure depicts the proportion of responses that rated the job 7/7 on the subjective income scale. The right figure depicts the proportion that indicated the maximum purchase likelihood (7/7) for the discretionary good. The dashed line at x = 3 indicates the region of very low anticipated wellbeing.

Subjective Income and Pay Satisfaction. Comparison of model results across various specifications in study 6 identifies the important role of pay satisfaction (e.g., see the specification curve in figure 1.6, where inclusion of pay satisfaction *as a control variable* greatly reduces the proportion of significant work wellbeing coefficients). Intuitively, this is consistent with the broader theorizing because pay satisfaction and subjective income both reflect—to some degree—an individual's perceptions about their income. Of course, pay satisfaction (and pay-level satisfaction) are clearly distinct from subjective income, with the most important point of distinction being that subjective income reflects a consumer's perceived adequacy of resources, whereas pay (or pay-level) satisfaction represents myriad additional organizational beliefs and is theoretically linked to organizational—not consumer—outcomes (Judge et al. 2010). Nevertheless, the points of conceptual similarity between subjective income and pay satisfaction may be a useful contribution for organizational researchers seeking to better understand holistic employee welfare.

Implications

For Marketers. The current research offers insights into how marketers might (i) identify and (ii) convert customers by strategically leveraging these findings on work wellbeing. First, considering only job-level variation in work wellbeing, marketers might choose to turn their limited attention to employees at happy firms. *Indeed* recently began collecting and publishing company-level data on work wellbeing. Marketers may leverage this data by recognizing employees at these high-wellbeing workplaces should systematically feel wealthier than their counterparts, other things equal. In terms of conversion, marketers should make efforts to make salient the high wellbeing aspects of a potential customer's work. This may resemble a just-in-time reminder of work wellbeing at pivotal moments, like when communicating with an indecisive customer. Industries that cultivate buyer-seller discourse (e.g., buying a home through realtor or an appliance through a salesperson) might benefit from shifting conversations towards not only work (a common theme of small talk), but work *wellbeing*, specifically.

For Financial Planning. Consumers constantly encounter spending decisions. Expenses big and small are ubiquitous in daily life. The current research suggests that in consumption contexts, consumers spend more if they experience high work wellbeing because they perceive their income to be larger. However, this mechanism leaves open the intriguing possibility that in other, non-consumption contexts, work wellbeing might be leveraged to increase savings. In contexts where saving or financial planning is focal, perceiving a larger income may encourage consumers to save more for the future. This highlights a potential application to increase savings, rather than discretionary spending.

Researchers and practitioners interested in increasing rates of saving might consider leveraging work wellbeing when consumers make financial plans, such as pre-committing to savings goals. For example, prior interventions have leveraged the distinct psychology of future income to increase savings through pre-commitment (Thaler and Benartzi 2004). The current research suggests one related avenue to explore is leveraging work wellbeing at the moment when consumers set their financial plans. Choice architects interested in encouraging savings might additionally consider naturally occurring within-person variation in work wellbeing (as in study 5) to align financial planning with episodes of high work wellbeing (when people perceive their income to be large). For example, this author reliably experiences work wellbeing differently during teaching and non-teaching semesters. Timing financial planning with episodes of predictably high work wellbeing might prove a useful strategy to increase pre-commitments to save, as consumers feel they have more income available. Such future directions may connect with recent explorations of the link between work evaluations and risky investing decisions (Bechler et al. 2024).

Open Questions and Future Directions

Subjective Income Updating. Work wellbeing and reservation wages both change over time (Chen et al. 2019; De Neve and Ward 2023). How closely does subjective income track these changes in work wellbeing and reservation wages? Does subjective income typically follow work wellbeing in lockstep, or does it lag behind, waiting to be updated? If so, what triggers this updating? Future research might explore whether work- or consumption-related activities are more prone to serve as the trigger. For example, is any deviation in work wellbeing reflected in subjective income, or do only sizeable departures in work wellbeing prompt such reevaluations? Alternatively, is subjective income more readily queried from the consumer side? For example, does considering a discretionary expense initiate the search for an individual's subjective wealth, in turn comparing actual income against reservation wages? These remain open questions and generative directions for future research.

Conclusion

This research contributes to a growing and important literature on how subjective financial perceptions impact consumer judgments, decisions, and behavior. (Gasiorowska 2014; Karlsson et al. 2005; De La Rosa and Tully 2022; Morewedge et al. 2007; Netemeyer et al. 2018;

Paley et al. 2019; Sharma and Alter 2012; Spiller 2011; Tully et al. 2015; Tully and Sharma 2022; Zauberman and Lynch 2005). It focuses on peoples' subjective impression of their work—specifically, work wellbeing—and explores the downstream impact on consumer behavior. This comes at a time when the very nature of work is rapidly changing, much in response to structural changes emerging from the COVID-19 pandemic, as well as the rapid adoption of artificial intelligence in the workplace. Recent years have witnessed the proliferation of work-from-home (Aksoy et al. 2022), a massive shrinking and reorganization of the workforce referred to as the "great resignation," all alongside a rapidly growing gig-economy (Statista 2024). The nature of work is quickly evolving, highlighting the growing importance and opportunity for research examining the links between work and consumption.

Chapter 1: Appendix

for

Happy Workers are “Rich” Consumers: Work Wellbeing Increases Discretionary Spending

APPENDIX A: ADDITIONAL MATERIALS FOR STUDY 1

Recent research has considered the relationship between work difficulty, risk tolerance, and investment decisions (Bechler et al. 2024). As a check that work difficulty—which is not necessarily expected to covary with work wellbeing, as discussed in the manuscript—is not driving the effects, this was measured on a 1-7 scale (“this work seems challenging” anchored on 1 = “strongly disagree” and 7 = “strongly agree”). There was no effect of condition on work challenge ($M_{\text{high wellbeing}} = 4.50$, $SD_{\text{high wellbeing}} = 1.69$, $M_{\text{low wellbeing}} = 4.50$, $SD_{\text{low wellbeing}} = 1.76$; $t(795) = 0.07$, $p = .943$).

Additional Analyses: Public Salary Request. An exploratory set of analyses consider the public salary request (the amount the participant would request when applying for the job). As expected, these salary requests were systematically higher than reservation wages ($M_{\text{salary request}}$ = \$56,751, $SD_{\text{salary request}}$ = \$17,882 vs. $M_{\text{res wage}}$ = \$49,987, $SD_{\text{res wage}}$ = \$17,193; $t(797) = 21.45$, $p < .001$). The two measures were highly correlated ($r = 0.87$), as should be expected if a prospective worker knows their reservation wages and negotiates upwards for their salary. An alternative—but unlikely—interpretation is that the prospective worker does not know their own reservation wage but is well-calibrated on how the market values work wellbeing. An *extremely conservative* test would be whether there is an effect of condition on reservation wages while controlling for the public salary requests. Even in this very restrictive model, there remains a marginally significant effect of condition (b = -1099, se = 609, t(795) = -1.80, p = .072), providing evidence work wellbeing decreases reservation wages beyond what can be explained through the salary request measure (table 1.A1).

TABLE 1.A1: CONTROLLING FOR SALARY REQUESTS (STUDY 1)

	Dependent variable:
	Reservation Wage
High Wellbeing Condition	$-1{,}098.973^{+}$ p = 0.072
Salary Request	0.832^{***} p = 0.000
Constant	$3{,}338.729^{**}$ p = 0.003
Observations	798
R^2	0.761
Adjusted R^2	0.760
Residual Std. Error	8,419.099 (df = 795)
Note:	+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

NOTE—Full regression results for a model regressing the reservation wage on the work wellbeing condition, while *controlling* for the public salary request. This is an extremely conservative model that adopts the strongest form of the alternative interpretation by attributing all of the shared variance to the salary request variable and considering only the remaining effect of condition. Despite this very high bar, there remains a marginally significant effect of condition (b = -1099, se = 609, t(795) = -1.80, p = .072).

Replication Without Reservation Wage Elicitation

One possible concern with study 2 is that the effect of the work wellbeing condition on subjective income arose only because participants were prompted to consider their reservation wage. To address such a possible concern, a replication was conducted using the exact same study design *except* participants were never asked to think about or state their public salary request or reservation wage. Therefore, this provides a more controlled test of the isolated effect of the work wellbeing manipulation on subjective income.

Method. Given the moderately large standardized effect size reported in study 1 (Cohen's $d = 0.39$), the study was run with half the prior sample size. The sample of 396 participants still provides greater than 95% power to detect an effect of similar size ($\alpha = .05$). All aspects of the study design were identical, except that participants were never prompted to consider their salary request or reservation wage.

Results. The results of the replication closely resembled those of study 1. Subjective income was larger in the high work wellbeing condition ($M = 60.5$, $SD = 24.0$) compared to the low work wellbeing condition ($M = 49.4$, $SD = 23.7$; $t(394) = 4.64$, $p < .001$, Cohen's $d = .47$). Therefore, the possible concern that the effect on subjective income was due to explicitly prompting salary demands seems highly unlikely.

APPENDIX B: ADDITIONAL MATERIALS FOR STUDY 3

Results

Objective Income Expectations. As reported in study 3 of the main manuscript, objective income expectations were 0.58 bins higher in the high work wellbeing condition, relative to the low work wellbeing condition (b = 0.58, se = 0.30, t(397) = 1.94, p = 0.053). As discussed in the manuscript footnote, this was an unexpected result based on a replication of the current study design (discussed subsequently in this appendix), which did not show an effect of condition on objective income expectations (p = .585).

Upon closer inspection, participants' actual income—also measured using the same 11 bins in $10,000 intervals—was 0.53 bins higher in the high work wellbeing condition, though this was not a significant difference (b = 0.53, se = 0.34, t(397) = 1.58, p = 0.115). Comparing the magnitude of these differences (0.58 vs. 0.53) suggests one explanation for the marginally significant difference in objective income expectations is that much of what is being estimated as the effect of condition is attributable to the random assignment of income.

To further explore this, I constructed a participant difference score comparing their objective income expectations and their actual reported income. This difference score is easily constructed (expectation – actual), as both forms of income are measured using the same 11-point scale. Regressing the difference score on work wellbeing suggests that (i) objective and actual income were not meaningfully different on average (a test of the intercept: b = 0.30, se = 0.19, t(397) = 1.61, p = .107) and that condition had no effect of changing income expectations relative to actual income (b = 0.05, se = 0.26, t(397) = 0.20, p = .844).

Preregistered Replication Using PANAS

Method

A total of 402 participants were recruited on Prolific to take part in this study. The method was the same as manuscript study 3 through the imagination exercise. Following this imagination exercise, participants were asked if they would be more or less likely to purchase the previously considered expenditure on a 1-7 purchase likelihood scale (1 = "much less likely", 7 = "much more likely"). To capture subjective income assessments, participants rated whether the income from the imagined job would feel small or large (1 = "very small", 7 = "very large").

The study included additional measures to address possible alternative explanations, as well as to provide a manipulation check. To capture objective income—which was not expected to differ by condition—participants reported the annual (pre-tax) income they expected from their job with the imagined changes from a drop-down menu with 11 income bins in $10,000 increments. Next, participants were asked "as you think about the imagined job, to what extent do you feel…?", where they rated the 10 items from the standard PANAS short form scale (Thompson 2007; Watson, Clark and Tellegen 1988) on a 1-5 scale (1 = "not at all", 5 = "very much"). These items are intended to capture positive affect (feeling alert, inspired, active, attentive, and determined) and negative affect (feeling upset, hostile, nervous, afraid, and ashamed). Finally, work wellbeing was measured on a 1-7 scale (1 = "not at all enjoyable", 7 = "very enjoyable") as a manipulation check.

Results

All 402 participants were included for analysis (M_{age} = 40.4, 49% female).

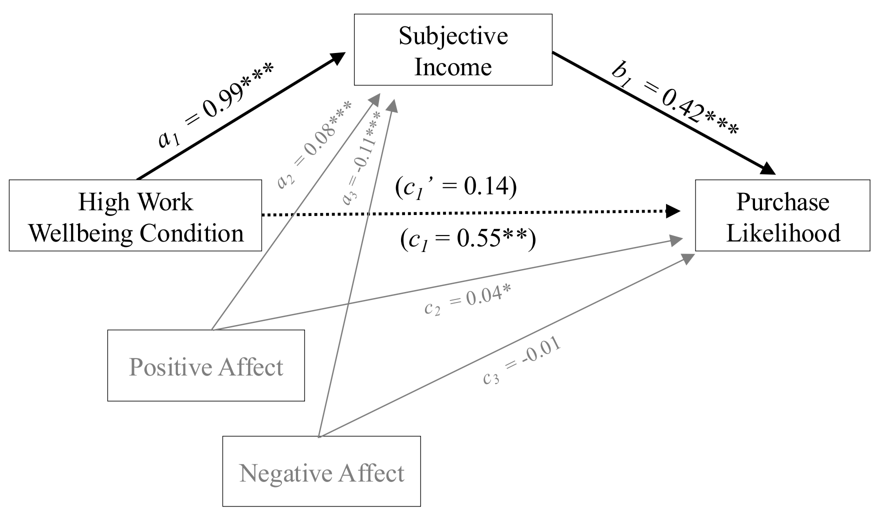
Manipulation and Confound Checks. The manipulation check identified a strong effect of condition on measured work wellbeing ($M_{\text{high wellbeing}} = 5.36$, $SD_{\text{high wellbeing}} = 1.57$, $M_{\text{low wellbeing}} = 2.58$, $SD_{\text{low wellbeing}} = 1.72$; $t(400) = 16.93$, $p < .001$). The study also included two additional sets of measures to address alternative explanations: objective income expectations and measured affect. There was no effect of condition on objective income expectations ($M_{\text{high wellbeing}} = 5.62$, $SD_{\text{high wellbeing}} = 2.83$, $M_{\text{low wellbeing}} = 5.47$, $SD_{\text{low wellbeing}} = 2.75$; $t(400) = 0.55$, $p = .585$). However, there was an effect of condition on both positive and negative affect, as measured using the PANAS short-form (Thompson 2007). Positive affect[8] ($\alpha = 0.87$) was greater in the high work wellbeing condition ($M_{\text{high wellbeing}} = 15.89$, $SD_{\text{high wellbeing}} = 5.55$) compared to the low work wellbeing condition ($M_{\text{low wellbeing}} = 13.20$, $SD_{\text{low wellbeing}} = 4.76$; $t(400) = 5.21$, $p < .001$). Negative affect ($\alpha = 0.84$) was lower in the high work wellbeing condition ($M_{\text{high wellbeing}} = 6.83$, $SD_{\text{high wellbeing}} = 3.00$), compared to the low work wellbeing condition ($M_{\text{low wellbeing}} = 10.22$, $SD_{\text{low wellbeing}} = 4.51$; $t(400) = -8.91$, $p < .001$). This suggests that while objective income expectations are unlikely to explain prior results, affect-based alternative explanations might. Therefore, all key analyses are reconsidered with controls for both positive and negative affect.

[8] As in Thompson (2007), the positive and negative affect scores are calculated as the sum of the set of 5 items.

Preregistered Analyses. Consistent with H3, there was an effect of condition on purchase likelihood. Participants in the high work wellbeing condition indicated they were more likely to make their previously identified purchase (M = 5.30, SD = 1.48) than those in the low work wellbeing condition (M = 4.38, SD = 1.89; $t(400) = 5.46$, $p < .001$, Cohen's d = 0.54). Furthermore, there was the expected effect on subjective income assessments, such that the imagined income was rated as larger in the high wellbeing condition ($M_{\text{high wellbeing}} = 4.62$, $SD_{\text{high wellbeing}} = 1.50$, $M_{\text{low wellbeing}} = 3.04$, $SD_{\text{low wellbeing}} = 1.67$; $t(400) = 9.99$, $p < .001$, Cohen's d = 1.00). According to H4, the effect of work wellbeing on purchase likelihood should operate through subjective income. To test this process, I conducted a mediation analysis with a bootstrapped 95% confidence interval around the indirect effect (5,000 resamples). Results are consistent with indirect-only mediation, with a significant indirect effect through subjective income assessments (ab = 0.72, $CI_{95\%}$ = [0.51, 0.95]) and no remaining direct effect of condition (c' = 0.21, p = .231).

Analyses Controlling for Affect. Reconsidering the preregistered analyses with controls for affect does not meaningfully change the conclusions of this study. The effects of the work wellbeing condition on purchase likelihood and subjective income remain sizeable and significant, though both are attenuated in absolute terms (purchase likelihood: b = 0.55, $t(398)$ = 3.02, p = .003, Cohen's d = 0.27; subjective income: b = 0.99, $t(398)$ = 6.06, $p < .001$, Cohen's d = 0.54). These relationships are depicted in figure 1.B1, which presents a full mediation path diagram for the models controlling for affect. Controlling for positive and negative affect, there remains a significant indirect effect of condition on purchase likelihood through subjective income (ab = 0.41, $CI_{95\%}$ = [0.25, 0.60]). While these analyses provide evidence for an effect of work wellbeing above and beyond the role of affect, it should be emphasized that affect clearly plays a role. But after partialing out this affective role, there remains a unique effect of work wellbeing on all outcomes of interest.

FIGURE 1.B1: MEDIATION PATH DIAGRAM FOR REPLICATION TO STUDY 3 (PANAS)



NOTE—The preregistered mediation pathways with estimates and significance levels. There is a significant indirect effect ($ab = 0.41$, $CI_{95\%} = [0.25, 0.60]$).

APPENDIX C: ADDITIONAL MATERIALS FOR STUDY 4

Method

Manipulation for Volunteering Domain. The guided imagination exercise wording for the *work* domain is provided in manuscript study 3. The wording for the leisure domain is provided in manuscript study 4. The wording for the volunteering domain (produced below and in the survey materials on ResearchBox) is nearly identical to the work domain. The key difference is that words like "work" are replaced with "experience."

High wellbeing

> Consider the aspects / tasks of this experience you most liked, and imagine they are your entire experience. Specifically, please imagine your whole volunteer experience is comprised of your most liked tasks. Everything else (e.g., hours, people, location) remains unchanged. The only difference with this scenario is that you get to spend all your time on the most enjoyable aspects of volunteering.

Low wellbeing

> Consider the aspects / tasks of this experience you most disliked, and imagine they are your entire experience. Specifically, please imagine your whole volunteer experience is comprised of your most disliked tasks. Everything else (e.g., hours, people, location) remains unchanged. The only difference with this scenario is that you have to spend all your time on the most unenjoyable aspects of volunteering.

Additional Measures. As another set of measures that might be less work specific and more robust across domains, I also measured domain-specific positive and negative affect. Specifically, participants indicated the extent to which the situation they imagined was associated with [positive/negative] feelings, anchored on 1 = "not at all associated with [positive/negative] feelings” and 7 = “very strongly associated with [positive/negative] feelings.” To clarify, these affect measures are distinct from those used in study 3. Whereas study 3 measured a participant’s current affect (to address concerns about affective spillover), study 4 measures whether the domain (work, volunteering, leisure) is associated with positive or negative affect (e.g., the affective component of work wellbeing, in the work domain).

Results

The primary and preregistered results are discussed in the main manuscript. Table 1.C1 includes the full regression output from those analyses.

TABLE 1.C1: FULL REGRESSION OUTPUT FROM MANUSCRIPT ANALYSIS

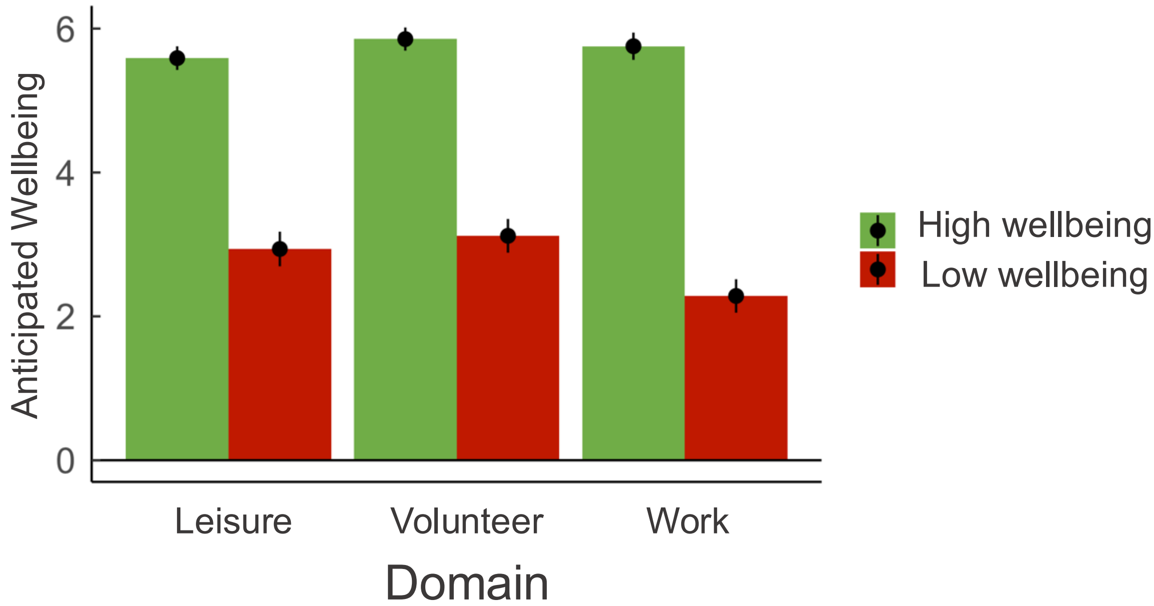
	Dependent variable:
	Purchase likelihood
High Wellbeing	0.40^{***} (0.07)
Volunteer	-0.33^{***} (0.10)
Leisure	-0.48^{***} (0.10)
High x Volunteer	-0.38^{***} (0.10)
High x TV	-0.27^{**} (0.10)
Constant	4.67^{***} (0.07)
Observations	1,201
F Statistic	12.28^{***} (df = 5; 1195)
Note:	*p<0.05;**p<0.01;***p<0.001

NOTE—Wellbeing was contrast coded (+1 = high, -1 = low), and Volunteer and Leisure were both dummy coded.

Robustness

Controlling for Measured Work Wellbeing. Because there were differences in measures of anticipated wellbeing across the three domains (figure 1.C1), it is important to consider whether the preregistered effects hold when controlling for such differences. This is achieved by adding the measured wellbeing variable to the main model (table 1.C2). In this new model, there is no longer a simple effect of wellbeing condition for participants in the work condition (b = 0.073, se = 0.082, t(1194) = 0.89, p = .374). This is to be expected, because controlling for the measured wellbeing explains the variance in purchase likelihood that was previously attributable to condition. The focal tests however are the tests of the two interactions. These interactions remain significant ($b_{\text{volunteer x wellbeing}}$ = -0.309, se = 0.097, t(1194) = -3.19, p = .001, Cohen's f = 0.07; $b_{\text{leisure x wellbeing}}$ = -0.200, se = 0.097, t(1194) = -2.02, p = .043, Cohen's f = 0.06). Therefore, this model suggests there remains a unique effect of work wellbeing (as opposed to non-work wellbeing), even after adjusting for differences in levels of anticipated wellbeing.

FIGURE 1.C1: ANTICIPATED WELLBEING



NOTE—Error bars represent 95% confidence intervals around the estimates of the group mean.

Controlling for Measured Affect. There were differences in measured affect across domains. The effect of the wellbeing manipulation on positive affect in the work domain was attenuated in both the volunteer ($b_{\text{volunteer x wellbeing}}$ = -0.203, *se* = 0.106, *t*(1195) = -1.92, *p* = .055) and leisure domains ($b_{\text{leisure x wellbeing}}$ = -0.385, *se* = 0.105, *t*(1195) = -3.65, *p* < .001). The effect of the wellbeing manipulation to reduce negative affect for work was also attenuated in both the volunteer ($b_{\text{volunteer x wellbeing}}$ = 0.187, *se* = 0.107, *t*(1195) = 1.76, *p* = .079) and leisure conditions ($b_{\text{leisure x wellbeing}}$ = 0.261, *se* = 0.106, *t*(1195) = 2.45, *p* = .014). Therefore, I reconsider the preregistered purchase likelihood regression including the positive and negative affect as two additional controls (table 1.C2). The two interactions remain significant ($b_{\text{volunteer x wellbeing}}$ = -0.343, *se* = 0.096, *t*(1193) = -3.56, *p* < .001, *f* = 0.08; $b_{\text{leisure x wellbeing}}$ = -0.200, *se* = 0.096, *t*(1193) = -2.03, *p* = .042, *f* = 0.06). Therefore, this model suggests there remains a unique effect of work wellbeing (as opposed to non-work wellbeing), even after adjusting for differences in the affective component of wellbeing.

TABLE 1.C2: ROBUSTNESS CHECKS

	Dependent variable:	
	Purchase likelihood	
Wellbeing Con	0.07 (0.08)	0.11 (0.08)
Volunteer	-0.42^{***} (0.10)	-0.38^{***} (0.10)
Leisure	-0.52^{***} (0.10)	-0.50^{***} (0.10)
Measured Wellbeing	0.19^{***} (0.03)	
Positive Affect		0.26^{***} (0.04)
Negative Affect		0.09^{*} (0.04)
Wellbeing Con x Volunteer	-0.31^{**} (0.10)	-0.34^{***} (0.10)
Wellbeing Con x Leisure	-0.20^{*} (0.10)	-0.20^{*} (0.10)
Constant	3.92^{***} (0.13)	3.26^{***} (0.31)
Observations	1,201	1,201
F Statistic	19.03^{***} (df = 6; 1194)	17.63^{***} (df = 7; 1193)
Note:		*p<0.05;**p<0.01;***p<0.001

NOTE—Two regressions corresponding to the two robustness checks. The first model controls for measured wellbeing (in addition to the contrast-coded condition variable for condition, “Wellbeing Con”). The second model controls for positive and negative affect.

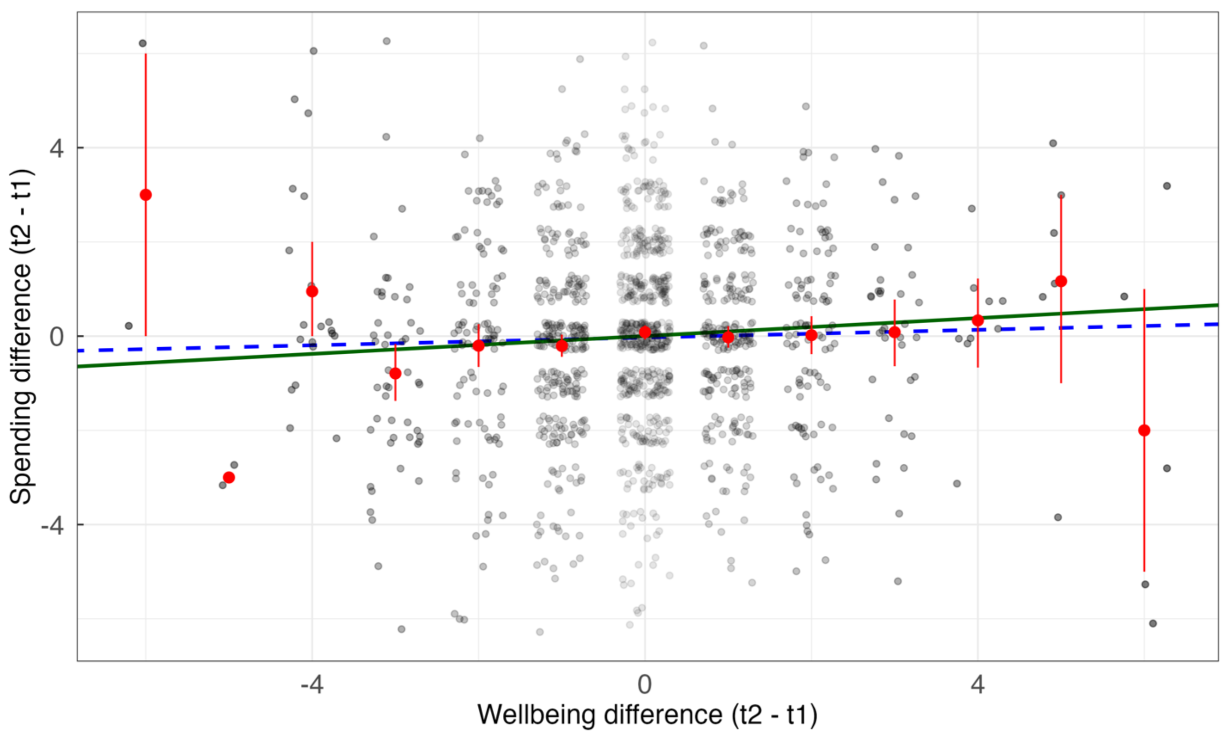
Exploratory Analyses

Exploratory Analyses of Affect with Connections to Retail Therapy. To better understand the role of affect, I regressed purchase likelihood on positive affect, negative affect, the two condition dummy variables, and the two-way interactions between affect and condition. In the work condition, the positive affect coefficient was significantly positive (b = 0.384, se = 0.073, t(1192) = 5.25, p < .001). Interestingly, the negative affect coefficient was also marginally positive (b = 0.131, se = 0.071, t(1192), p = .068), which is perhaps consistent with the theoretically opposing role of retail therapy. The coefficient on positive affect is significantly larger than the negative affect coefficient (F(1,1192) = 73.54, p < .001). Furthermore, this relationship between positive affect and greater purchase likelihood in the work condition appears to be significantly or directionally attenuated in both of the non-work conditions ($b_{\text{Positive Affect x Volunteer}}$ = -0.228, *se* = 0.101, $t(1192) = -2.26$, $p = .023$; $b_{\text{Positive Affect x Leisure}}$ = -0.158, *se* = 0.096, $t(1192) = -1.64$, $p = .102$). There is no evidence the role of negative affect depends on situation (smallest $p = .68$). This pattern of results is consistent with the explanation that positive affect has the greatest impact on purchase likelihood in the work situation (where it may operate on subjective income assessments), whereas negative affect may also have a positive impact on purchasing (retail therapy) that does not differ between work vs. non-work.

APPENDIX D: ADDITIONAL MATERIALS FOR STUDY 5

As discussed in the manuscript, the coefficient on the difference in work wellbeing was not significant using OLS regression. A quick visual inspection of the data suggests something potentially unusual: The pattern that visually emerges for the majority of the data (e.g., difference scores in [-5, 5]) appears to be a positive correlation between the work and spending difference scores. However, the pattern appears to flip—and drastically so—for the most extreme difference scores. This is visualized in figure 1.D1.

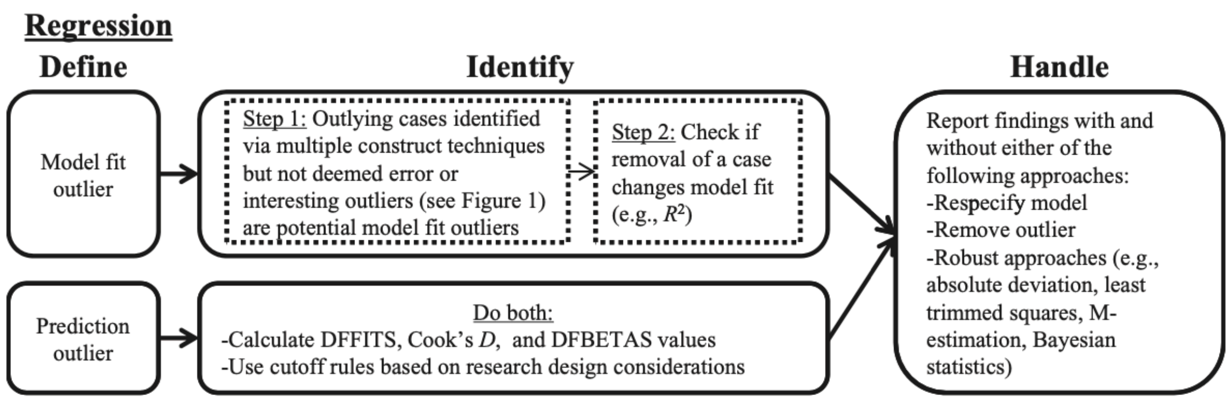
FIGURE 1.D1: ESTIMATES OF SLOPES WITH AND WITHOUT ROBUST REGRESSION



NOTE—A jittered scatterplot of the raw data includes mean spending differences (with 95% confidence intervals) at each level of wellbeing difference. The blue dashed line shows the estimated regression slope using an OLS model. The green solid line shows the estimated regression slope using a robust regression (MM-estimation), to account for any prediction outliers.

Following the exact recommendations of Aguinis, Gottfredson and Joo (2013), who meta-analyze and define best-practices for the treatment of outliers in regression analysis, I conduct and report on (i) an alternative model specification; (ii) removal of outliers based on analyses unrelated to the focal hypothesis (i.e., DFFITS, Cook's Distance, DFBETAS, and Studentized Deleted Residuals); and (iii) use robust regressions that will be less sensitive to the presence of outliers (i.e., least-trimmed squares, M- and MM-estimations). These are summarized in figure 1.D2.

FIGURE 1.D2: BEST PRACTICES FOR OUTLIERS, FROM AGUINIS ET AL. (2013)



NOTE—Portion of "Figure 2" (pp. 289) specifically refers to instructions for addressing outliers in regression models (Aguinis et al. 2013).

Alternative Model Specification. Instead of asking "how does the difference in work wellbeing correspond to the difference in spending?", an alternative question is "do people who experience more work wellbeing also spend more?" This translation involves discretizing the difference scores to reflect whether the *wave 2* measure was higher (+1), lower (-1), or no difference (0), relative to *wave 1*. The benefit of this approach is that observations will have more equitable "votes" on the slope of the regression line, whereas in the raw data an observation at +6 would exert substantially more leverage on the "slope vote" than an observation near the data mean (0). The downside of this approach is that meaningful variation—the degree of the difference—is lost in an effort to only estimate the directionality of the relationship. Three different approaches to discretizing the difference scores involve transforming X (work difference), Y (spending difference), or both. Table 1.D1 presents the results of these three alternative specifications.

TABLE 1.D1: ALTERNATIVE SPECIFICATIONS

	transformation	df	b	p	removed
1	Discrete X	1437	0.115	0.099	0
2	Discrete Y	1437	0.040	0.011	0
3	Discrete XY	1437	0.067	0.024	0

NOTE—Three alternative specifications of the model with either X (difference in work wellbeing), Y (difference in discretionary spending), or both are discretized as -1 if < 0; 0 if 0; +1 if > 0.

Identification and Removal of Outliers. The second approach recommended in Aguinis et al. (2013) is to identify prediction outliers and remove them based on established cutoff rules. Table 1.D2 presents four outlier metrics (studentized deleted residuals, DFFITS, DFBETA, and Cook's Distance), along with the model predictions following the recommended cutoffs.

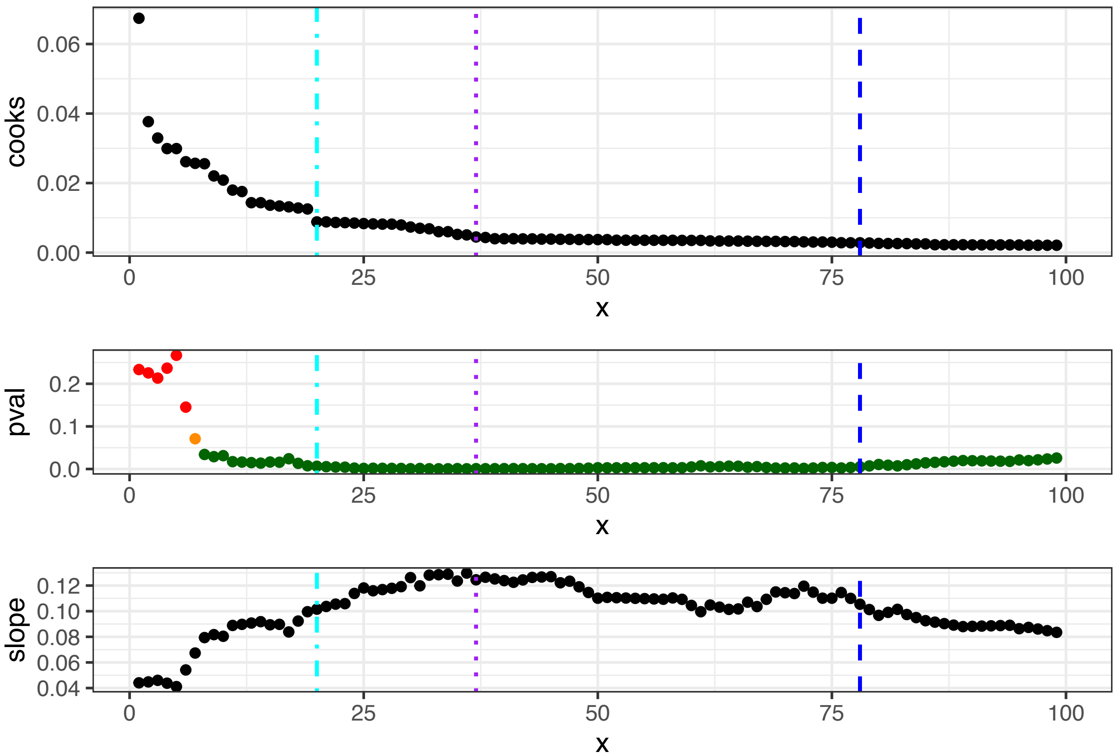
TABLE 1.D2: OUTLIER DETECTION AND REMOVAL

	metric	cutoff	df	b	p	removed
1	Studentized Deleted Residual	3	1430	0.083	0.024	7
2	Studentized Deleted Residual	2.5	1403	0.108	0.002	34
3	Studentized Deleted Residual	2	1353	0.135	0.000	84
4	DFFITS	$\pm 2\sqrt{\frac{k+1}{n}}$	1375	0.105	0.005	62
5	DFBETA	$\pm 2\sqrt{n}$	1344	0.097	0.016	93
6	Cook's Distance	$\frac{4}{n}$	1359	0.105	0.005	78

NOTE—Identification and removal of outliers based on various metrics with established cutoffs. Subsequent regressions were run using each outlier metric and the associated cutoff. For each new model, table 1.D2 reports degrees of freedom (df), the estimated slope (b), the *p*-value for the estimated slope (p), and the number of removed observations.

Of course, these cutoff values are somewhat arbitrary, and this may be especially true for cutoffs that are not based on distributional characteristics (e.g., studentized deleted residuals) but rather sample size. To this end, an additional suggestion in Aguinis et al. (2013) is to consider cutoff values where there are breaks in the data. To assist in this approach, figure 1.D3 plots Cook's Distance, the model *p*-value, and the estimated slope (*b*) for the 100 highest-impact observations. Figure 1.D3 marks two additional points at which there are visual breaks in Cook's Distance (the teal dot-dash line, and the purple dotted line), which could serve as alternative cutoff values, rather than (4/n: the blue dashed line).

FIGURE 1.D3: VISUALIZATION OF COOK'S *D*, *P*, AND *B* ACROSS RANKED OBSERVATIONS



NOTE—Cooks' Distance (top panel), *p*-value for a model excluding all observations with a Cook's Distance equal to or larger than the current observation (middle panel), and the corresponding estimate of the slope (bottom panel) for the 100 most impactful observation, ordered along the x-axis. The conventional exclusion threshold of (4/n) is marked with a blue dashed line. Alternatively, one could consider cutoffs based on breaks in Cook's D, such as around the dotted purple line or the dot-dashed teal line.

Robust Regression Techniques. A third approach involves using regression techniques that are designed to be more robust to influential outliers. The feasible methods suggested by Aguinis et al. (2013) are reported in table 1.D3. Conceptually, the least trimmed squares regression resembles the outlier removal methods already discussed, as the estimator drops highly influential observations (as reflected in the reduced degrees of freedom). M-estimation and MM-estimations are a class of robust regressions that are often used when examining longitudinal data (Aguinis et al. 2013). These approaches use a loss function for the residuals (as opposed to operating on the sum of squared residuals, as in OLS) making them more robust to influential outliers. In addition to being frequently used for longitudinal data, this approach may be preferred because it does not drop any data, while preserving an interpretable estimate of the coefficient in the model that also includes the additional day-of-week controls. For completeness, I also consider a Bayesian estimator using a flat prior, though this method appears to be less consistent with the other three techniques.

TABLE 1.D3: ALTERNATIVE ESTIMATORS

	name	df	b	inference	removed
1	Least Trimmed Squares	1382	0.119	0	55
2	Robust Regression (M)	1437	0.091	0.009	0
3	Robust Regression (MM)	1437	0.095	0.009	0
4	Bayesian (flat prior)		0.042	87.4%	0

NOTE—Alternative estimator models. Inference refers to the *p*-value for rows 1-3 and the posterior probability of a positive coefficient (row 4).

APPENDIX E: ADDITIONAL MATERIALS FOR STUDY 6

No additional robustness is presented here, given the complete access to all models via the online specification table (https://specificationtable.shinyapps.io/shiny/).

Additional Details on Measures from *Survey of Consumer Expectations* Labor Market Survey.

The following questions were used in the Labor Market Survey. Items Q1 (pay satisfaction), Q4 (promotion opportunities), and Q5 (overall job satisfaction) were excluded as pertaining either to compensation (Q1, Q4) or composite job satisfaction (Q5). Q2 (satisfaction with non-financial aspects) and Q3 (person-job-fit) were combined to measure work enjoyment (though all conclusions hold using either measure in isolation). Q2 was measured on a 1-5 satisfaction scale, which was standardized to a 7-point scale to provide equal weight with Q3.

Q1: "How satisfied would you say you are with your level of compensation at your [current/main] job?"

*Q2: "And how satisfied would you say you are with other aspects of the job, such as benefits, maternity/paternity leaves, flexibility in work hours, etc?"

*Q3: "On a scale from 1 to 7, how well do you think this job fits your experience and skills?"

Q4: "On a scale from 1 to 7, how would you rate the opportunities for a promotion or other career progression with your current employer, over the next three years?"

Q5: "Taking everything into consideration, how satisfied would you say you are, overall, in your [current/main] job?"

Spending Category Question Wordings

The following question was presented in the Household spending survey.

Approximately what proportion of your current monthly household spending falls in each of the following categories? Please enter a number between 0 and 100 for each category. The numbers need to add up to 100%.

- *Housing (including mortgage, rent, maintenance and home owner/renter insurance) (1)*
- *Utilities (including water, sewer, electricity, gas, heating oil) (2)*
- *Food (including groceries, dining out, and beverages) (3)*
- *Clothing, footwear and personal care (4)*
- *Transportation (including gasoline, public transportation fares, and car maintenance) (5)*
- *Medical care (including health insurance, medical bills, prescription drugs) (6)*
- *Recreation and entertainment (7)*
- *Education and child care (8)*
- *Other (including gifts, child support or alimony, charitable giving, and other miscellaneous) (9)*

APPENDIX F: ATTRITION

This section presents the contingency tables examining complete and incomplete responses by condition. Analyses of attrition may be especially important for the current research if wellbeing manipulations lead to differential drop-out rates. This does not appear to be the case. All studies with manipulated work wellbeing are analyzed below.

Study 1

There was no evidence of differential attrition in study 1 based on the number of complete and incomplete responses, as shown by table 1.F1 ($\chi^2(1) = 1.36$, $p = .381$).

TABLE 1.F1: COMPLETE AND INCOMPLETE RESPONSES IN STUDY 1

	Complete	Incomplete
High wellbeing	402	4
Low wellbeing	397	8

Study 2

There was clear evidence of differential attrition in study 2 based on the number of complete and incomplete responses, as shown by table 1.F2 ($\chi^2(1) = 52.28$, $p < .001$).

TABLE 1.F2: COMPLETE AND INCOMPLETE RESPONSES IN STUDY 2

	Complete	Incomplete
High wellbeing	281	8
Low wellbeing	219	67

This is not surprising, given (1) the low-wellbeing CAPTCHA manipulation was glitchy (by design) and (2) the low-wellbeing manipulation was designed to be boring, redundant, and annoying. Are these concerning? I would argue not. Many participants likely left in response to believing the survey was broken, as evidenced by several participant messages, such as the one included in the main manuscript.

Additionally and most importantly: This differential attrition that is due to differences in experienced wellbeing *should conservatively bias the estimates*. Consider, if low wellbeing leads to drop out, then it should be true that participants who drop out are more likely to experience very low wellbeing, relative to those who do not drop out. If this is the case, then those participants who left *should have provided the lowest scores on subjective income*, in expectation*, because they experienced the lowest wellbeing*, in expectation. Therefore, the selective removal of these individuals also selectively removes observations of subjective income that are likely to skew lower than the observed distribution. Therefore, the direction of differential attrition should reduce the difference between the two experimental conditions (thus reducing the chance of a false positive inference, not increasing it).

Study 3

There was no evidence of differential attrition in study 3 based on the number of complete and incomplete responses, as shown by table 1.F3 ($\chi^2(1) = 0.00$, $p = 1$)

TABLE 1.F3: COMPLETE AND INCOMPLETE RESPONSES IN STUDY 3

	Complete	Incomplete
High work wellbeing	199	12
Low work wellbeing	200	12

Study 4

There was no evidence of differential attrition in study 4 based on the number of complete and incomplete responses, as shown by table 1.F4 ($\chi^2(5) = 3.25$, $p = .662$).

TABLE 1.F4: COMPLETE AND INCOMPLETE RESPONSES IN STUDY 4

	Complete	Incomplete
High wellbeing (Work)	200	6
High wellbeing (Volunteer)	199	7
High wellbeing (Leisure)	203	3
Low wellbeing (Work)	201	4
Low wellbeing (Volunteer)	198	8
Low wellbeing (Leisure)	200	5

APPENDIX REFERENCES

Aguinis, Herman, Ryan K. Gottfredson, and Harry Joo (2013), “Best-Practice Recommendations for Defining, Identifying, and Handling Outliers,” *Organizational Research Methods*, 16(2), 270–301.

Bechler, Christopher J., Samina Lutfeali, Szu-chi Huang, and Joshua I. Morris (2024), “Working Hard for Money Decreases Risk Tolerance,” *Journal of Consumer Psychology*, 34(1), 110–18.

Thompson, Edmund R. (2007), “Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS),” *Journal of Cross-Cultural Psychology*, 38(2), 227–42.

Watson, David, Lee A. Clark, and Auke Tellegen (1988), “Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales,” *Journal of Personality and Social Psychology*, 54(6), 1063.

CHAPTER 1 REFERENCES

Aguinis, Herman, Ryan K. Gottfredson, and Harry Joo (2013), "Best-Practice Recommendations for Defining, Identifying, and Handling Outliers," *Organizational Research Methods*, 16(2), 270–301.

Ahn, Namkee, Victoria Ateca-Amestoy, and Arantza Ugidos (2014), "Financial Satisfaction from an Intra-Household Perspective," *Journal of Happiness Studies*, 15(5), 1109–23.

Aksoy, Cevat G., Jose M. Barrero, Nicholas Bloom, Steven J. Davis, Mathias Dolls, and Pablo Zarate (2022), "Working from Home Around the World," *Working Paper*, w30446.

André, Quentin and Nicholas Reinholtz (2024), "Pre-Registered Interim Analysis Designs (PRIADs): Increasing the Cost-Effectiveness of Hypothesis Testing," *Journal of Consumer Research*.

Ariely, Dan, Emir Kamenica, and Dražen Prelec (2008), "Man's Search for Meaning: The Case of Legos," *Journal of Economic Behavior and Organization*, 67(3–4), 671–77.

Astebro, Thomas, Holger Herz, Ramana Nanda, and Roberto A. Weber (2014), "Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics," *Journal of Economic Perspectives*, 28(3), 49–69.

Atalay, A. Selin and Margaret G. Meloy (2011), "Retail Therapy: A Strategic Effort to Improve Mood," *Psychology & Marketing*, 28(6), 638–60.

Aziri, Brikend (2011), "Job Satisfaction: A Literature Review," *Management Research and Practice*, 3(4), 77–86.

Bechler, Christopher J., Samina Lutfeali, Szu-chi Huang, and Joshua I. Morris (2024), "Working Hard for Money Decreases Risk Tolerance," *Journal of Consumer Psychology*, 34(1), 110–18.

Bellet, Clément S., Jan-Emmanuel De Neve, and George Ward (2023), "Does Employee Happiness Have an Impact on Productivity?," *Management Science*, 70(3), 1656–79.

Bhattacharjee, Amit and Cassie Mogilner (2014), "Happiness from Ordinary and Extraordinary Experiences," *Journal of Consumer Research*, 41(1), 1–17.

Bowling, Nathan A., Kevin J. Eschleman, and Qiang Wang (2010), "A Meta-Analytic Examination of the Relationship between Job Satisfaction and Subjective Well-Being," *Journal of Occupational and Organizational Psychology*, 83(4), 915–34.

Burbano, Vanessa C. (2016), "Social Responsibility Messages and Worker Wage Requirements: Field Experimental Evidence from Online Labor Marketplaces," *Organization Science*, 27(4), 1010–28.

Bureau of Labor Statistics (2024), *American Time Use Survey - 2023 Results*.

Cassar, Lea and Stephan Meier (2018), "Nonmonetary Incentives and the Implications of Work as a Source of Meaning," *Journal of Economic Perspectives*, 32(3), 215–38.

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.

Chen, Kuan-Ming, Claire Ding, John A. List, and Magne Mogstad (2020), "Reservation Wages and Workers' Valuation of Job Flexibility: Evidence from a Natural Field Experiment," *Working Paper,* w27807.

Chen, M. Keith, Judith A. Chevalier, Peter E. Rossi, and Emily Oehlsen (2019), "The Value of Flexible Work: Evidence from Uber Drivers," *Journal of Political Economy*, 127(6), 2735–2794.

Clark, Andrew E. and Andrew J. Oswald (1996), "Satisfaction and Comparison Income," *Journal of Public Economics*, 61(3), 359–81.

Cryder, Cynthia E., Jennifer S. Lerner, James J. Gross, and Ronald E. Dahl (2008), "Misery Is Not Miserly: Sad and Self-Focused Individuals Spend More," *Psychological Science*, 19(6), 525–30.

Cunningham, Sarah, William Fleming, Cherise Regier, Micah Kaats, and Jan-Emmanuel De Neve (2024), *Work Wellbeing Playbook: A Systematic Review of Evidence-Based Interventions to Improve Employee Wellbeing*. World Wellbeing Movement.

Davis, Wayne (1982), "A Causal Theory of Enjoyment," *Mind*, 91(362), 240–56.

De La Rosa, Wendy and Stephanie M. Tully (2022), "The Impact of Payment Frequency on Consumer Spending and Subjective Wealth Perceptions," *Journal of Consumer Research*, 48(6), 991–1009.

De Neve, Jan-Emmanuel, Micah Kaats, and George Ward (2023), *Workplace Wellbeing and Firm Performance*. World Wellbeing Movement.

De Neve, Jan-Emmanuel and George Ward (2023), *Measuring Workplace Wellbeing*. World Wellbeing Movement.

Dias, Rodrigo, Eesha Sharma, and Gavan J. Fitzsimons (2022), "Spending and Happiness: The Role of Perceived Financial Constraints," *Journal of Consumer Research*, 49(3), 373–88.

Diener, Ed and Katherine Ryan (2009), "Subjective Well-Being: A General Overview," *South African Journal of Psychology*, 39(4), 391–406.

Diener, Ed, Christie N. Scollon, and Richard E. Lucas (2009), "The Evolving Concept of Subjective Well-Being: The Multifaceted Nature of Happiness," *Assessing well-being: The collected works of Ed Diener*, 67-100.

Diener, Ed, Eunkook M. Suh, Richard E. Lucas, and Heidi L. Smith (1999), "Subjective Well-Being: Three Decades of Progress," *Psychological Bulletin*, 125(2), 276–302.

Dik, Bryan J., Zinta S. Byrne, and Michael F. Steger (2013), *Purpose and Meaning in the Workplace*, American Psychological Association.

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

Federal Reserve Bank of New York (2024), *Survey of Consumer Expectations.*

Fernbach, Philip M., Christina Kan, and John G. Lynch (2015), "Squeezed: Coping with Constraint through Efficiency and Prioritization," *Journal of Consumer Research*, 41(5), 1204–27.

Friedman, Milton (1957), *A Theory of the Consumption Function*, Princeton, NJ: Princeton University Press.

Gasiorowska, Agata (2014), "The Relationship between Objective and Subjective Wealth Is Moderated by Financial Control and Mediated by Money Anxiety," *Journal of Economic Psychology*, 43, 64–74.

Goldstone, Robert L. (1994), "The Role of Similarity in Categorization: Providing a Groundwork," *Cognition*, 52(2), 125–57.

Haisley, Emily, Romel Mostafa, and George Loewenstein (2008), "Subjective Relative Income and Lottery Ticket Purchases," *Journal of Behavioral Decision Making*, 21(3), 283–95.

Hamilton, Rebecca W., Chiraag Mittal, Anuj Shah, Debora V. Thompson, and Vladas Griskevicius (2019), "How Financial Constraints Influence Consumer Behavior: An Integrative Framework," *Journal of Consumer Psychology*, 29(2), 285–305.

Indeed (2024), "*Global Work Wellbeing Report 2024.*"

Judge, Timothy A., Charles L Hulin, and Reeshad S. Dalal (2012), "Job Satisfaction and Job Affect," in *The Oxford Handbook of Industrial and Organizational Psychology*, ed. S.W.J. Kozlowski, New York, NY: Oxford University Press, 496–525.

Judge, Timothy A., Ronald F. Piccolo, Nathan P. Podsakoff, John C. Shaw, and Bruce L. Rich (2010), "The Relationship between Pay and Job Satisfaction: A Meta-Analysis of the Literature," *Journal of Vocational Behavior*, 77(2), 157–67.

Kahneman, Daniel and Alan B. Krueger (2006), "Developments in the Measurement of Subjective Well-Being," *Journal of Economic Perspectives*, 20(1), 3–24.

Karlsson, Niklas, Tommy Gärling, Peter Dellgrain, and Birgitta Klingander (2005), "Social Comparison and Consumer Behavior: When Feeling Richer or Poorer Than Others Is More Important Than Being So," *Journal of Applied Social Psychology*, 35(6), 1206–22.

Leete, Laura (2001), "Whither the Nonprofit Wage Differential? Estimates from the 1990 Census," *Journal of Labor Economics*, 19(1), 136–70.

Lerner, Jennifer S, Deborah A. Small, and George Loewenstein (2004), "Heart Strings and Purse Strings: Carryover Effects of Emotions on Economic Decisions," *Psychological science*, 15(5), 337–41.

Loewenstein, George E. and Drazen Prelec (1993), "Preferences for Sequences of Outcomes," *Psychological Review*, 100(1), 91–108.

Lyubomirsky, Sonja (2001), “Why Are Some People Happier than Others? The Role of Cognitive and Motivational Processes in Well-Being,” *American Psychologist*, 56(3), 239.

Mishra, Himanshu, Arul Mishra, and Dhananjay Nayakankuppam (2006), “Money: A Bias for the Whole,” *Journal of Consumer Research*, 32, 541–49.

Mogilner, Cassie, Sepandar D. Kamvar, and Jennifer Aaker (2011), “The Shifting Meaning of Happiness,” *Social Psychological and Personality Science*, 2(4), 395–402.

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.

Naber, Andrew (2017), *One Third of Your Life Is Spent at Work.*

Oh, Travis T. and Michel T. Pham (2022), “A Liberating-Engagement Theory of Consumer Fun,” *Journal of Consumer Research*, 49(1), 46–73.

Paley, Anna, Stephanie M. Tully, and Eesha Sharma (2019), “Too Constrained to Converse: The Effect of Financial Constraints on Word of Mouth,” *Journal of Consumer Research*, 45(5), 889–905.

Rick, Scott I., Beatriz Pereira, and Katherine A. Burson (2014), “The Benefits of Retail Therapy: Making Purchase Decisions Reduces Residual Sadness,” *Journal of Consumer Psychology*, 24(3), 373–80.

Riffkin, Rebecca (2014), *In U.S., 55% of Workers Get Sense of Identity From Their Job*.

Rosen, Sherwin (1986), “The Theory of Equalizing Differences,” *Handbook of Labor Economics*, 1(C), 641–92.

Ryan, Richard M. and Edward L. Deci (2000), “Self-Determination Theory and the Facilitation of Intrinsic Motivation,” *American Psychologist*, 55(1), 68–78.

Ryan, Richard M. and Edward L. Deci (2001), "On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being," *Annual Review of Psychology*, 52, 141–66.

Shah, Anuj K., Eldar Shafir, and Sendhil Mullainathan (2015), "Scarcity Frames Value," *Psychological Science*, 26(4), 402–12.

Sharma, Eesha and Adam L. Alter (2012), "Financial Deprivation Prompts Consumers to Seek Scarce Goods," *Journal of Consumer Research*, 39(3), 545–60.

Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson (2020), "Specification Curve Analysis," *Nature Human Behaviour*, 4(11), 1208–14.

Spector, Paul E. (1997), *Job Satisfaction: Application, Assessment, Causes, and Consequences*, 3, Sage.

Spector, Paul E. (1985), "Measurement of Human Service Staff Satisfaction: Development of the Job Satisfaction Survey," *American Journal of Community Psychology*, 13(6), 693–713.

——— (2022), *Job Satisfaction: From Assessment to Intervention*, Routledge.

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

Statista (2024), *Gig Economy in the U.S. - Statistics & Facts*.

Stern, Scott (2004), "Do Scientists Pay to Be Scientists?," *Management Science*, 50(6), 835–53.

Taber, Tom D. and George M. Alliger (1995), "A Task-Level Assessment of Job Satisfaction," *Journal of Organizational Behavior*, 16(2), 101–21.

Thaler, Richard H. and Shlomo Benartzi (2004), "Save More Tomorrow ™: Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, 112(S1), 164–87.

Tully, Stephanie M., Hal E. Hershfield, and Tom Meyvis (2015), “Seeking Lasting Enjoyment with Limited Money: Financial Constraints Increase Preference for Material Goods over Experiences,” *Journal of Consumer Research*, 42(1), 59–75.

Tully, Stephanie M. and Eesha Sharma (2022), “Consumer Wealth,” *Consumer Psychology Review*, 5(1), 125–43.

Tversky, Amos (1977), “Features of Similarity,” *Psychological Review*, 84(4), 327–52.

Weiss, Howard M. (2002), “Deconstructing Job Satisfaction. Separating Evaluations, Beliefs and Affective Experiences,” *Human Resource Management Review*, 12(2), 173–94.

Wertenbroch, Klaus, Dilip Soman, and Amitava Chattopadhyay (2007), “On the Perceived Value of Money: The Reference Dependence of Currency Numerosity Effects,” *Journal of Consumer Research*, 34(1), 1–10.

Zauberman, Gal and John G. Lynch (2005), “Resource Slack and Propensity to Discount Delayed Investments of Time Versus Money” *Journal of Experimental Psychology: General*, 134(1), 23.

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

CHAPTER 2: CONSUMER RESPONSES TO INCOME VERSUS BALANCE INFORMATION

With Stephanie Smith and Stephen A. Spiller

ABSTRACT

Consumers do not spend indefinitely, but rather they calibrate their spending against their available funds. But what are these *available funds*? Two likely measures are a person's income or their current bank balance. While both are capable of describing a consumer's financial situation, income is a *flow* and balance is a *stock*. Yet, prior research on stock-flow reasoning suggests this distinction may not be fully accounted for. As such, we find consumers who attend relatively more to one metric over tend to treat this as their amount of available funds. Because consumers are debt averse, they attempt to spend less than whatever amount they attend to (income or balance). Furthermore, because of systematic differences in the relative magnitudes of income and balance, attending to whichever form is smaller in magnitude leads to less spending in the moment. We conclude that this underspending (when attending to the smaller amount) is inadvertent accumulation, explained by the asymmetric memory of flows and stocks.

Keywords: stock-flow reasoning, debt-aversion, attention, income, consumer spending

Consumers seek to acquire products through spending but also seek to avoid debt (Prelec and Loewenstein 1998). Balancing these opposing motivations requires spending to be constrained at some limit of available funds. But what metrics do consumers use to assess their *available funds*? Two likely metrics are income and bank account balance. While both measures convey related aspects of a person's financial situation, these metrics differ along an important dimension: Income is a *flow* and balance is a *stock*; therefore, income and balance can describe (equivalent) financial information through different formats. If either metric leads to different financial judgments, then income versus balance information could have implications for consumer spending. Understanding such implications is important because consumers' financial service providers—ranging from traditional banks to modern fintech applications—may communicate financial information using either an income or balance frame. Furthermore, the decision to present information as income or balance may be, at present, largely arbitrary. We aim to explore consumers' responses to these two presentation formats to better understand whether, and why, income versus balance information affects consumer spending.

Why might consumers be sensitive to learning about their finances in terms of either income or balance? Drawing from research describing the challenges of stock-flow reasoning (Cronin, Gonzalez and Sterman 2009; Kainz and Ossimitz 2002; Korzilius et al. 2014; Newell et al. 2016; Reinholtz, Maglio and Spiller 2021; Spiller, Reinholtz and Maglio 2020; Sweeney and Sterman 2000), we propose consumers may not fully recognize the distinction between income and balance when thinking about their spending constraints. But this distinction *is* important. Income is a flow and represents a change in wealth. Balance is a stock and represents a level of wealth. If consumers do not translate between the two formats, their judgments and decisions regarding their own finances may depend simply on whether this information is presented in terms of flows or stocks. Under certain conditions (e.g., when income is small, relative to balance), this may lead to predictable differences in spending and accumulation. We consider these propositions in contexts in which only income or balance information is available, as well as contexts in which both are available. In cases when both metrics are available, we propose that a consumers' attention can both reveal and induce how they perceive their available funds. Therefore, we propose attention to income or balance information will affect spending and accumulation over time.

We begin by discussing the literature on stock-flow reasoning, which highlights that judgments and decisions are sensitive to whether information is presented in terms of flows or stocks. We contribute to this literature by incorporating attention (to either flows or stocks) as a component to stock-flow reasoning, which is crucial for contexts in which both flow and stock information might be accessible. Next, we discuss consumer spending and debt avoidance. In particular, we focus on how debt avoidance strategies (e.g., budgeting) depend on calibrating spending against some level of constraint. We clarify that income and balance—which correspond to a flow and a stock—are candidate metrics for inclusion in the consumers' budget constraint. If consumers are insufficiently sensitive to the differences implied by income (as a flow) and balance (as a stock), then downstream judgments, decisions, and behavior should depend on whether consumers engage relatively more with income or balance information. We explore these ideas across three studies. The first study tests whether consumer spending decisions differ depending on whether a monthly bank statement presents equivalent finances using either an income or balance frame. Studies 2 and 3 use an experimental game in which we both measure (study 2) and manipulate (study 3) attention and observe the resulting spending decisions and outcomes.

FLOWS, STOCKS, AND THE ROLE OF ATTENTION

Imagine a consumer has $2,625 in their bank account, which grows to $3,625 after $1,000 of income is deposited. This financial situation can be described in terms of either flows or stocks. Flows communicate changes over time—in this case, the $1,000 increase of funds. Stocks communicate levels of a resource—in this case, $2,625 as the starting level and $3,625 as the ending level. Therefore, flows and stocks are *different* presentations of *equivalent* data. Prior work on stock-flow reasoning suggests people are sensitive to flow versus stock presentations.

People Do not Spontaneously Translate Between Flows and Stocks. One potential explanation is that people do not spontaneously translate between flows and stocks. In such cases, judgments may be sensitive to the trends made salient by a given presentation format. For example, Spiller, Reinholtz, and Maglio (2020) present time-series data of US employment numbers from 2007-2013, with a particular focus on a subperiod during which there was job loss (falling stock), but at a decreasing rate (increasing flow). Participants who considered this information in term of stocks judged the state of the economy as worse than participants who considered this information in terms of flows. This and related work (Reinholtz et al. 2021; Weber et al. 2025) suggest judgments are sensitive to stock versus flow information because people do not automatically translate between the two formats, but rather make inferences about the information as given.

People Are Inaccurate in their Stock-Flow Translations. In addition to the findings that people do not spontaneously translate between formats, a related body of research suggests that people are inaccurate in making stock-flow translations, even when prompted. Research documents persistent shortcomings of stock-flow reasoning across a variety of domains, including CO2 emissions, business logistics, and simple thought experiments (Newell et al. 2016; Ossimitz 2002; Sweeney and Sterman 2000). Difficulty in stock-flow reasoning persists even for highly educated people (Cronin et al. 2009), and improvements may require specialized training programs (Kainz and Ossimitz 2002) or elaborate problem-solving processes, such as thinking aloud (Fischer, Degen and Funke 2015; Korzilius et al. 2014).

Flows and Stocks Differ in Amount and Memory. Examining this prior literature, we call attention to two key dimensions upon which stocks and flows necessarily differ. First, presenting the same data in terms of either flows or stocks necessarily implies presenting different numerical amounts,[9] where both the direction and the magnitude of the presentation forms may differ. In both existing research (Spiller et al. 2020) and in our own theorizing, these different amounts conveyed by flows and stocks are important determinants of judgments and decisions.

A more nuanced and perhaps lesser appreciated distinction between flows and stocks is their asymmetric memory for past events. Flows are rates of change. Information about the current flow does not communicate anything about past flows. This is not the case for stocks, which are levels. Stocks "remember" and summarize past events, because past flows accumulate to the current level of a resource. This dimension of memory should be particularly important when decision makers consider information over time, as opposed to merely observing time-series data. This is an important distinction from much of the past literature, which typically

[9] The sole exception over time is when a 0 level changes by 0 between time periods.

presents decision makers with time series data, which is necessarily constructed in terms of either flows or stocks (but not both). Such an approach is appropriate for contexts in which people make judgments or forecasts regarding trends (Reinholtz et al. 2021; Spiller et al. 2020; Weber et al. 2025); however, perhaps not contexts like making spending or savings decisions.

The Role of Attention, in the Moment. When people make decisions in real time (as opposed to viewing summaries of time-series data), they may have access—or be exposed—to both flow and stock information. For example, consumers can easily check their paystubs for income information or access their banking portal for current balances. Therefore, we consider how consumers' attention relates to stock-flow decisions. Specifically, attention research suggests that people put more weight on the information they spend more time looking at (Amasino et al. 2019; Fiedler and Glöckner 2012; Fisher 2017; Glickman et al. 2019; Rramani et al. 2020; Smith and Krajbich 2018; Yang and Krajbich 2021). Therefore, consumers who attend relatively more to flows should be more likely to use flow information as an input for their decisions; and consumers who attend relatively more to stocks should be more likely to use stock information as an input for their decisions.

THE CONSUMER BUDGET CONSTRAINT

Our basic argument is that consumers consider their personal financial constraint in terms of flows or stocks, depending on whether their finances are presented in terms of income or balance, respectively. To the extent that considering constraint based on income or balance may involve different numerical amounts or asymmetric memory of past financial events, consumer spending should depend on presentation format.

In considering financial constraints, we first acknowledge consumers are constantly tempted by the pleasure of immediate consumption. This makes it necessary to limit present consumption to allow for future consumption (Ando and Modigliani 1963; Browning and Crossley 2001; Prelec and Loewenstein 1998; Shefrin and Thaler 1988). The consumer budget constraint (i.e., constraining spending to some level of available funds) is pivotal for combating myopia and distributing consumption over time. Calibrating spending against available funds may be difficult for at least two different reasons.

Controlling Spending. One challenge is whether consumers *can* limit their spending. A large and separate literature documents failures of self-control to limit present consumption (Ainslie 1975; Frederick, Loewenstein and O'Donoghue 2002; O'Donoghue and Rabin 2000; Thaler 1980). Various interventions can assist consumers desiring more savings to overcome issues of self-control. For example, people might be nudged to save more if they can pre-commit to future saving (Thaler and Benartzi 2004), visualize their future self as the recipient of saved funds (Hershfield et al. 2011), receive persistent reminders of their savings goals (Karlan et al. 2016; Soman and Cheema 2011), hold fewer savings goals (Soman and Zhao 2011), or if spendable money is perceived as less liquid (Ashraf et al., 2006; Shefrin & Thaler, 1988).

Limiting Spending to a Specific Amount. A separate challenge, which is the focus of the present research, concerns consumers' assessment of *how much* they can spend. This is because consumers do not typically think about their available funds in terms of overall assets or net worth (Olafsson and Pagel 2018; Shefrin and Thaler 1988). Instead, consumers tend to operate on a more localized conceptualization of money, such as the distribution of funds across various mental accounts (Heath 1995; Helion and Gilovich 2014; Morewedge, Holtzman and Epley 2007; Okada 2002; Prelec and Loewenstein 1998; Soster, Monga and Bearden 2010; Thaler 1980, 1985, 1999; Thaler and Johnson 1990; Zhang and Sussman 2018a, 2018b). This suggests consumers' constraint likely depends on how they organize and track their finances. We suggest this depends—in part—on how financial information is conveyed.

A key question is how consumers—even those with sufficient self-control to limit spending—perceive their available funds. Based on consumers' preference to calibrate current spending against current resources (as opposed to future resources, such as an eventual pension; (Hirst, Joyce and Schadewald 1994; Shefrin and Thaler 1988), we focus primarily on present funds. Based on consumers' relative sensitivity to liquid over illiquid funds (Olafsson and Pagel 2018), we focus primarily on liquid resources. Therefore, it follows that consumers are likely to think about their personal budget constraint based on liquid funds in the present moment. The common ways of describing information about present liquid resources is to communicate information about *income* or *balance*.

Of course, income and balance are two sides of the same coin: The former is a flow and the latter is a stock. Income mechanically raises balance, and either format can be translated to the other using simple arithmetic. Yet, because people do not spontaneously—or accurately—translate between flows and stocks (Cronin et al. 2009; Spiller et al. 2020; Sweeney and Sterman 2000), we expect consumers will be sensitive to the specific presentation of their finances in terms of either income or balance. Specifically, consumers will tend to think of their available funds in whichever format they are attended to: as either income or balance.

H1: Attending relatively more to income or balance increases the likelihood of using that amount as a measure of available funds.

The Role of Debt Aversion. Consumers are typically debt averse, meaning they experience disproportionately large psychological discomfort from spending more money than they have available, compared to spending that does not incur debt. Debt aversion is reference dependent, meaning consumers attempt to avoid overspending their financial reference point (Prelec and Loewenstein 1998). For example, spending that does not exhaust a consumer's total resources but does exceed the funds in a given mental account will be psychologically painful (Soster, Gershoff and Bearden 2014). To the extent that consumers use either income or balance as their measure of available funds, debt aversion implies spending may approach—but should not exceed—this amount.

Therefore, consumers who attend relatively more to income (and perceive income as their measure of funds) should be motivated to spend less than their income. And consumers who attend relatively more to balance (and perceive balance as their measure of funds) should be motivated to spend less than their balance.

H2: Attending relatively more to income or balance increases the likelihood of underspending that amount.

The Role of Magnitude. Whereas H2 suggests consumers will behave in a manner consistent with debt aversion for both income and balance information, it does not directly predict how attending to income or balance information will affect the level of consumer spending. Whether attention to income or balance predicts greater spending likely depends on the relative magnitudes of the two amounts.

Flows and stocks will *always* convey different amounts, making differences in magnitude inseparable from income versus balance presentations. When either income or balance is systematically larger in amount (greater numerosity) and that amount is attended to, consumers should perceive more available funds. This may arise because they interpret their constraint as having more available funds (e.g., $1,000 income is a smaller limit than $3,625 in balance). Alternatively, thinking about constraint with larger numbers (greater numerosity) may lead to differences in the perceived value or magnitude of their resources (Morewedge et al. 2007; Wertenbroch, Soman and Chattopadhyay 2007).

H3: Attending relatively more to the larger of income and balance increases consumer spending.

The Role of Memory. Income and balance differ on another dimension, regardless of whether either metric is systematically larger than the other. This dimension reflects the very nature of how flows and stocks communicate information *over time*. Specifically, income-based and balance-based constraints may differ in their “memory” of past events. Income—as a flow—is a rate of change, and balance—as a stock—is a level that “remembers” prior accumulation. Therefore, income-based constraints “forget” any past under- or overspending, and balance-based constraints do not. This asymmetric memory of income versus balance implies a greater memory burden on consumers using income-based constraints. Of course, an income-based consumer who exceeds their constraint by $100 during a past pay period may remember to reduce their current budget constraint by $100 (to maintain equivalence with the balance constraint); however, we suspect this might not always be the case. Therefore, we expect attending relatively more to income information will result in more unexpected accumulation (whether positive or negative) over time.

> **H4:** Attending relatively more to income (vs. balance) information leads to unexpected accumulation (over time).

In summary, we first predict that relative attention to either income or balance increases the likelihood consumers use that metric as the measure of available funds (H1). This reflects peoples' general tendency to accept flow and stock information as given, rather than translating between the two (Spiller et al. 2020). A consequence for consumers who are debt-averse is that attending to either income or balance should decrease the likelihood of spending more than that amount (H2). Whether this relates to an increase or decrease in consumer spending at any given moment depends on whether income or balance is systematically larger in magnitude (H3). The nature of flows and stocks also implies a difference from attending to income versus balance information over time. Namely, income "forgets" past instances of underspending or overspending, relative to an individual's personal constraint. But balance "remembers" past financial events. The memory asymmetry of the information itself suggests over time, consumers who attend to income will experience a higher degree of unexpected accumulation (whether positive or negative) (H4).

OVERVIEW OF STUDIES

We begin with a descriptive survey examining consumers' tendencies, preferences, and beliefs regarding income or balance information. We then present three studies exploring the role of attention to income or balance information on consumer spending outcomes. The first study considers the hypothetical choice of a cheaper or more expensive purchase for participants who interpret their finances through flows or stocks. This study provides a test of H1 and H2. The second and third studies use a multiperiod, incentive-compatible spending game to test H2, H3, and H4. The paradigm captures repeated measures of spending alongside attention and finances. In study 2, we measure attention to income and balance through mouse-tracking. To address potential endogeneity concerns, study 3 uses the same paradigm but directly manipulates attention through the availability or absence of financial information. Therefore, study 2 provides an endogenous measure of proportional attention (to income or balance) and study 3 provides an exogenous measure of attention to either income or balance. All three studies were preregistered[10], and all materials are available at https://researchbox.org/3968&PEER_REVIEW_passcode=EFOMGE.

[10] Study 2 deviates from our preregistration by excluding participants with highly erratic spending behavior inconsistent with game mechanics. This deviation does not impact the conclusions of the majority of our analyses. One analysis of aggregate-level spending is impacted, and we argue that deviating from our preregistration provides a more appropriate test. We formally adopt this exclusion rule in our preregistration for study 3. Study 3 deviates from our preregistration in terms of the condition codes we use, largely for ease of explication and robustness. These discussions, along with all preregistered analyses, will be included in the supplementary materials.

DESCRIPTIVE SURVEY

Given the motivation of the current research is to consider how the presentation of financial information (in terms of either flows or stocks) affects consumer judgments and decisions, we begin by asking consumers about their own experience, preference, and beliefs regarding income and balance information. To assess consumers' personal experiences thinking in terms of either income or balance, we asked 100 Prolific survey participants (M_{age} = 36, 48% female) to describe "the exact approach you use to determine how much money you can spend. There are no right or wrong answers, but please be as specific and precise as possible in describing your approach." Participants later self-coded their responses as pertaining to "mostly [their] paycheck," "mostly [their] bank account balance/s," "both," or "neither." After completing the open-ended response (but prior to the self-coding), survey respondents also imagined signing up for a new bank account. As part of this exercise, participants were informed the bank sends text message notifications whenever a direct deposit occurs. Respondents indicated whether they would prefer this information displayed as "the amount of the direct deposit (e.g., 'You have just received a payment of $800')" or "the amount in your bank account after the deposit hits (e.g., 'You now have $1,700 in your account/s')." Finally, and on a new page, participants considered the prior numbers ($800 income and $1,700) as their own and reflected upon whether income or balance would be "a more important factor as you decide how to manage your spending."

The descriptive survey summarizes the following responses regarding income versus balance information: (1) what metric consumers typically use (based on their self-coding); (2) what they prefer (based on their bank notification preference); (3) and what they feel is more important (given a specific financial scenario). Responses to the self-coded measure (reflecting the information type that is typically used) reveal 33% focus mostly on their paycheck, 27% focus mostly on their bank balance/s, 31% focus on both, and 9% focus on neither. When asked to identify which they prefer (if they were to sign up for a new bank with direct deposit text message notifications), 64% indicated a preference for income information and 36% indicated a preference for balance information. When asked to reveal whether they believed income or balance information was more important (to manage spending) under the imagined scenario of $800 in weekly income and $1,700 in a bank account balance, participants were nearly evenly split at 51% and 49%, respectively.

Across three questions intended to capture what consumers do, prefer, and believe to be important, we find evidence that consumers rely on the metrics of income and balance to manage their finances (only 9% report using neither). This supports our decision to examine income and balance, as opposed to other financial metrics (e.g., retirement savings or net worth). While the vast majority of respondents report using income or balance in their own financial decisions, we observe substantial heterogeneity in which metric people actually use, prefer to know, and believe is most important. This heterogeneity underscores our interest in this topic, as consumers may naturally encounter either (or both) types of information—perhaps as the result of arbitrary decisions by banks or financial service providers. We seek to understand whether this seemingly innocuous difference in presentation format matters for consumers.

STUDY 1: BANK STATEMENTS

Study 1 seeks to test whether attending to either income or balance information increases the likelihood of using that metric (income versus balance) as a measure of available funds (H1), and therefore underspending that amount (H2). If consumers do not structure their budget constraints around the information they attend to or exhibit decision making inconsistent with localized debt aversion, then the subsequent hypotheses (requiring repeated decisions over time) cannot hold. Therefore, we test the first two hypotheses in a simple paradigm where we present financial time series data summarized in terms of either flows or stocks.

We assess H1 by asking participants to identify whether their own spending constraints are based more on income or balance. We assess H2 by examining rates of purchasing a cheaper ($300) or more expensive ($1,200) car maintenance service. The $1,200 option exceeds the weekly $1,000 of income. Therefore, by H2, we predict participants who attend relatively more to income information will consider their constraint as having $1,000 in available funds (H1) and be more likely to choose the cheaper $300 option rather than the $1,200 option (H2).

Method

A total of 499 participants were recruited from Prolific Academic to take part in this study in exchange for a small payment.

Participants imagined owning a malfunctioning vehicle and receiving the following two options from their mechanic: They could repair the problem for $300 or perform a comprehensive maintenance (that would also repair the problem) for $1,200. In order to decide which option to pursue, participants were asked to inspect their hypothetical bank statement from the prior month. The statement showed an initial bank balance of $1,075, which grew to $3,625 by the end of the month, after accounting for income and expenditures. Participants were randomly assigned to see this statement presented in terms of either income (income deposits and expense withdrawals) or balances, as depicted in figure 2.1. These two statements depict *equivalent* financial information (a starting balance of $1,075 growing to $3,625 due to four income deposits of $1,000 each and two expenses totaling $1,450).

FIGURE 2.1: BANK STATEMENTS PRESENTED AS FLOWS OR STOCKS (STUDY 1)

Monthly Bank Statement (Jan. 15 - Feb. 14)

Beginning Balance: $1,075

	Event Description	Withdrawals	Deposits
01/22	Monthly Auto Loan	-$350	
01/24	Income		+$1,000
01/31	Income		+$1,000
02/01	Monthly Rent	-$1,100	
02/07	Income		+$1,000
02/14	Income		+$1,000

Monthly Bank Statement (Jan. 15 - Feb. 14)

	Event Description	Balance
01/15	***Beginning balance***	$1,075
01/22	Monthly auto loan is withdrawn	$725
01/24	Income is deposited	$1,725
01/31	Income is deposited	$2,725
02/01	Monthly rent is withdrawn	$1,625
02/07	Income is deposited	$2,625
02/14	Income is deposited	$3,625

NOTE—Bank statements for the income condition (left) and balance condition (right) present equivalent information using different presentation formats. The statement period was Jan. 15 – Feb. 14 because data was collected on Feb. 15.

After considering the bank statement, participants decided whether they would choose the cheaper car repair ($300) or the comprehensive and more expensive servicing ($1,200). Subsequently and on a new page, participants indicated the maximum amount of money they felt they could spend (right now), given the financial profile. The purpose of this exercise was to have participants think about their spending limit. At the end of the study (following comprehension checks), participants self-coded their open-ended response as pertaining to the starting balance ($1,075), ending balance ($3,625), income ($1,000), other, or none of the above, by selecting all that applied.[11] Prior to self-coding their own response, participants encountered two comprehension questions (on separate pages). The first question asked participants to identify the starting balance on Jan. 15 out of a set of three options ($575; $1,075 = correct; $1,575). The second question asked participants to identify the entire financial scenario out of a set of three options (the correct response had the starting balance, ending balance, total income, and total expenses; for exact wordings see appendix A).

Results

All participants were included for analysis (M_{age} = 39; 63% female). We first examine whether participants were more likely to consider income or balance information when considering their spending limits (as indicated by the self-coded responses). Consistent with H1, participants in the income condition were more likely to say their spending limits would be based on the $1,000 weekly income (60.4% vs. 45.4%; $t(498) = 3.36$, $p < .001$). Conversely, participants in the balance condition were more likely to indicate their spending limits would relate to their starting or ending balances (66.1% vs. 49.6%; $t(498) = 3.72$, $p < .001$).

Does directing attention towards income or balance information affect the likelihood of making a $300 purchase (H2)? In the income condition, 79.84% of participants (198/248) chose the $300 option, compared to only 61.75% (155/251) in the balance condition ($b_{logistic}$ = 0.897, *se*

[11] The self-coding was positioned at the end of the survey to ensure that participants were not reminded of aspects of the financial statements (i.e., starting and ending balance, and income) prior to the comprehension questions, which asked about these very metrics.

= 0.205, $t(498) = 4.38$, $p < .001$, Cohen's $d = 0.40$). This is consistent with our prediction, because participants who think about their available funds as $1,000 of income should be averse to overspending this amount on the expensive $1,200 car maintenance.

To ensure these results are not based solely on participants' misunderstanding of the financial context, we reconsider our results based on participant comprehension. There was a meaningful rate of incorrect responses to the comprehension questions (as discussed shortly), though arguably not to an alarming level, given participants had to recall specific amounts from memory based on the financial statements they read. Importantly, there were no differences in the rates of incorrect responses to comprehension questions across conditions (ps > .46; appendix A). Furthermore, the key analysis on spending decision remains significant when excluding then 94 participants who incorrectly recalled the starting balances ($b_{logistic} = 0.723$, $se = 0.228$, $t(404) = 3.18$, $p = .002$, Cohen's $d = 0.32$), the 115 participants who incorrectly recalled the entire financial sequence ($b_{logistic} = 0.682$, $se = 0.233$, $t(383) = 3.18$, $p = .003$, Cohen's $d = 0.30$), or the 181 who recalled either incorrectly ($b_{logistic} = 0.582$, $se = 0.260$, $t(317) = 2.24$, $p = .025$, Cohen's $d = 0.25$).

Discussion

Study 1 describes finances in a manner consumers might naturally encounter them: as a monthly bank statement. Regardless of whether the statement was presented in terms of flows (income and expenditures) or stocks (running balances), the underlying finances were equivalent. All participants saw the same starting balance, number of financial events, and event descriptions, and funds accumulated at the same rate and to the same level. Consistent with H1, attending relatively more to income information (by being assigned to the income condition) increased the likelihood that participants self-reported considering income when thinking about the maximum amount of money they could spend. Consistent with H2, participants in the income condition—who should prefer to avoid an income deficit by spending less than $1,000—were less likely to choose a $1,200 purchase than those in the balance condition.

These results illustrate how the simple presentation format of personal finances can affect participants' judgments and decisions. These findings integrate prior research on both stock-flow reasoning (e.g., Cronin et al. 2009; Spiller et al. 2020; Sweeney and Sterman 2000) and consumer debt aversion (e.g., Prelec and Loewenstein 1998). The pattern of results suggests it is not merely that participants interpret income and balance information as measures of available funds (without sufficient translation between forms), but also that this becomes the reference point for experiencing debt aversion. Consumers attempt to underspend the amount they attend to, even if that number is a somewhat arbitrary reflection of their financial state.

Nevertheless, study 1 is somewhat of a caged test of both attention *and* income versus balance. We manipulate attention by providing income or balance statements; however, this falls short of consumers' endogenous direction of attention. We summarize a time series of financial transactions in a single bank statement; however, this falls short of experiencing information over time. Furthermore, the single choice in study 1 does not permit a true test of spending (H3), and the single-shot nature does not allow us to consider accumulation outcomes over time (H4). Therefore, the subsequent study introduces mouse-tracking—as a proxy for attention—to understand how self-directed attention to either income or balance corresponds to incentivized spending decisions over time.

STUDY 2: INCENTIVIZED GAME WITH ENDOGENOUS ATTENTION

We developed an incentivized game to observe how attention to either income or balance relates to spending decisions over time. The game is constructed around several desirable properties.

First, we make both income and balance information available and measure attention to each (proxied by mouse-tracking). This offers a potential contribution to the stock-flow literature, which has traditionally considered presentations of one format or the other. We believe allowing for attention to either metric is particularly important in the domain of financial decisions, where consumers can choose to attend to their direct deposits or log in to their bank accounts.

Second, the game involves incentivized spending decisions. This allows us to test whether relative attention to income or balance is associated with higher or lower levels of spending (H3).

Third, there are many repeated observations, allowing participants to experience (and the researchers to observe) financial information and spending decisions over time. We impose a clear (and incentivized) goal to the game, allowing us to consider the extent to which attending relatively more to income or balance is associated with deviations from this goal (H4).

In summary, study 2 is an experimental game in which participants make repeated spending decisions over time. The paradigm is engaging (for participants) and data-rich (for researchers).

Method

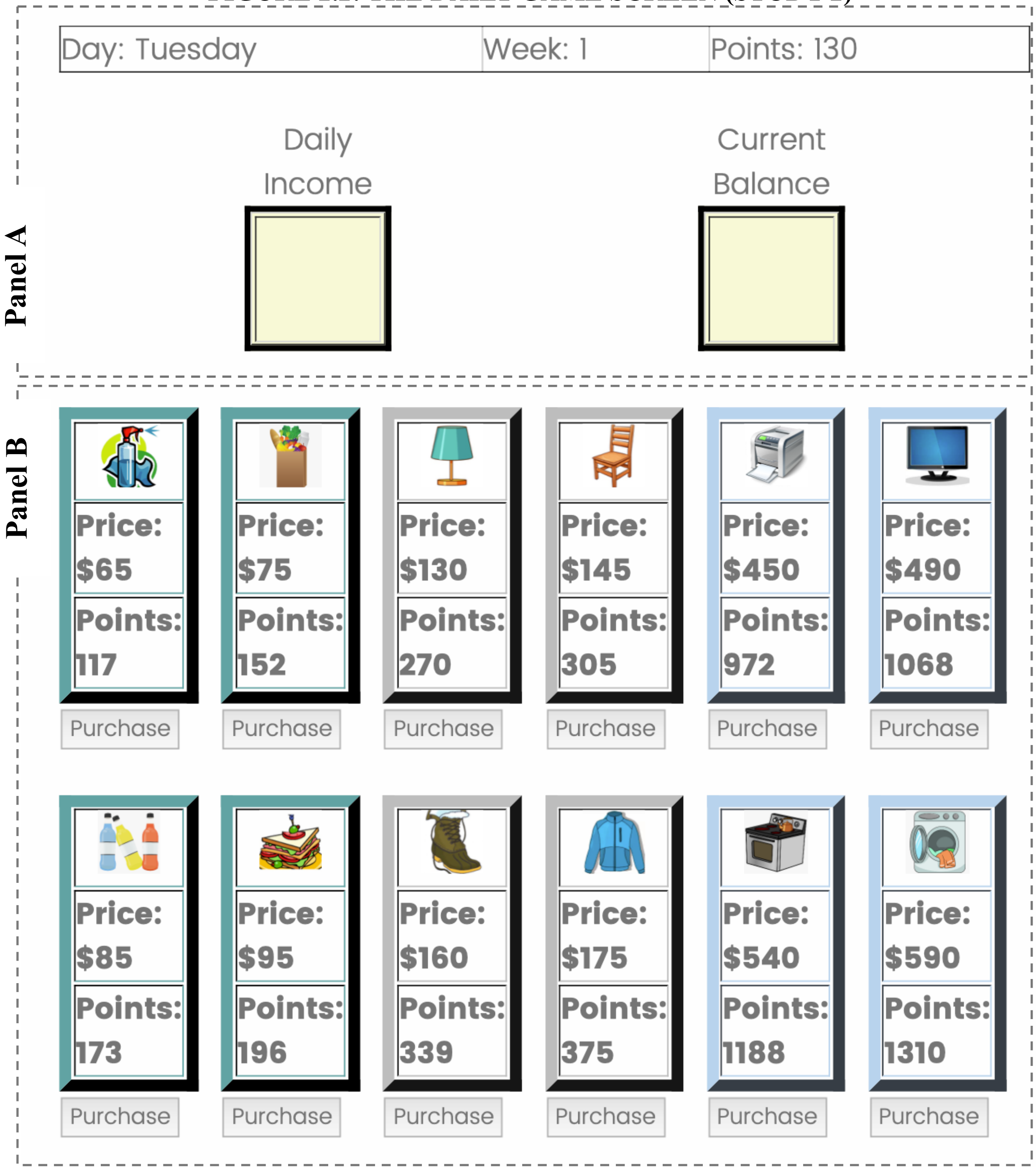
Participants. A total of 400 participants were recruited from Amazon Mechanical Turk (AMT) through CloudResearch (Litman, Robinson and Abberbock 2017). All participants were compensated $1.50 for completing the study and had the opportunity to earn an additional bonus payment. We preregistered a plan to exclude participants who never checked either their income or balance information during the incentivized four-week game. Such behavior is likely to signal inattention (it also prevents estimating an attention variable due to a lack of variation). The exclusion resulted in a remaining sample of 387 participants. We deviated from our preregistration by excluding an additional 36 participants who had extreme spending behavior extraordinarily inconsistent with the incentives of the game. This deviation does not meaningfully impact the interpretations of the majority of our results. One additional participant was excluded for having incomplete data, resulting in a final sample of 350 subjects (M_{age} = 40.84; 38% female).

Game Structure. The spending game was structured as a series of simulated "days" organized into five-day "weeks" (Monday-Friday), though the entire experimental paradigm took place within a single session. Every day began with an income payment, ostensibly representing daily wages from work. Subsequently, participants had the opportunity to earn points by purchasing consumer products with various costs. More spending earned more points, which participants maximized to earn a larger bonus. There were five weeks; the first was described as "practice" and the final four constituted the incentivized game. As preregistered, we only consider the incentivized four weeks in our analysis.

Daily Income. Dollars were earned each day in exchange for clicking a button at the "Click Factory" ten times to simulate work. We used variable income—rather than fixed income—to make the game more engaging and reflect the experience of an increasing number of US households with non-salaried, part-time, or volatile income (West, Whillans and DeVoe 2020). Daily income was repeatedly drawn in whole-dollar increments from the uniform distribution bound by [$100, $200]. Participants were informed of the lower and upper income bounds, as well as the expected value of draws from this distribution ($150 each day; $3,000 over the span of 20 days). After simulating work each day, participants progressed to the main spending screen.

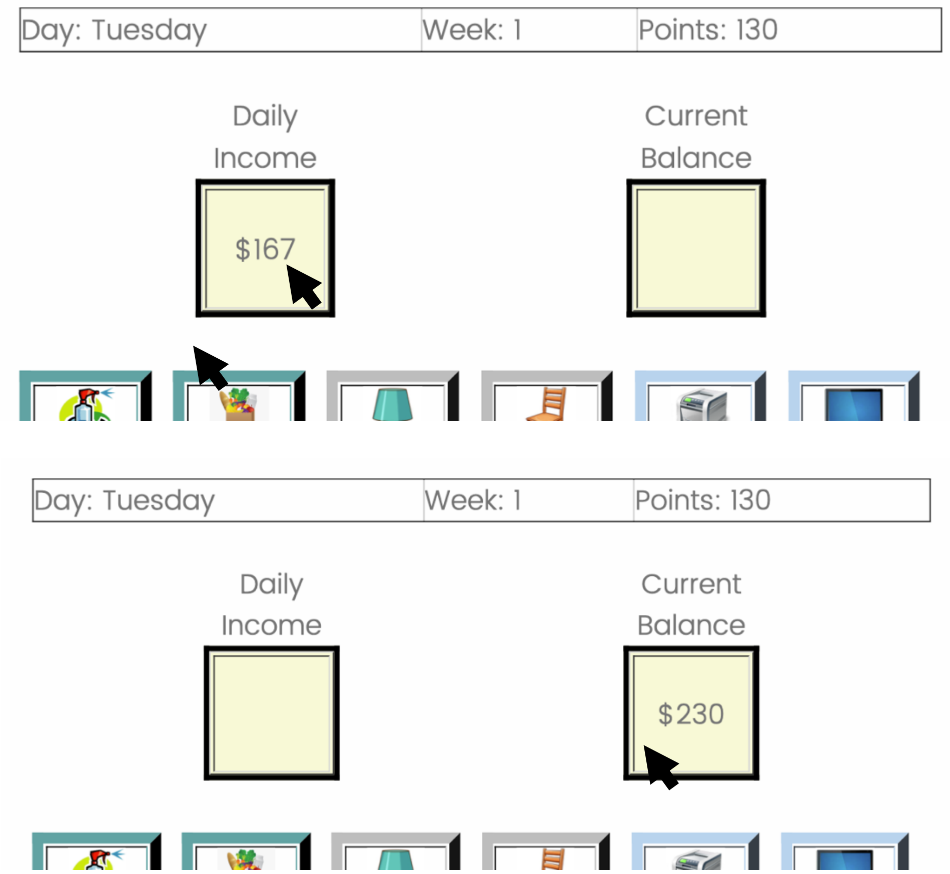
Mouse-Tracking. Daily income information was not immediately available on the spending screen. Instead, participants checked this information by hovering their mouse over a square box labelled "Daily Income." On the other side of the screen, hovering over another box labelled "Current Balance" revealed the bank balance (figure 2.2, panel A). Balance always integrated the current day's income with any past accumulation. For example: a daily income of $167 would be added to the previous day's accumulated balance of $63 to display a current balance of $230. Either piece of financial information remained visible only while the participant's mouse remained within the perimeter of the information area (figure 2.3). The sequence of dwell times was recorded for both the daily income and current balance regions of interest.

FIGURE 2.2: THE DAILY GAME SCREEN (STUDY 2)



NOTE—The daily game screen, as viewed by participants in study 2. Panel A permanently displayed game information (day of week, week, points). It also required a mouse hover to reveal income or balance information. Panel B is the grid of 12 possible purchases. Each purchase had a fixed price during the game, but the associated points diminished exponentially through repeat consumption.

FIGURE 2.3: MOUSE-TRACKING FOR INCOME AND BALANCE (STUDY 2)



NOTE—Hovering a mouse over either region of interest temporarily displayed the associated daily income or current balance information until the mouse was removed. The examples depict hovering over daily income (top) and current balance (bottom).

Daily Purchase Menu. Each day, participants repeatedly encountered a menu of 12 possible purchases with prices ranging from $65 to $590 (figure 2.2, panel B). These potential purchases were marked by images depicting ordinary consumer goods (e.g., household cleaning supplies, clothing, and home appliances). The images and product categories were intended to enhance participant engagement and game playability; these were not relevant to the game's incentive structure. In addition to price, the important attribute was the item's points. Higher points were associated with higher prices. Participants maximized their real bonus by collecting as many points as possible through purchases. Points decayed exponentially through repeated consumption of the same item. Additional information about the precise point mechanics is available in appendix B.

Financing Purchases. During the incentivized four-week game, participants had access to zero-interest financing, allowing them to spend beyond their balance at any moment without penalty. However, there was a substantial penalty for ending the game in debt, with borrowed money being repaid directly from points at a rate of $1 of debt : 3 points. Therefore, it was in each player's interest to spend entirely without incurring substantial debt by the end of the game. This incentive structure reinforces debt-aversion at the game-level while eliminating any competing savings goals. This allows us to interpret any accumulation of unspent resources as inadvertent, not strategic.

Participants knew they would have an opportunity to earn an additional bonus based on their income and spending decisions. Prior to completing the game, participants knew only that the maximum bonus was $0.20 and that they could increase their likelihood of receiving a larger bonus by scoring as many points as possible. Bonuses were issued to participants who scored 5,550 points or more and the amount increased by $0.01 for every 50 additional points until reaching the limit of $0.20.

Learning and Feedback. Game mechanics were learned through a series of instructions with five comprehension checks. After completing the instruction exercises but before playing the incentivized game, participants tested the game for one practice week. At the conclusion of the practice week, all participants received feedback about their daily incomes, end-of-week balance, total points, and total score, which was the number of points adjusted for any debt penalty. This feedback was not provided during the incentivized portion of the game.

Analysis Plan

Levels of Analysis. The nature of the spending game allows us to consider the data at multiple levels. The trial-level data consist of repeated daily observations. These data can be analyzed to consider how daily attention to current resources is associated with various patterns of spending. These repeated measures can be aggregated into game-level measures. We use aggregate-level analyses to test whether patterns at the trial level emerge for the total game. Such analyses are important because participants face incentives at the level of the game, not individual days. Trial- and aggregate-level analyses could diverge if participants use inconsistent strategies across the game (e.g., spend nothing for 19 days, then spend $3,000 on a single day). Therefore, the aggregate-level analyses summarize whether relationships at the daily level hold at the game level. The aggregate-level analysis of spending is essential to testing whether attention to either income or balance is associated with unexpected accumulation at the end of the game (H4).

Independent Variables. Mouse-tracking data serves as a proxy for attention, which we use to construct the key independent variable. Specifically, we construct a proportional attention measure, which we frame as either "proportional attention to income" (total dwell time in the income region of interest / total dwell time in both regions of interest) or "proportional attention to balance" (1 – proportional attention to income). We emphasize these are equivalent and perfectly translatable measures of proportion, where the naming convention is for ease of explication. We construct a proportional measure (rather than using raw mouse-tracked timings), because proportion is a useful proxy for the allocation of attention that is not affected by individual differences in response time.

Income Surplus as the Dependent Variable for H2. Recall, H2 proposes relative attention to either income or balance increases the likelihood of underspending that amount. Specifically, we can observe whether participants spend less than the daily income or balance (implying adherence to an income-based or balance-based budget constraint). We use the terminology *income surplus* to signify underspending daily income ($Spending_t < Income_t$) and *balance surplus* to signify underspending relative to the current balance ($Spending_t < Balance_t$). H2 predicts proportional attention to income (balance) will be positively associated with the likelihood of an income (balance) surplus.

As preregistered, our focus is the income surplus (for testing H2), not the balance surplus. While both are of equal conceptual interest, there are mechanical limitations to hypothesis testing the balance surplus in this paradigm, as discussed in appendix B. The core challenges are that (i) we use a proportional measure of attention and (ii) balances tend to be larger than income. On the 81% of days in which balance > income, it follows that an income surplus necessarily implies a balance surplus. Therefore, attending proportionally more to income (equivalent to *attending proportionally less to balance*) should be associated with more income surpluses, which also means *more balance surpluses*. Of course, this is illogical and arises mechanically due to the described circumstances. We consider alternative and exploratory methods for analyzing balance surpluses by considering both model-free evidence and bunching analyses on the distributions of spending (Allen et al. 2017).

Daily Spending as the Dependent Variable for H3. To address H3 (attending relatively more to the larger of income and balance increases spending), we consider the relationship between the proportional attention to income and daily-level spending. Within the experimental game, income should be systematically smaller than balance (see appendix B). Therefore, we expect proportional attention to income to correspond with less spending at the daily level.

Aggregate Spending as the Dependent Variable for H4. The key difference between H3 and H4 is the time horizon. This distinction is meaningful, because it separates consumers' momentary spending decisions from their longer-term financial strategy. If consumers are not fully aware of how income and balance communicate differently about past accumulation, then even consumers with the same goals will spend differently over time. The incentivized mechanics of study 2 allow us to interpret deviations from the spending goal (spending entirely, without incurring debt) as inadvertent. We consider aggregate spending throughout the entire game to test H4, which predicts attention to income will be associated with more unexpected accumulation (in other words: less spending).

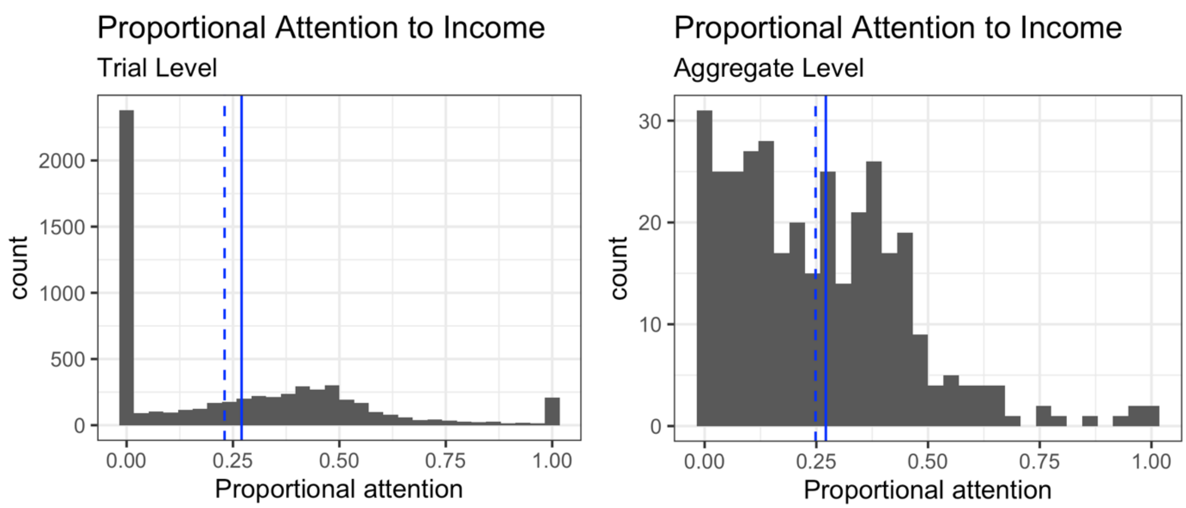
Preregistered Models. We preregistered a trial-level and an aggregate-level model for our set of dependent variables (income surpluses and spending). The trial-level model regresses the focal dependent variable on the following set of predictors: daily mean-centered proportional attention to income, daily mean-centered income, their interaction, the subject's average daily proportional attention to income, the subject's total income, and a set of day fixed effects. Mean-centering occurred at the participant level. To address issues of non-independence between observations, we use participant-level cluster-robust standard errors. The aggregate-level model does not use repeated observations and instead regresses the focal variable on the subject's average proportional attention to income and the subject's total income. Therefore, our full preregistered set consists of four models, across two dependent variables (income surplus and spending) and two levels of analysis (trial level and aggregate level).

Results

Attention. At the trial level, participants generally attended to financial information, with at least one piece of information considered on 86% of days. Across all participants and days, the mean proportion of time examining income information was 0.259 (SD = 0.275). Averaging the daily proportional attention to income measure at the participant level produced a nearly identical measure of subject-level attention to income (M = 0.256, SD = 0.198). Both trial-level and aggregate-level distributions of proportional attention to income are visualized in figure 4. A further analysis of the trial-level data suggests the proportional attention to income information decreased over time during the 20-day game (b = -0.009, se = 0.001, t(331) = -12.60, p < .001). This is addressed by the day fixed effects in our trial-level analyses.

Clearly, balance was the preferred metric by participants in this paradigm. This is to be expected, given the incentives of the game suggest balance is a more diagnostic metric (as previously discussed). We reiterate, however, that consumers in our descriptive survey were roughly evenly split about whether income or balance information was more important, in other scenarios, as well as their own lives. Therefore, the relative importance of one metric over the other is likely context dependent.

FIGURE 2.4: ATTENTION HISTOGRAMS (STUDY 2)



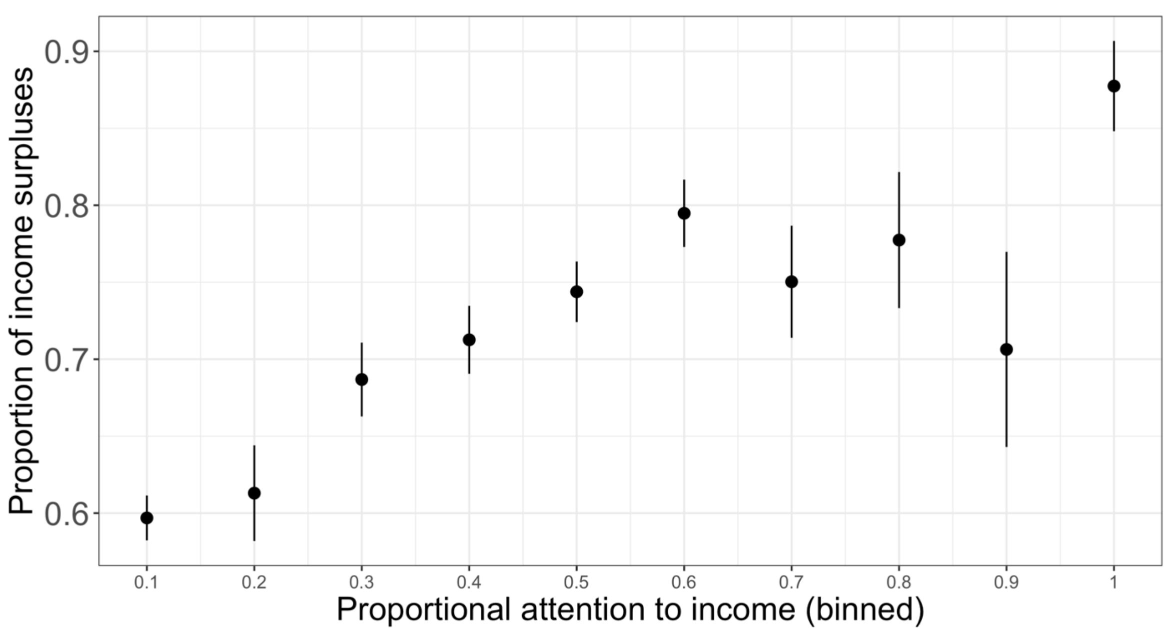
NOTE—Distributions of proportional attention to income at the trial level (left) and aggregate level (right). Solid lines represent mean values and dashed lines represent median values.

At the aggregate level, participants generally attended to both pieces of financial information. The mean number of days checking income during the 20-day game was 10.55 (*SD* = 6.99), compared to a mean of 16.66 days checking balance (*SD* = 4.83). These relatively high frequencies imply participants often checked both income and balance on a given day; on average, both pieces of information were checked on 9.96 days during the game (*SD* = 6.94).

Earning and Spending. Participants earned an average of $3,002 during the 20-day game (*SD* = $130), coinciding with the expected value of $3,000 ($\sum_{i=1}^{20} X_i \sim \mathrm{U}[100, 200]$). On average, participants spent a total of \$2,905 (*SD* = \$328), reflecting significant underspending of \$97, relative to their total income (b = -97.28, se = 17.57, $t(349)$ = -5.54, $p < .001$). Some participants who underspent their income by less than \$65 may have done so strategically (see appendix B). However, 105 of all participants (30%) ended the game with a surplus greater or equal to \$65, which could have been spent to earn additional points without any penalty. Therefore, the observed underspending at the aggregate level is not easily rationalized or explained by strategic motivations. As expected, balance exceeded income on most days after the first day (81%).

Trial-Level Income Surpluses (H2). Income surpluses occurred on 71% of days. H2 predicts higher levels of proportional attention to income will be associated with greater instances of income surpluses. This increasing pattern is visualized in figure 2.5, which plots the mean proportion of income surpluses across 10 levels of proportional attention to income.

FIGURE 2.5: PROPORTION OF INCOME SURPLUSES (STUDY 2)



NOTE—Collapses observations into 10 bins of trial-level proportional attention to income, ranging from [0, 0.10), [0.10, 0.20), … , [0.90, 1.00]. Error bars are trial-level standard errors calculated across participants within each bin.

Our formal analysis regressed an indicator variable for income surpluses on the set of trial-level predictors, discussed previously. The linear probability model[12] identified a significant coefficient on the proportional attention to income variable (b = 0.206, se = 0.031, $t(217)$ = 6.57, p < .001), which was not qualified by a significant interaction with income (p = .110) (table 2.1). This suggests greater attention to income is associated with an increased likelihood of spending less than income on a given day, after accounting for the daily draw of income, the subject-level average attention, total income, and the day of the game. The model predicts about a 20-percentage point greater likelihood of an income surplus for someone who exclusively attends to income versus someone who exclusively attends to balance. As expected, there is also a similarly-sized positive association between a participant's average proportional attention to income (a model covariate) and the likelihood of income surplus (b = 0.179, se = 0.034, $t(82)$ = 5.25, p < .001).

Spending Discontinuities (H2). The preceding analyses consider a binary measure of income surpluses. This measure is appropriate given the hypotheses of underspending available funds to avoid the psychological pain of debt. However, this approach is not sensitive to whether consumers underspend their available funds by a small or large margin, which may be useful to understand. Furthermore, we were not able to reliably consider balance surpluses (as previously

[12] For ease of interpretation, we present coefficient estimates from a linear probability model. We will also report the preregistered logistic regression results in our supplementary analyses (see appendix B).

discussed; also see appendix B). Therefore, we consider an additional exploratory method to visualize and analyze surpluses for both income and balance.

TABLE 2.1: TRIAL-LEVEL REGRESSION RESULTS (STUDY 2)

	Dependent variable:	
	Income Surplus	Spending
	(1)	(2)
Prop. Attn. to Income (MC)	0.206***	−102.757***
	(0.031)	(15.678)
Income (MC)	0.001***	0.512***
	(0.0002)	(0.099)
Avg. Prop. Attn. to Income	0.179***	−43.469***
	(0.034)	(8.590)
Total Income	−0.00001	0.062***
	(0.00004)	(0.010)
Prop. Attn. to Income (MC) x Income (MC)	0.002	−0.908+
	(0.001)	(0.517)
Constant	0.847***	−66.762*
	(0.118)	(29.998)
Observations	6,040	6,040
R^2	0.071	0.066
Adjusted R^2	0.068	0.063
Residual Std. Error (df = 6015)	0.447	218.924
Note:	+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001	

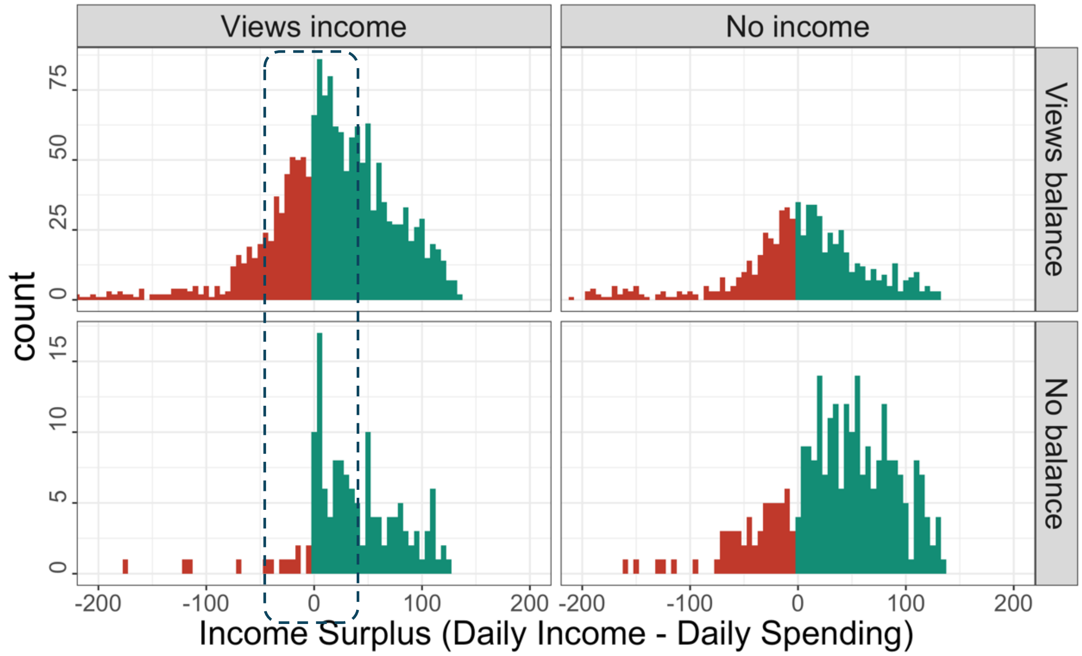
NOTE—Trial-level regression results for the two dependent variables. Standard errors are cluster-robust.

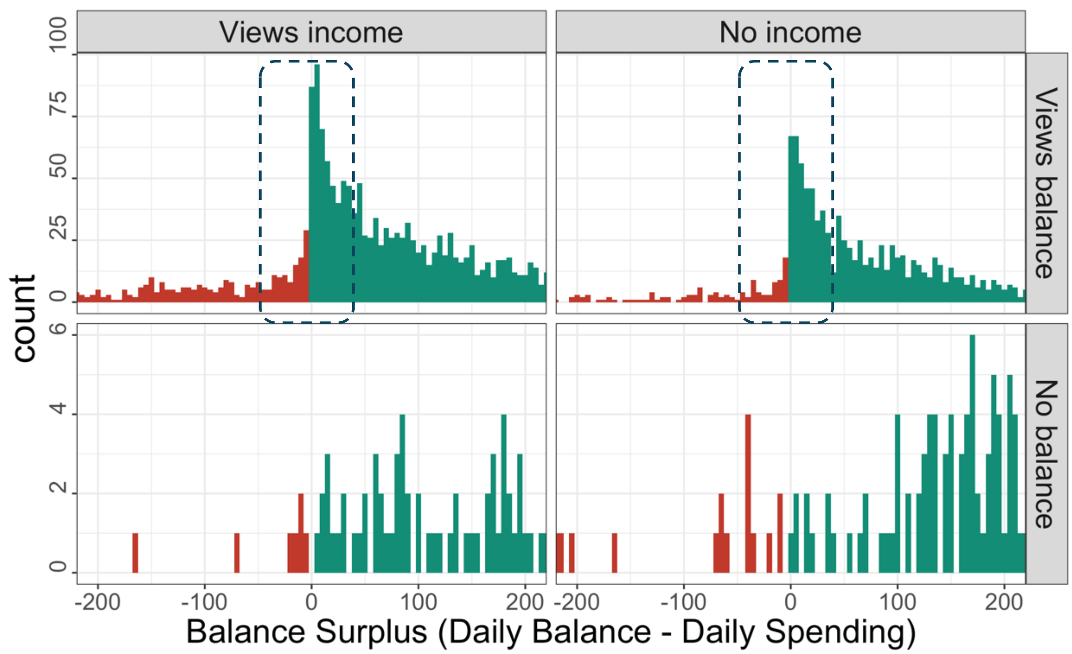
We visualize daily surpluses for both income and balance by considering whether or not participants view each type of financial information on days with non-zero spending (figures 2.6 and 2.7). The visual discontinuities are apparent and consistent with our predictions.

There are discontinuities at zero in the income surplus distributions when participants viewed income but not when they did not view income. Beyond visual inspection, we can test this discontinuity using the bunching analysis procedure detailed in Allen et al. (2017). We use bins of $5, a smoothing window of [-100, 100], and a bunching window of [0, 20], and we find consistent results. There was bunching as expected for income surpluses just above 0 when participants looked at income (total: excess mass = 28%, 95% bootstrapped CI = [11%, 47%]; income-only: 89%, [11%, 175%]), but not when participants did not look at income (total: 15%, [-1%, 34%]; balance-only: 15%, [-4%, 36%]).

Similarly, there are discontinuities at zero in the balance surplus distributions when participants viewed balance but not when they did not view balance. Using the analogous analysis, there was bunching as expected for balance surpluses just above 0 when participants looked at balance (total: 101%, [82%, 122%]; balance-only: 116%, [87%, 148%]), but no bunching or less bunching when participants did not look at balance, although these latter intervals are non-informative due to small samples (total: 47%, [-22%, 247%]; income-only: 55%, [-64%, 705%]).

FIGURES 2.6 (UPPER) AND 2.7 (LOWER): INCOME AND BALANCE SURPLUS DISTRIBUTIONS



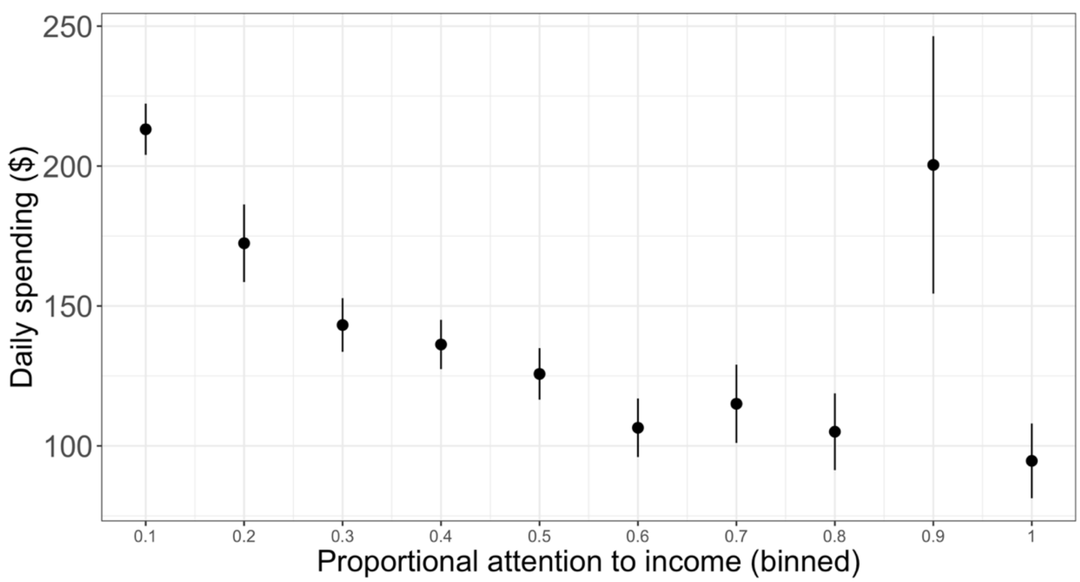


NOTE—Visualizations of the income surplus (upper) and balance surplus (lower) for observations categorized by whether they viewed income (left column) or not (right column), and whether they view balance (top row) or not (bottom row). Non-negative surpluses are green and negative surpluses (deficits) are red. These plots only include data from days with a non-zero amount of spending. We are especially interested in the discontinuities highlighted by the dashed outlines.

Aggregate-Level Surpluses (H2). Examining the aggregate data, we find a consistent pattern of results. Controlling for the total amount of income received, participants higher in average proportional attention to income throughout the game incurred a higher proportion of income surpluses ($b = 0.121$, $se = 0.026$, $t(347) = 4.60$, $p < .001$) (table 2.2).

Trial-Level Spending (H3). Does attention to income versus balance relate to how much participants spend on a given day? Specifically, H3 predicts relatively greater attention to income should be associated with less spending because income is systematically smaller than balance in our experimental game (income < balance on 81% of days). This is indeed what we observe when considering the relationship between proportional attention to income and daily spending (figure 2.8).

FIGURE 2.8: DAILY SPENDING



NOTE—Collapses observations into 10 bins of trial-level proportional attention to income, ranging from [0, 0.10), [0.10, 0.20), … , [0.90, 1.00]. Error bars are standard errors calculated across participants within each bin.

To formally test this relationship, we consider spending as the dependent variable in a linear regression using the set of preregistered trial-level predictors. Higher proportional attention to income was associated with significantly less spending (b = -102.76, se = 15.68, $t(211) = -6.55$, $p < .001$) (table 2.1). This represents a $103 difference in expected spending between two people with an average draw of income, where one attends exclusively to income and the other attends exclusively to balance (while also accounting for the nonsignificant interaction, p = .081, subject-level average attention and income, and the day of game).

Aggregate-Level Spending (H4). While the trial-level analyses detect a negative relationship between proportional attention to income and daily spending, daily patterns do not guarantee patterns at the aggregate level. This is an important nuance for understanding and interpreting the relationships at the daily level. One possibility is that participants remember their accumulation over time (e.g., when they underspend by $10, and then by $0, and then by $5, they keep a running tab of their $15 of accumulation). If this is the case, participants in our incentivized game should ultimately spend these accumulated funds before ending the game. Therefore, we might observe a general tendency of attention relating to spending (H3) without observing inadvertent accumulation at the end of the game. This is not the case, as we do find evidence of inadvertent accumulation (H4).

Higher average proportional attention to income was associated with less spending throughout the course of the game (b = -315.48, se = 78.68, $t(347) = -4.01$, $p < .001$) (table 2.2); this represents a $315 difference in expected spending between two people with an average draw of income, where one attends exclusively to income and the other attends exclusively to balance.

As expected, total income is highly predictive of spending (*b* = 1.09, *se* = 0.12, $t(347) = 9.10$, $p < .001$).

TABLE 2.2: AGGREGATE-LEVEL REGRESSION RESULTS (STUDY 2)

	Dependent variable:	
	Income Surplus	Spending
	(1)	(2)
Avg. Prop. Attn. to Income	0.121***	−315.476***
	(0.026)	(78.683)
Total Income	−0.00000	1.091***
	(0.00004)	(0.120)
Constant	0.688***	−289.590
	(0.120)	(360.359)
Observations	350	350
R^2	0.058	0.219
Adjusted R^2	0.052	0.215
Residual Std. Error (df = 347)	0.097	291.315
Note:	+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001	

NOTE—Aggregate-level regression results for the preregistered dependent variables.

Game Performance. Related to aggregate spending, we conduct exploratory analyses of the relationships between attention and both score and bonus payments (including total income as a covariate). As expected, there are negative associations between proportional attention to income and both score (*b* = -1169.62, *se* = 181.08, $t(347) = -6.459$, $p < .001$) and the bonus payment (*b* = -0.084, *se* = 0.017, $t(347) = -4.87$, $p < .001$). These results align with the aggregate-level spending analysis and suggest participants who attended relatively more to income deviated further from the target spending, as reflected by both score and bonuses. Given these incentivized mechanics, this lends further support for our interpretation of the accumulated funds as unexpected and inadvertent.

Discussion

The experimental game in study 2 provides participants with funds, and assortment of options to buy, and a simple spending goal. Within this paradigm, we observe how participants navigate information and decisions over time. We neither force nor prescribe a specific type of information, and participants are free to attend to either income or balance as they see fit.

Within this paradigm, we observe several patterns that are consistent with our hypotheses. Higher proportional attention to income is associated with more income surpluses at the daily and aggregate levels. Model-free evidence and bunching analyses suggest there may be spending discontinuities right at the zero-surplus level for participants who view income or balance. All of these results support H2.

We also observe a relationship between proportional attention to income and decreased daily spending. This is consistent with H3, as income was smaller than balance on 81% of days. Therefore, we explain this result as participants who attend to income systematically perceive a lesser amount of available funds, leading to lesser spending. This daily-level pattern persists in the aggregate spending data. This is important, because it suggests daily underspending was inadvertent (H4).

The study 2 paradigm allows us to observe spending decisions over time while participants actively allocate their attention between income and balance information. One trade-off of the paradigm is that while we benefit from the flexibility of endogenous attention, we are limited in our ability to make causal claims (within this paradigm). While study 1 demonstrates a causal effect through a different study design generally resembling a stimulus-based choice, it remains useful to understand the direction of causality in study 2, which involves aspects of memory-based choice over the set of repeated decisions (e.g., Lynch and Srull 1982). Furthermore, the imbalanced allocation of endogenous attention imposed an unanticipated constraint on certain analyses (i.e., very few observations in which balance was not attended to; see figure 2.7). Therefore, we conducted study 3 to explore causality (within the study 2 paradigm) as well as to ensure a balanced distribution of (manipulated) attention to income and balance.

STUDY 3: INCENTIVIZED GAME WITH EXOGENOUS ATTENTION

Study 3 uses the same experimental game developed for study 2, except we independently manipulate attention to income and balance information. This allows for (1) separate measures of attention to income and attention to balance, (2) while clarifying the direction of causality between income and spending, and (3) ensuring roughly equal numbers of observations for attending to versus not attending to income and balance information.

Method

Participants. We recruited 479 participants from AMT to take part in this experiment in exchange for $1.65 plus the opportunity to earn an additional bonus payment. A total of 53 participants were removed following our preregistered plan to exclude subjects who dramatically overspent or underspent[13], as this is likely a signal of inattention or misunderstanding the game dynamics. This is the same exclusion rule implemented (but not preregistered) in study 2. The game failed to fully load for one additional participant, who was therefore removed from the data. This resulted in a final sample of 425 participants (M_{age} = 38.15; 47% female).

Manipulating Availability of Financial Information. All aspects of the game were identical to study 2, except for the manipulation of available financial information. Whereas study 2 measured attention to either income or balance information, study 3 manipulated the availability of such information and thus, we argue, attention to it. Participants were assigned across a 2 (Income Information: available, absent) x 2 (Balance Information: available, absent) factorial design. Condition assignment determined whether each piece of financial information was available when participants made their consumption decisions. We equate the availability of information with attention to information. While attention and availability should be extremely closely aligned when only one piece of information is available, we recognize the cases when both or neither pieces of information are available are more opaque. Study 2 suggests people overwhelmingly look at balance over income; therefore, the same pattern may arise when both are available. When neither piece of information is available, participants are not attending to daily variation in income or balance; instead, they must use some alternative strategy (e.g., using average daily income as a spending target).

[13] We preregistered to excluded participants who spent less than \$1,500 or more than \$4,500 during the course of the 20-day game. All participants were informed that their expected earnings (therefore target spending) during the game was \$3,000.

When available (based on condition assignment), income and balance information were presented in the top-left and top-right corners of the screen, respectively. As in the prior study, the instructions provided sufficient information about expected income for a reasonable strategy, regardless of condition (see appendix B). Even in the no-information condition, participants knew they could expect to earn $3000 over the course of the game.

This experimental design avoids prior issues of endogeneity between attention and spending. Here, we exogenously manipulate whether participants *can* attend to either type of information; we expect people *will* attend to whatever is presented.

Analysis Plan

The same core set of models from study 2 are used in study 3. Again, we consider four analyses: the trial- and aggregate-level analyses for income surpluses and spending. In study 3, we include a measure of balance surplus as an exploratory measure. Unlike study 2, which used a proportional measure of attention (which is difficult to interpret for balance surpluses; see appendix B), study 3 randomly assigns attention at the participant level. The effect of condition assignment on balance surpluses is clearer to interpret, so we include balance surpluses as an exploratory measure in our analyses at both the trial and aggregate level.

All predictors and model specifications remain the same[14], except (1) we no longer have a measure of participant-level average attention, and (2) we use a set of dummy-coded condition

[14] We did not preregister the participant's total income as part of the trial-level model. We include it for these analyses to align studies 1 and 2. Excluding this variable in study 1 or including it in study 2 does not meaningfully change our pattern of results.

variables to capture attention to income vs. balance. Specifically, we construct a dummy variable for the condition with only income information available (*Income Only*), only balance information available (*Balance Only*), and no information available (*No Information*). This leaves the condition with both available income and balance information as the reference group. We use the case of full information as the reference group because it allows for easy comparison against the two conditions that are of greatest interest: Income Only (the comparison represents making balance information unavailable) and Balance Only (the comparison represents making income information unavailable). The model also compares the reference group against the case in which no information is available. Though the No Information condition may be informative, we note that participants assigned to this condition may adopt a qualitatively different game strategy than those in the other three cells, who react to daily information.

Results

Earning and Spending. Participants earned an average of $3,004 during the 20-day game (SD = $128) and spent a total of $2,828 ($SD$ = $507), on average. As in the prior study, this reflects a significant degree of underspending relative to realized income ($b = -176$, $se = 25$, $t(424) = -7.16$, $p < .001$). 45% of participants ended the game with a balance surplus greater or equal to $65 (the cost of the lowest-priced item), meaning this group of participants could have earned more points by making at least one additional purchase.

Trial-Level Income Surpluses (H2). Participants ran income surpluses on 73% of days. According to H2, the frequency of these surpluses should be highest when participants attend relatively more to income. Within the context of the 2 x 2 experimental design, attention to income should be highest in the Income Only condition, followed by the reference-group condition in which both types of information are available, followed by the conditions in which only balance information is available and no information is available. Given these different levels of manipulated attention to income, we expect a positive effect of being in the Income Only condition on the likelihood of maintaining an income surplus. Similarly, we expect a negative effect of being in the Balance Only condition. This is what we observe. Compared to having both income and balance information, having only income information increased the likelihood of maintaining an income surplus ($b = 0.047$, $se = 0.016$, $t(209) = 2.94$, $p = .004$) and having only balance information decreased the likelihood of maintaining an income surplus ($b = -0.042$, $se = 0.013$, $t(223) = -3.12$, $p = .002$). There was no additional effect of being assigned to the No Information condition ($b = -0.004$, $se = 0.015$, $t(205) = -0.25$, $p = .799$), relative to the full information condition (table 2.3).

Trial-Level Balance Surpluses (H2). Because of some cases of extreme balances, we limited the sample to trials for which balances were in [$0, $1000] and included balance as a covariate. There were fewer balance surpluses when only income information was available, compared to when all information was available ($b = -0.047$, $se = 0.019$, $t(190) = -2.50$, $p = .013$). There was no effect of having only balance information ($b = 0.006$, $se = 0.015$, $t(208) = 0.38$, $p = .705$), but there was an effect of having no information ($b = -0.101$, $se = 0.018$, $t(165) = -5.46$, $p < .001$). We interpret these results as suggesting an effect of balance information on the likelihood of balance surpluses—regardless of whether in isolation or alongside income information—thus providing additional support for H2.

TABLE 2.3: TRIAL-LEVEL REGRESSION RESULTS

	Dependent variable:		
	Income Surplus	Balance Surplus	Spending
	(1)	(2)	(3)
Income Only	0.047**	−0.047*	−5.763+
	(0.016)	(0.019)	(3.064)
Balance Only	−0.042**	0.006	6.009**
	(0.013)	(0.015)	(2.144)
No Information	−0.004	−0.101***	−0.763
	(0.015)	(0.018)	(3.999)
Current Income (MC)	0.001***	−0.0001	0.283+
	(0.0003)	(0.0002)	(0.147)
Total Income	0.0001+	−0.0001**	0.025*
	(0.00005)	(0.00005)	(0.010)
Current Balance		0.0004***	
		(0.00003)	
Income Only x Current Income (MC)	−0.001	−0.0004	0.464+
	(0.0005)	(0.0003)	(0.246)
Balance Only x Current Income (MC)	0.0001	−0.0004	0.375
	(0.0005)	(0.0003)	(0.234)
No Information x Current Income (MC)	0.002**	0.001	−0.502*
	(0.0005)	(0.0004)	(0.207)
Constant	0.552***	1.177***	82.085*
	(0.138)	(0.142)	(33.796)
Observations	8,500	6,787	8,500
R^2	0.037	0.088	0.027
Adjusted R^2	0.034	0.084	0.024
Residual Std. Error	0.438 (df = 8472)	0.300 (df = 6758)	217.862 (df = 8472)
Note:			+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

NOTE—Trial-level regression results for the three dependent variables. Income surplus (column 1) and spending (column 3) were preregistered. Balance surplus (column 2) was exploratory and uses a constrained sample where balance is non-negative and less than $1,000. Standard errors are cluster-robust.

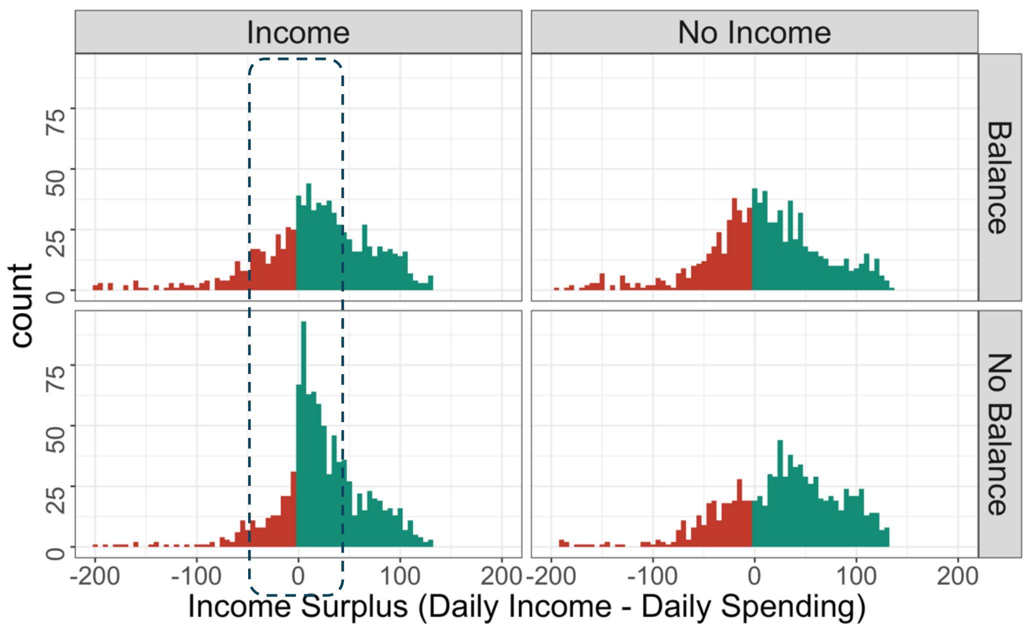
Spending Discontinuities (H2). Following the same approach as in study 2, we again consider the surplus distributions by attention condition (figures 2.9 and 2.10), focusing on trials on which participants spend money. Because we propose attention to either income or balance increases the likelihood of setting that amount as a limit, manifesting in a non-negative surplus, we expect discontinuities around zero surplus. The visual discontinuities are apparent and generally consistent with our predictions.

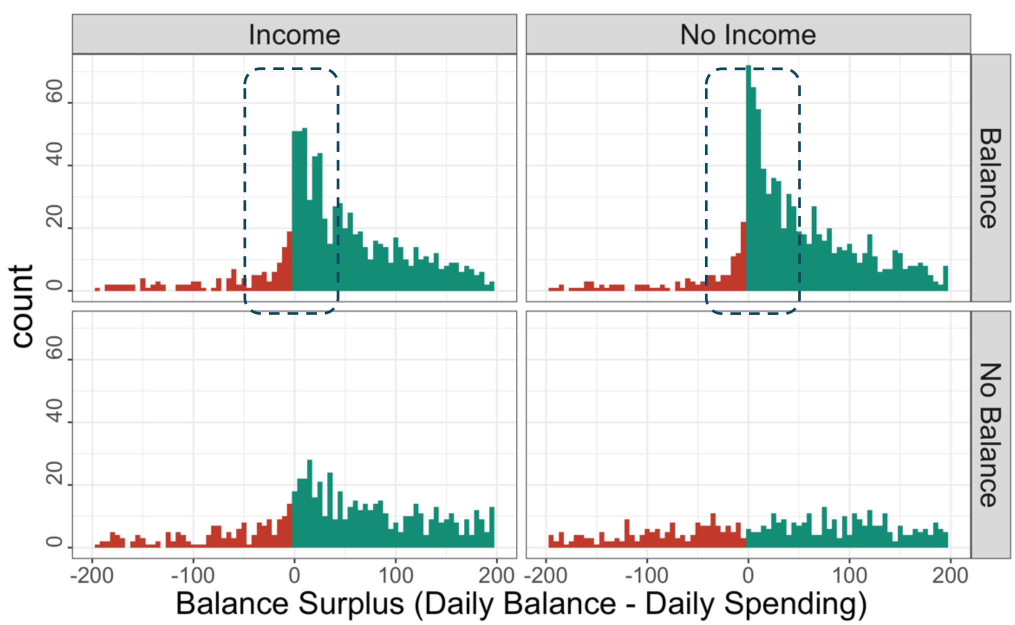
There are discontinuities at zero in the income surplus distributions when participants viewed income (total: excess mass = 57%, 95% bootstrapped CI = [37%, 80%]; income-only: 83%, [51%, 117%]), but not when they did not view income (total: -9%, [-23%, 5%]; balance-only: 5%, [-15%, 27%]). Similarly, there are discontinuities at zero in the balance surplus distributions when participants viewed balance (total: 89%, [70%, 108%]; balance-only: 98%, [74%, 124%]), but less-so when they did not view balance (total: 28%, [1%, 61%]; income-only: 52%, [16%, 96%]). In this latter case, the excess mass is potentially attributable to the ability to roughly track balance over time using income and spending information.

Aggregate-Level Surpluses (H2). The estimated effects of condition on surpluses were similar at the aggregate level of analysis (table 2.4). Compared to seeing all information, there was a positive effect of attending to only income (b = 0.047, se = 0.016, $t(420)$ = 2.96, p = .003) and a negative effect of attending only to balance (b = -0.042, se = 0.015, $t(420)$ = -2.78, p = .006) on the average proportion of income surpluses. As with the trial-level analyses, there was no aggregate-level effect of being assigned to the No Information condition (b = -0.004, se = 0.015, $t(420)$ = -0.25, p = .804). When considering balance surpluses, there was an aggregate-level effect of *not* attending to balance information; there were fewer balance surpluses for those in the Income Only condition (b = -0.088, se = 0.042, $t(420)$ = -2.07, p = .039) and the No

Information condition (b = -0.230, se = 0.043, t(420) = -5.38, p < .001). There was no difference for those in the Balance Only condition (b = -0.066, se = 0.041, t(420) = -1.60, p = .110).

FIGURES 2.9 (UPPER) AND 2.10 (LOWER): INCOME AND BALANCE SURPLUS DISTRIBUTIONS





NOTE—Visualizations of the income surplus (upper) and balance surplus (lower) across the conditions assigned in this experiment. Income information was either available (left column) or not (right column), and balance information was either available (top row) or not (bottom row). Non-negative surpluses are green and negative surpluses (deficits) are red. We plot only data on days with a non-zero amount of spending. We are especially interested in the discontinuities around zero, highlighted by the dashed outlines.

TABLE 2.4: AGGREGATE-LEVEL REGRESSION RESULTS

	Dependent variable:		
	Income Surplus	Balance Surplus	Spending
	(1)	(2)	(3)
Income Only	0.047**	−0.088*	−115.260+
	(0.016)	(0.042)	(68.243)
Balance Only	−0.042**	−0.066	120.186+
	(0.015)	(0.041)	(66.280)
No Information	−0.004	−0.230***	−15.268
	(0.016)	(0.043)	(68.750)
Total Income	0.0001*	−0.0001	0.502**
	(0.00004)	(0.0001)	(0.189)
Constant	0.470***	1.120**	1,317.704*
	(0.130)	(0.354)	(568.820)
Observations	425	425	425
R^2	0.080	0.069	0.043
Adjusted R^2	0.072	0.060	0.034
Residual Std. Error (df = 420)	0.114	0.310	498.039
Note:	+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001		

NOTE—Aggregate-level regression results for the three dependent variables. Income surplus (column 1) and spending (column 3) were preregistered variables. Balance surplus (column 2) was exploratory.

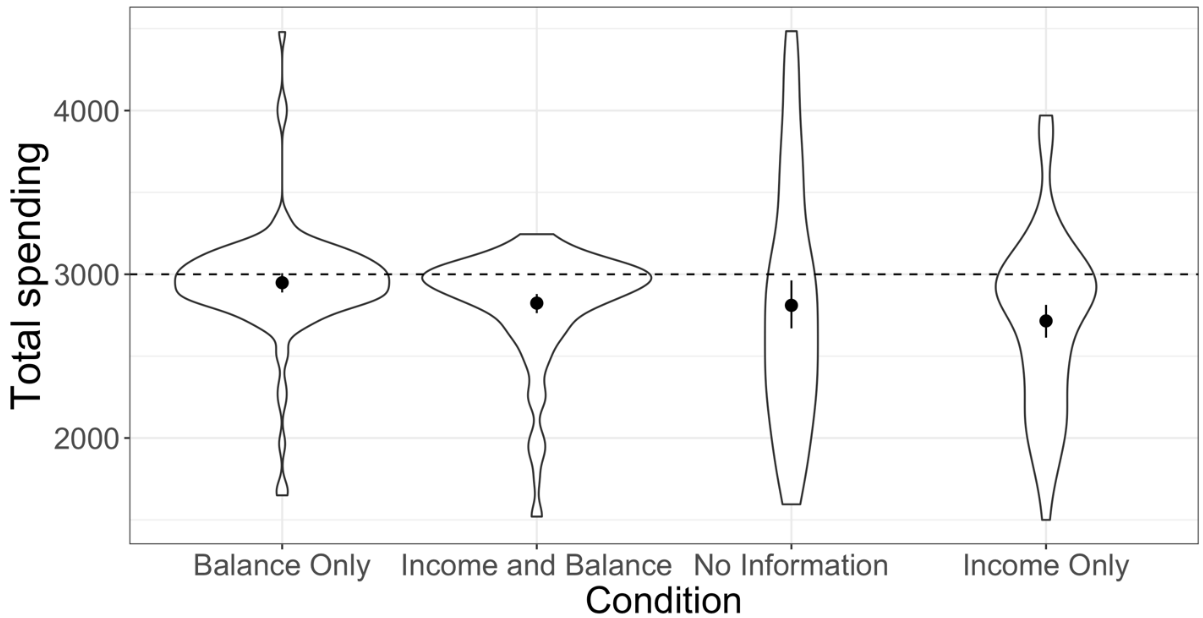
Trial-Level Spending (H3). Do participants who see income information spend less?

There is a marginally significant negative effect of having only income information on spending (b = -5.76, se = 3.06, t(209) = -1.87, p = .062). Furthermore, there is a positive effect of having only balance information on spending (b = 6.01, se = 2.14, $t(224)$ = 2.80, p = .006). There is no difference between having full information and having no information (b = -0.763, se = 4.00, $t(205)$ = -0.19, p = .849) (table 2.4).

Aggregate-Level Spending (H4). Average total spending was highest when only balance information was provided (M = \$2,949, SD = \$332) and lowest when only income information was provided (M = \$2,715, SD = \$529). Spending was somewhere in between when both income and balance were available (M = \$2,824, SD = \$337) or no information was available (M = \$2,810, SD = \$738). These distributions are shown in figure 2.11. We formally assess spending across these groups using the aggregate-level model, controlling for total income. Compared to having both types of information, spending was marginally lower in the Income Only condition (b = -115.26, se = 68.24, $t(420)$ = -1.69, p = .092) and marginally higher in the Balance Only condition (b = 120.19, se = 66.28, $t(420)$ = 1.81, p = .071). There was no effect of being in the No Information condition (b = -15.27, se = 68.75, $t(420)$ = -0.22, p = .824) (table 2.4).

Game Performance. Similar to the prior study, game scores and bonus payments were lower among participants who attended more to income and less to balance. Scores and bonuses were greatest when only balance information was available (M_{score} = 5,971, SD_{score} = 701; M_{bonus} = \$0.12, SD_{bonus} = \$0.07), compared to when both were available (M_{score} = 5,874, SD_{score} = 814; M_{bonus} = \$0.11, SD_{bonus} = \$0.07), or when only income information was available (M_{score} = 5,250, SD_{score} = 1013; M_{bonus} = \$0.05, SD_{bonus} = \$0.07). Scores and bonuses were lowest in the condition with no information (M_{score} = 5,134, SD_{score} = 911; M_{bonus} = \$0.04, SD_{bonus} = \$0.07). This pattern is consistent with a negative effect of not having balance information (*p*s < .001 for Income Only and No Information condition for both score and bonus).

FIGURE 2.11: TOTAL SPENDING BY CONDITION (STUDY 3)



NOTE—Total spending by condition, arranged in order of descending mean total spending. Mean total income was $3,004 (dashed horizontal line). Violin plots depict the distribution of spending observations around the estimate of the mean with a 95% confidence interval.

Discussion

This experiment extends the study 2 results by independently manipulating attention to income and balance through the availability or absence of information. By exogenously manipulating income and balance separately across a 2 x 2 design, the current study investigates the independent causal role of attention to income and attention to balance on spending behaviors. The results are broadly consistent with those from the prior study, increasing our confidence in the set of results as well as the results from study 2.

Compared to study 2, the study 3 design both strengthens and weakens our ability to test H2. The major benefit is that we move away from the endogenous proportional attention measure, which allows us to better consider balance surpluses. Specifically, we observe that not having balance information (as in the Income Only and No Information conditions) leads to fewer balance surpluses. However, there was no effect of having both income and balance information versus having only balance information. This seems initially at odds with our account, because having both income and balance information should *decrease* the relative attention to balance, compared to having only balance. But we cannot fully know this, without measures of attention. Hence, the exogenous manipulation of attention in study 3 also weakens our ability to fully test H2, as we cannot be sure whether participants who were presented with both income and balance *actually attended* to both, or perhaps *only* attended to balance.

We find general support for H3 and H4. Compared to having both income and balance information, providing only income led to less spending and providing only balance led to more spending at both the daily and aggregate levels of analysis (though we note several of these effects were marginally significant; see tables 2.3 and 2.4). But the general pattern persists: Attending more to income leads to marginally less spending, and attending more to balance leads to more (or marginally more) spending.

Studies 2 and 3 are important complements. One measures and the other manipulates attention. Across both studies, we observe broadly consistent results.

GENERAL DISCUSSION

Consumers who desire the pleasure of consumption but want to avoid the pain of present or future debt must take an active role in managing their finances. Rather than spending freely, consumers may calibrate their spending against some measure of available funds. The current research considers how the use of either income or balance information as a measure of available funds affects spending. Specifically, we consider how the mechanisms of accumulation contribute to different—and unintended—spending patterns.

Discussion of Findings

We focus on how attention to income or balance affects consumption through the selection of spending limits. We propose attending to income increases the likelihood of perceiving income as a measure of available funds (and the equivalent process for balance) (H1). To test this, we ask participants to identify their current level of constraint and later had them self-code their response as pertaining to income or balance (study 1). An implication of perceiving either income or balance as a measure of available funds is that consumers should systematically attempt to underspend that amount (to avoid the psychological pain of debt) (H2). We observe this in study 1, where participants who think in terms of income ($1,000) disproportionately avoid spending $1,200 on a car maintenance service, compared to participants who think in terms of balance ($3,625). We corroborate this finding with several analyses regarding income and balance surpluses from studies 2-3, in which attention to either income or balance tends to be positively associated with spending less than that amount.

A key difference between flows and stocks is that they differ in magnitude, and we find that attention to the larger metric (balance) is associated with more spending in our experimental games (H3). In our incentivized paradigm, this amounts to greater attention to income being associated with suboptimally less spending (compared to greater attention to balance) (H4). This reflects a fundamental aspect of flows versus stocks, which facilitates small daily differences accumulating into large unexpected differences: Flows "forget" past decisions and stocks "remember" them.

Contributions

These findings offer multiple contributions to the literature. First, we advance the understanding of how consumers establish their spending constraints. Whereas much of the prior work has focused on limiting consumption (Krishnamurthy and Prokopec 2010; Shefrin and Thaler 1988; Thaler 1980, 1985, 1999; Wertenbroch 1998) or allocation of available funds through budgets or spending plans (Antonides, Manon de Groot and Fred van Raaij 2011; Heath and Soll 1996; Stilley, Inman and Wakefield 2010; Zhang et al. 2022; Zhang and Sussman 2018a), we seek to clarify how consumers think about their available funds in the first place. We find consumers represent their available funds—at least in part—as they are presented, as either income or balance. Whereas consumers could translate income to balance or balance to income, our findings suggest this is not the case.

Second, we contribute to the literature on accumulation and stock-flow reasoning by demonstrating the role of attention—particularly proportional attention—on the use of stock versus flow information. Whereas prior literature considers the importance of either stock versus flow presentation form on judgments and decisions (Cronin et al. 2009; Fischer et al. 2015; Kainz and Ossimitz 2002; Korzilius et al. 2014; Newell et al. 2016; Ossimitz 2002; Reinholtz et al. 2021; Spiller et al. 2020; Sweeney and Sterman 2000), we consider how the allocation of attention to stock versus flow information impacts judgments and decisions (study 2). This distinction is meaningful for situations in which people may be exposed to both stock and flow information. Our results suggest the more impactful presentation form will be the one that is relatively more attended to.

Third, we contribute to the literature on consumer spending. We find evidence that consumer spending is sensitive to the seemingly innocuous, perhaps arbitrary presentation of financial information. While income and balance are equally capable of conveying financial information (e.g., study 1) and both commonly used and preferred by consumers (descriptive survey), they differ in magnitude and memory. In single-shot decisions, the magnitude differences may imply differences in available funds, hence affecting spending now (H3). Over time, any prior underspending or overspending will be less apparent if attending to flows, rather than stocks. For this reason, we predict (and observe) that attending to income leads to more unexpected accumulation trajectories, over time (H4).

Implications

Every consumer learns about their finances through the information that is provided to them. Therefore, financial service providers might reevaluate whether their communication strategies are likely to help or hinder clients in achieving their financial goals. As previously discussed, whether an income or balance presentation leads to more spending is likely to depend on the relative magnitudes of the two metrics. For constantly indebted or low-balance consumers, highlighting income might encourage more spending in the short term (through perceiving more available funds). For high-balance consumers, the effect should run in the opposite direction. However, in the long-term, consumers who attend to income information may unexpectedly accumulate or decumulate resources. Understanding whether this accumulation or decumulation is desirable to clients is also context dependent and merits further exploration.

Limitations

Generalizability. As previously discussed, the observed effects are likely context dependent. Different consumers segments who differ along dimensions of income size, income variability, liquid balances, and savings goals may have different spending responses to income versus balance information. However, the previously stated goal of this research is not to prescribe a one-size-fits-all intervention. Instead, the purpose of the present research is to identify that there are two common ways consumers learn about their finances, and attention to one versus the other has a cascading set of implications for how consumers perceive constraint, attempt to constrain spending to avoid debt, how much they spend, and how accurately they manage their spending, over time. Therefore, while we caution that the results from our spending game are not intended to be generalizable, the theory (in the form of H1-H4) should be. We leave it to the information architects who understand their target audience and the segments they serve to apply our hypotheses to make specific predictions.

Spending vs. Financial Management. Although the game captures behavior over a number of trials, it is focused on repeated spending with limited opportunity for consumers to "take a step back" and consider management of their overall finances (e.g., through an annual budget-setting exercise). In addition to the impact of attention on spending we document here, attention also affects budget-setting (e.g., Mrkva & Van Boven, 2017). Some of the accumulation effects we observe may be different if consumers periodically assess their overall finances, leading to reevaluation of available funds. Given that such reevaluation is also likely sensitive to such attentional effects, this may change, rather than reduce, the cumulative effects of attention.

Small Incentives. The maximum achievable bonus in the spending game was $0.20. This is an admittedly modest amount, and future research may potentially benefit from raising the stakes to feel more consequential. Yet, reviews of the literature find little to no evidence that incentive sizes alone predict decisions or performance in experimental contexts (Camerer et al. 1999; Carpenter, Verhoogen and Burks 2005), and participation in a portfolio of similar experiments on AMT over time would lead repeated earning vs. foregoing this sort of bonus to be equivalent to about $1 per hour in earnings. An important distinction in our sample may be between excluded and included participants. Across both studies, 11% of participants were excluded on the basis of game performance (spending < $1,500 or spending > $4,500). By virtue of such erratic spending behavior, none of these participants earned any bonus. Therefore, it is plausible that a sizeable proportion of these participants were *not* motivated by the small stakes of our game. However, considering the participants included for analysis, their behaviors seem consistent with the incentivized, point-maximizing strategy.

Conclusion

The current work sheds further light on attention to different sources of funds and consumer spending behavior. It suggests minor details that direct attention to or from specific financial forms can play a major role in how consumers assess their funds and spend. This identifies a potential opportunity for information architects working with banks and other financial service providers to guide attention in ways that will help consumers towards their stated financial goals.

Chapter 2: Appendix

for

Consumer Responses to Income Versus Balance Information

APPENDIX A: ADDITIONAL MATERIALS FOR STUDY 1

Question Wordings and Response Rates for Comprehension Questions:

Question 1: "Recall: According to the statement, what was the balance at the start of the statement period (Jan. 15)?"

- *$575*
- *$1,075*
- *$1,575*

Question 2: "Recall: Which of the following accurately captures the financial scenario you saw?"

- A beginning balance of $1,075 grew to a closing balance of $3,625 ($4,000 in income and $1,450 in expenses)

- A beginning balance of $1,075 grew to a closing balance of $4,725 ($5,000 in income and $1,350 in expenses)
- A beginning balance of $1,075 grew to a closing balance of $2,825 ($2,000 in income and $250 in expenses)

Response Rates

Question 1. A total of 94 participants answered question 1 incorrectly (44 income, 55 balance). There was no significant difference in incorrect response rate across conditions ($t(497) = 0.621$, $p = .535$).

Question 2. A total of 94 participants answered question 1 incorrectly (58 income, 57 balance). There was no significant difference in incorrect response rate across conditions ($t(497) = 0.179$, $p = .858$).

Question 1 or Question 2. A total of 181 participants answered either question 1 or question 2 incorrectly (86 income, 95 balance). There was no significant difference in incorrect response rate across conditions ($t(497) = 0.736$, $p = .462$).

APPENDIX B: ADDITIONAL MATERIALS FOR STUDY 2

The Spending Game was designed in Qualtrics and is available to run online, including through standard survey and microtask platforms (e.g., AMT).

Goals of the Spending Game

Participant Perspective. The spending game is intended to be engaging and strategic. We make the game engaging through the use of button-clicking to simulate work, familiar consumer categories and items to resemble normal household shopping, and dynamic points (utility) over time.

Researcher Perspective. This paradigm offers a rich set of data including daily draws of income, daily balances, total daily spending, the items purchased, the order of items purchased, and measured (study 2) or manipulated (study 3) attention.

Game Mechanics. All information necessary to play the game strategically was given in the pre-game instructions (table 2.B1, bolded columns). Therefore, we have various exploratory ways of comparing in-game performance to normative benchmarks (in addition to comparing spending across conditions).

Total Costs Equal Expected Income. By design, the total costs of all 12 items sums to $3,000, which is the precise amount of income earned during the game (in expectation).

Point Decay. To simulate the diminishing marginal utility of consumption, we programmed points to decay exponentially with each purchase. Specifically, there was a 10% decay of point value with each purchase (rounded to the nearest whole number). Examples of point decay are given by the five decay columns in table 2.B1.

Optimal Strategy. "Solving" the game requires considering the ratio of points to dollars (accounting for decay) and spreading total income across items to maximize points. This ratio is calculable given information about costs and points, provided in the instructions. Column 2 (table 2.B1) lists these ratios and shows an increasing pattern, such that greater value is derived from more expensive purchases. This column alone is insufficient for optimal gameplay. To find the optimal allocation, a player must verify whether there are incentives to purchase repeat items, given the 10% point decay. A savvy player may quickly verify there is never an incentive to purchase repeat items by comparing the lowest starting point-to-dollar ratio (2, for the $65 item) to the highest point-to-dollar ratio for a repeat purchase (1.998, for the $590 item: 1179/590). Therefore, optimal gameplay conditional upon expected income is simply spreading allocation across all items by purchasing one of each.

TABLE 2.B1: POINT STRUCTURE OF THE SPENDING GAME

Cost ($)	Points/$	**Points (starting)**	Points *Decay 1*	Points *Decay 2*	Points *Decay 3*	Points *Decay 4*	Points *Decay 5*
65	2	**130**	*117*	*105*	*95*	*85*	*77*
75	2.02	**152**	*136*	*123*	*110*	*99*	*89*
85	2.04	**173**	*156*	*140*	*126*	*114*	*102*
95	2.06	**196**	*176*	*159*	*143*	*128*	*116*
130	2.08	**270**	*243*	*219*	*197*	*177*	*160*
145	2.1	**305**	*274*	*247*	*222*	*200*	*180*
160	2.12	**339**	*305*	*275*	*247*	*223*	*200*
175	2.14	**375**	*337*	*303*	*273*	*246*	*221*
450	2.16	**972**	*875*	*787*	*709*	*638*	*574*
490	2.18	**1068**	*961*	*865*	*779*	*701*	*631*
540	2.2	**1188**	*1069*	*962*	*866*	*779*	*702*
590	2.22	**1310**	*1179*	*1061*	*955*	*859*	*773*

NOTE—Cost and point structure of the Spending Game. Bolded columns were known to participants at the start of the game. Non-bolded columns were not known but could be imputed at the start of the game. The decay columns represent a 10% exponential decay from the prior point value.

Balance Surpluses

Analyzing balance surpluses presents three unique challenges.

The first challenge involves the relationship between balance surpluses and balance itself. Whereas the income surplus compares spending only against a daily draw of income, the balance surplus comparison includes prior accumulation. This is the aspect of "memory" that is unique to balance. Therefore, the likelihood of maintaining a balance surplus should relate to both the daily draw of income and the amount of prior accumulation. We account for this by controlling for daily balance when assessing balance surpluses. However, the relationship between balance and balance surpluses is necessarily non-linear when considering negative balances (a negative balance of any size guarantees a deficit, rather than surplus). To address this, we constrain our analysis of balance surpluses to observations with non-negative balances. Similar concerns arise when balances are sufficiently large that any spending is unlikely to affect the likelihood of surplus. We further limit our balance surplus analyses to the 96% of observations in which balance is less than or equal to $1,000. Constraining our sample in this way ($\$0 \leq$ balance $\leq$ $\$1,000$) restricts our analysis to regions where we expect the relationship between balance (as a covariate) and balance surpluses (as the dependent variable) to be approximately linear.

The second challenge arises when balance is larger than income. When balance > income (as it generally would be for a debt-averse participant), then an income surplus *necessarily* implies a balance surplus. Therefore, to the extent that proportional attention to *income* is positively associated with the likelihood of running an income surplus (H2), proportional attention to *income* will also be positively associated with the likelihood of running a balance surplus. Yet, because proportional attention to *balance* is (1 – proportional attention to income), this implies proportional attention balance may be negatively associated with the likelihood of a balance surplus; this is the opposite of what is predicted by H2. Therefore, we recognize our exploratory analyses of balance surpluses may be limited in their ability to detect the hypothesized effect.

Analyses

We use a nearly identical approach in assessing our exploratory balance surplus measure. For reasons previously discussed, the trial-level model controls for balance and limits the analysis to the range of the data where the relationship between balance and balance surplus is most likely to be linear ($\$0 \leq$ balance $\leq \$1,000$). The aggregate-level model is unchanged from the version preregistered for the other two dependent variables, aside from the naming convention of proportional attention to balance.

Trial-Level Balance Surpluses. The parallel analysis on balance surpluses does not detect a meaningful relationship between the proportional attention to balance and the likelihood of a balance surplus (b = -0.021, se = 0.021, t(195) = -0.98, p = .326). There is suggestive evidence of the expected positive relationship between a participant's average proportional attention to balance and their likelihood of incurring a balance surplus (b = 0.079, se = 0.041, t(88) = 1.93, p = .057) (table 2.B2). We do not find evidence that the likelihood of a balance surplus is greater when participants attend relatively more to balance (H2). However, we note four reasons why this may not be surprising.

First, balance surpluses depend on not only the random draw of income, but also prior accumulation. While we attempt to control for this through the inclusion of balance as a covariate, there are issues of endogeneity. These may be particularly concerning if participants deploy intertemporal spending strategies (e.g., deliberately accumulating balance today with the expectation of making a large purchase tomorrow) or if balance can be somewhat accurately inferred without attention (e.g., balance today can be predicted from balance yesterday).

Second, balance exceeded income on the majority of days (77% in total; 81% when ignoring the first day, on which both measures are guaranteed to be equivalent). On such days, any income surplus mechanically generates a balance surplus. Therefore, to the extent that proportional attention to income is associated with more income surpluses, it is necessarily associated with more balance surpluses. Recognizing the proportional attention to balance is 1 –

proportional attention to income (because they are the same measure of proportion), this implies proportional attention to balance may be negatively associated with balance surpluses.

Third, the issue of balance regularly exceeding income should be more pronounced when subjects attend relatively more to income and therefore spend less (H3). This is because attending to income is predicted to decrease spending, thus leading to larger accumulated balances over time. Larger balances are more likely to remain in surplus, thus relating systematic attention to income with an increased likelihood of balance surplus. Equivalently, because greater proportional attention to balance is associated with less savings, systematically smaller balances may be overspent, reflective of a negative association between attention to balance and balance surpluses.

Fourth, alternative methods of analysis—such as the bunching analysis (Allen et al., 2017)—may help to disentangle the relationship between measures of attention and either income or balance surpluses. However, the described approach is statistically constrained by the limited number of observations in which participants do not attend to balance information. We experimentally manipulate attention in study 2 to address this and other concerns.

TABLE 2.B2: ADDITIONAL REGRESSION ANALYSES

	Dependent variable:		
	Income Surplus	Balance Surplus	Spending
	(1)	(2)	(3)
Prop. Attn. to Income (MC)	0.206^{***}		-102.757^{***}
	(0.031)		(15.678)
Prop. Attn. to Balance (MC)		-0.021	
		(0.021)	
Income (MC)	0.001^{***}	-0.0002^{+}	0.512^{***}
	(0.0002)	(0.0001)	(0.099)
Avg. Prop. Attn. to Income	0.179^{***}		-43.469^{***}
	(0.034)		(8.590)
Avg. Prop. Attn. to Balance		0.079^{+}	
		(0.041)	
Total Income	-0.00001	-0.0001	0.062^{***}
	(0.00004)	(0.0001)	(0.010)
Prop. Attn. to Income (MC) x Income (MC)	0.002		-0.908^{+}
	(0.001)		(0.517)
Balance		0.0003^{***}	
		(0.00004)	
Prop. Attn. to Balance (MC) x Income (MC)		-0.0002	
		(0.001)	
Constant	0.847^{***}	1.042^{***}	-66.762^{*}
	(0.118)	(0.170)	(29.998)
Observations	6,040	5,214	6,040
R^2	0.071	0.058	0.066
Adjusted R^2	0.068	0.053	0.063
Residual Std. Error	0.447 (df = 6015)	0.278 (df = 5188)	218.924 (df = 6015)
Note:			+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

NOTE—Trial-level regression results for the three dependent variables. Income surplus (column 1) and spending (column 3) were preregistered. Balance surplus (column 2) was exploratory and uses a constrained sample where balance is non-negative and less than $1,000. Standard errors are cluster-robust.

CHAPTER 2 REFERENCES

Ainslie, George (1975), "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control," *Psychological Bulletin*, 82(4), 463.

Allen, Eric J., Patricia M. Dechow, Devin G. Pope, and George Wu (2017), "Reference-Dependent Preferences: Evidence from Marathon Runners," *Management Science*, 63(6), 1657–72.

Amasino, Dianna R., Nicolette J. Sullivan, Rachel E. Kranton, and Scott A. Huettel (2019), "Amount and Time Exert Independent Influences on Intertemporal Choice," *Nature Human Behaviour*, 3(4), 383–92.

Ando, Albert and Franco Modigliani (1963), "The Life Cycle Hypothesis of Saving: Aggregate Implications and Tests," *American Economic Review*, 53, 55–84.

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

Ashraf, Nava, Dean Karlan, and Wesley Yin (2006), "Household Decision Making and Savings Impacts: Further Evidence from a Commitment Savings Product in the Philippines."

Browning, Martin and Thomas F. Crossley (2001), "The Life-Cycle Model of Consumption and Saving," *Journal of Economic Perspectives*, 15(3), 3–22.

Camerer, Colin F. and Robin M. Hogarth (1999), "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty*, 19, 7-42.

Carpenter, Jeffrey, Eric Verhoogen, and Stephen Burks (2005), "The Effect of Stakes in Distribution Experiments," *Economics Letters*, 86(3), 393–98.

Cronin, Matthew A., Cleotilde Gonzalez, and John D. Sterman (2009), "Why Don't Well-Educated Adults Understand Accumulation? A Challenge to Researchers, Educators, and Citizens," *Organizational Behavior and Human Decision Processes*, 108(1), 116–30.

Fiedler, Susann and Andreas Glöckner (2012), "The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis," *Frontiers in Psychology*, 3(OCT).

Fischer, Helen, Christina Degen, and Joachim Funke (2015), "Improving Stock-Flow Reasoning With Verbal Formats," *Simulation and Gaming*, 46(3–4), 255–69.

Fisher, Geoffrey (2017), "An Attentional Drift Diffusion Model over Binary-Attribute Choice," *Cognition*, 168, 34–45.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue (2002), "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature*, 40(2), 351–401.

Glickman, Moshe, Orian Sharoni, Dino J. Levy, Ernst Niebur, Veit Stuphorn, and Marius Usher (2019), "The Formation of Preference in Risky Choice," *PLoS Computational Biology*, 15(8).

Heath, Chip (1995), "Escalation and De-Escalation of Commitment in Response to Sunk Costs: The Role of Budgeting in Mental Accounting," *Organizational Behavior and Human Decision Processes*, 62(1), 38–54.

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

Helion, Chelsea and Tom Gilovich (2014), "Gift Cards and Mental Accounting: Green-lighting Hedonic Spending," *Journal of Behavioral Decision Making*, 27(4), 386–93.

Hershfield, Hal E., Daniel G. Goldstein, William F. Sharpe, Jesse Fox, Leo Yeykelis, Laura L. Carstensen, and Jeremy N. Bailenson (2011), "Increasing Saving Behavior Through Age-

Progressed Renderings of the Future Self," *Journal of Marketing Research*, 48 (SPL), S23–37.

Hirst, D. Eric, Edward J. Joyce, and Michael S. Schadewald (1994), "Mental Accounting and Outcome Contiguity in Consumer-Borrowing Decisions," *Organizational Behavior and Human Decision Processes*, 58(1), 136–52.

Kainz, Daniel and Günther Ossimitz (2002), "Can Students Learn Stock-Flow-Thinking? An Empirical Investigation," *Proceedings of the 20th International Conference of the System Dynamics Society*, (January 2002), 1–34.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman (2016), "Getting to the Top of Mind: How Reminders Increase Saving Center," *Management Science*, 62(12), 3393–3411.

Korzilius, Hubert, Stephan Raaijmakers, Etiënne Rouwette, and Jac Vennix (2014), "Thinking Aloud While Solving a Stock-Flow Task: Surfacing the Correlation Heuristic and Other Reasoning Patterns," *Systems Research and Behavioral Science*, 31(2), 268–79.

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

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.

Lynch, Jr., John G. and Thomas K. Srull (1982), "Memory and Attentional Factors in Consumer Choice: Concepts and Research Methods," *Journal of Consumer Research*, 9(1), 18.

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.

Newell, Ben R., Arthur Kary, Chris Moore, and Cleotilde Gonzalez (2016), “Managing the Budget: Stock-Flow Reasoning and the CO2 Accumulation Problem,” *Topics in Cognitive Science*, 8, 138–59.

O’Donoghue, Ted and Matthew Rabin (2000), “The Economics of Immediate Gratification,” *Journal of Behavioral Decision Making*, 13(2), 233–50.

Okada, Erica Mina (2002), “Trade-Ins, Mental Accounting, and Product Replacement Decisions,” *Journal of Consumer Research*, 27(4), 433–46.

Olafsson, Arna and Michaela Pagel (2018), “The Liquid Hand-to-Mouth: Evidence from Personal Finance Management Software,” *The Review of Financial Studies*, 31(11), 4398–4446.

Ossimitz, Günther (2002), “Stock-Flow-Thinking and Reading Stock-Flow-Related Graphs: An Empirical Investigation in Dynamic Thinking Abilities,” *Proceedings of the 2002 international systems dynamics conference.* Albany, NY: Systems Dynamics Society, 2002.

Prelec, Drazen and George Loewenstein (1998), “The Red and the Black: Mental Accounting of Savings and Debt,” *Marketing Science*, 17(1), 4–28.

Reinholtz, Nicholas, Sam Maglio, and Stephen A. Spiller (2021), “Stocks, Flows, and Risk Response to Pandemic Data,” *Journal of Experimental Psychology: Applied*, 27(4), 657.

Rramani, Qëndresa, Ian Krajbich, Laura Enax, Lisa Brustkern, and Bernd Weber (2020), “Salient Nutrition Labels Shift Peoples’ Attention to Healthy Foods and Exert More Influence on Their Choices,” *Nutrition Research*, 80, 106–16.

Shefrin, Hersh M. and Richard H. Thaler (1988), “The Behavioral Life Cycle Hypothesis,” *Economic Inquiry*, 26(4), 609–43.

Smith, Stephanie M. and Ian Krajbich (2018), “Attention and Choice across Domains,” *Journal of Experimental Psychology: General*, 147(12), 1810–26.

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

Soman, Dilip and Min Zhao (2011), “The Fewer the Better: Number of Goals and Savings Behavior,” *Journal of Marketing Research*, 48(6), 944–57.

Soster, Robin L., Andrew D. Gershoff, and William O. Bearden (2014), “The Bottom Dollar Effect: The Influence of Spending to Zero on Pain of Payment and Satisfaction,” *Journal of Consumer Research*, 41(3), 656–77.

Soster, Robin L., Ashwani Monga, and William O. Bearden (2010), “Tracking Costs of Time and Money: How Accounting Periods Affect Mental Accounting,” *Journal of Consumer Research*, 37(4), 712–21.

Spiller, Stephen A., Nicholas Reinholtz, and Sam J. Maglio (2020), “Judgments Based on Stocks and Flows: Different Presentations of the Same Data Can Lead to Opposing Inferences,” *Management Science*, 66(5), 2213–31.

Stilley, Karen M, J. Jeffrey Inman, and Kirk L Wakefield (2010), “Spending on the Fly: Mental Budgets, Promotions, and Spending Behavior,” *Journal of Marketing*, 74(3), 34–47.

Sweeney, Linda Booth and John D. Sterman (2000), “Bathtub Dynamics: Initial Results of a Systems Thinking Inventory,” *System Dynamics Review*, 16(4), 249–86.

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 Shlomo Benartzi (2004), "Save More Tomorrow ™: Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, 112(S1), 164–87.

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

Weber, Megan E., Stephen A. Spiller, Hal E. Hershfield, and Suzanne B. Shu (2025), "The Role of Stock-Flow Reasoning in Understanding the Social Security Trust Fund," *Working Paper.*

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

Wertenbroch, Klaus, Dilip Soman, and Amitava Chattopadhyay (2007), "On the Perceived Value of Money: The Reference Dependence of Currency Numerosity Effects," *Journal of Consumer Research*, 34(1), 1–10.

West, Colin, Ashley Whillans, and Sanford DeVoe (2020), "Income Volatility Increases Financial Impatience," Cambridge, MA: Harvard Business School.

Yang, Xiaozhi and Ian Krajbich (2021), "Webcam-Based Online Eye-Tracking for Behavioral Research," *Judgment and Decision Making*, 16(6), 1485–1505.

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

CHAPTER 3: AVERAGE VALUE AFFECTS CONSUMER BUDGETS

With Stephen A. Spiller

ABSTRACT

Consumers frequently use budgets to manage their spending. Money in budgets is treated as though it is not fungible, so budget allocations matter. What determines budget allocations? The authors propose budgeters are sensitive to the *average value* of the set of products constituting a budget category. This represents a departure from normative budget setting, which is based solely on products' *marginal value*. Results indicate budget allocations are indeed sensitive to the average value of products within a budget category, beyond what can be explained by standard economic theory. This finding is unique to budgeting, such that budgeters are more sensitive to average value than are non-budgeters. Consequently, the mere act of budgeting affects the composition of purchases, even when spending levels and preferences remain unchanged.

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 consumers' perceptions of value (average value versus marginal value) guide these budget allocations.

We propose consumers will be sensitive to a category's average value when setting budgets. As a result, they will allocate more funds to the categories they perceive as more valuable, on average. This contrasts with the standard economic model, which contends marginal value is the basis for normative approaches to budgeting and consumption. Marginal value is the incremental value of the next-best, most-valuable available option. The principle of marginal analysis guides normative theories of budgeting and purchasing. The simple rule to maximize value is to select the options providing the greatest marginal value. Despite this straightforward approach prescribed by normative theories, we argue that consumers (and the budgets they set) will be sensitive to the average value of budget categories.

Why might budgets deviate from the value-maximizing principle presumed to guide both budget setting and ordinary purchase behavior? We propose that the process of setting a budget encourages category-level evaluations, in which consumers extract and make decisions based on summary information (i.e., average values). This differs from marginal analysis, which involves evaluating the value of individual items. As a result, we expect budget allocations to reflect the average value of budget categories, whereas purchase decisions should reflect the marginal values of individual items. The implication is that the mere act of budgeting (versus purchasing) alters how consumers assess value. And because budget allocations are sticky, consumers with identical preferences and identical levels of spending 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 introduce and distinguish two different measures of value: average and marginal. This is followed by two studies designed to assess budgeters' consideration of average value in addition to marginal value when allocating budgets. Study 1 surveys consumers about their actual budgeting behavior. Respondents described their own discretionary budgets and indicated the extent to which those budgets reflected both average value and marginal value. Study 2 measures participants' preferences and valuations of various activities and examines how average and marginal value relate to hypothetical budget allocations. Next, we characterize why these findings are unique to budgeting, as opposed to general consumer decision making. Specifically, the task of budget allocation engages relatively more category-level evaluation, compared to ordinary purchase decisions. We present evidence for this shift in evaluation mode through a supplemental experiment involving stimuli from Amazon.com. We then present Study 3, which uses a tightly controlled experimental game to manipulate average category values while holding marginal value constant. The set of results suggests that allocation is uniquely sensitive to a category's average value, above and beyond marginal value. This has 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

The word "budget" is commonly used in everyday language to refer to myriad forms of planning, behavior, and constraint for both individuals and organizations. We adopt the definition of budgeting as a two-stage process involving: (1) the allocation of funds (planning how to spend), and (2) the subsequent spending of those funds (Heath and Soll 1996; Lukas and Howard 2023; Thaler 1985; Zhang and Sussman 2018). 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. This form of budgeting resembles how many consumers divide and spend their discretionary funds, whereby they first decide how much to allocate and later decide exactly what to buy. The current research focuses on this form of budgeting. Note that two-stage budgeting is distinct from automatic bill payments, which are more accurately described as automated transactions.

Budgets can be used to manage both the total level of spending as well as the composition of spending across budget categories. Whereas other consumer research considers how the act of budgeting affects the *level of spending* (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*. In particular, for a given level of total consumption, we focus on how funds are allocated and used across different budget categories. A representative survey conducted by Zhang et al. (2022) finds that consumers use categories to organize their budgets. 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 such categories (appendix A). These findings underscore the importance of understanding how consumers allocate their budgets across specific categories, separate from their overall level of constraint.

The Effect of Budgeting on Consumption. Budgeting matters 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 affects consumption in multiple ways. 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 within-category spending. But 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 entertainment if they have two budgets devoted to entertainment and food (where food encompasses both groceries and dining out), than if they have three budgets devoted to entertainment, 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 the greatest value. That is, each of these literatures implicitly or explicitly acknowledge that consumers allocate more to budget categories that are perceived as more valuable. But how do consumers assess the value of budget categories? This is the question we seek to answer.

Assessing the Value of Budget Categories

Consider the task of setting a budget for discretionary entertainment expenses. Given some set of considered options (e.g., bowling, going to the movies, mini golf, attending a concert, and taking a pottery class), how valuable is this budget category? One approach to answering this question is to consider the value offered by each individual prospect. For a consumer who highly values bowling and mini golf but does not value the other activities enough to purchase them, the value of the budget category should reflect anticipated consumption of bowling and mini golf. This approach is consistent with valuing budget categories in terms of *marginal value*.

An alternative approach is to zoom out and consider the average value of the set of entertainment options. That same consumer might perceive this set to have a medium value on average (high-value items like bowling and mini golf balanced against low-value items like movies, concerts, and pottery classes). This describes how consumers might think about budget categories in terms of *average value*.

Normative models of decision making suggest consumers think in terms of marginal value, which is the incremental value of an additional unit of consumption. In the preceding example, this entails iteratively asking "how much value would I get out of my favorite available entertainment option from the set I have not yet decided to purchase?" If a consumer does indeed consider value in this way, then the "marginal principle" from economics prescribes making consumption decisions according to these marginal values (Samuelson and Nordhaus 2009; Colander 2019). Strict adherence to this principle ensures consumers will get the "best bang for the buck" by only making purchases that confer the best possible value, relative to the alternatives. The consumer who derives value from paying to go bowling and mini golfing but does not derive value from going to a concert will consume the former activities but not the latter. The marginal principle is a powerful model of decision making that implicitly guides the standard assumption that consumption follows preferences.

Yet, we suggest budgeters in particular might not think and act in terms of marginal value alone, but rather average value as well. Drawing upon prior work in ensemble perception and the evaluation of sets, we expect consumers to value budget categories based on the *average value* of their options. Furthermore, we expect budgeters to allocate in accordance with this value.

The key insight driving these predictions is that budgeting is distinct from ordinary purchasing because it encourages the organization of individual purchases (e.g., bowling, going to the movies, mini golf, attending a concert, and taking a pottery class) into a set of purchases (e.g., entertainment options). This subtle yet meaningful shift in how options are represented (as a budget category, rather than individual items) should encourage consumers to focus on the average value of the category rather than the marginal values of the items.

Extracting Averages. People automatically extract average (mean)[15] representations from sets, categories, and groups of people with little to no effort (Ariely 2001; Haberman and Whitney 2009; Whitney and Yamanashi Leib 2018; Yamanashi Leib et al. 2020). This has been demonstrated in relatively simple contexts (e.g., basic visual perception involving size, color, and motion paths; Ariely 2001; Chong and Treisman 2003; Watamaniuk and Duchon 1992) and with complex assortments (Whitney and Yamanashi Leib 2018; Woiczyk and Le Mens 2021). For example, people quickly and automatically extract average features and expressions from group of faces (Haberman and Whitney 2007, 2009), and consumers extract the average value from assortments of products (Yamanashi Leib et al. 2020).

Decisions Based on Averages. The literature reviewed above indicates people quickly assess and encode the average representation of a set, including its average value. We expect budgeters will similarly assess the average value of a budget category. Though extracting an

[15] Ensemble perception generally discusses "average" as the mean representation of a set. While it is beyond the scope of the current research to consider other measures of central tendency (e.g., median and mode), we discuss some of the implications of sensitivity to different distributional features in the general discussion.

average representation does not guarantee it will serve as the basis for downstream decisions, there is ample evidence across related disciplines that averages indeed inform decision making. In studies of distributed choice (i.e., consumption over time), people tend to choose in proportion to the average long-run benefits they receive from those choices (Davison and McCarthy 1988; Herrnstein and Prelec 1991; Herrnstein et al. 1993; McDowell 2013; Rachlin and Laibson 1997). Facing nonlinear cost structures, people often make economic decisions based on average costs (Liebman and Zeckhauser 2004). This tactic is 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). For example, energy consumption is sensitive to changes in average price when marginal price is held constant (Ito 2014; Shin 1985), and tax expectations reflect average rather than marginal tax rates (Rees-Jones and Taubinsky 2020).

Consumers may also occasionally 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 2024).

Average Value Diverges from Marginal Value. Evidence from across disciplines suggests there are occasions on which people make decisions based on averages rather than marginal analysis. This raises the possibility that budgeting may also be sensitive to average values, thus deviating from purely marginal thinking and decision making. We suggest it is, because budgets are themselves sets of options, organized within a categorical structure. Such cognitive representation aligns with ensemble perception, which favors representing category values by their mean. It is important to understand how consumers perceive value when making allocation decisions because budgets based on average value versus marginal value can diverge in important ways. This can be the case even for consumers with identical preferences, facing identical option sets, and spending equivalent amounts of money. Consider that average value reflects the evaluation of the entire category, which might be pulled up by high-valued options or dragged down by low-valued options. When considering average value—as we propose budgeters do—composition of the full set matters. This is not the case for marginal value, which does not depend on the entirety of the set, but rather on the most-valued items. Because average value and marginal value differ, if budgeters rely on average value, their allocations will systematically deviate from allocations based on marginal value. And because budgeting is a two-stage process in which allocations guide downstream consumption, any sensitivity to average value while setting a budget will have a downstream effect on spending from a budget.

H1: Budget allocations are sensitive to a category's average value, above and beyond the category's marginal value.

H2: Because budget allocations are sticky, spending from allocated funds will be sensitive to a category's average value, above and beyond marginal value.

We first provide evidence for H1, our central hypothesis, in two studies. Study 1 surveys consumers about their own budgeting behavior. Consistent with H1, consumers report high levels of attention to average category value in describing how they originally set their budgets. Study 2 provides a more nuanced test of H1 by asking study participants to make allocation decisions based on their own preferences for hypothetical vacation activities. After providing evidence that budgeting is sensitive to average value (H1), we next explain why the observed pattern of results should be unique to budgeting—in contrast to purchasing in the absence of budgeting. This argument considers how budgeting and purchasing engage distinct modes of evaluation. A supplementary experiment using stimuli from Amazon.com suggests budgeters evaluate categories relatively more, whereas purchasers evaluate items relatively more. We return to H2—alongside H3 and H4, introduced later—in Study 3, which presents an incentivized consumption game. Key results are summarized in table 3.1, and all preregistrations, materials, data, and code are available at https://researchbox.org/3968&PEER_REVIEW_passcode=EFOMGE.

TABLE 3.1: SUMMARY OF STUDY METHODS AND KEY FINDINGS

Study	Method	*N*	Key Findings
1	Descriptive survey of actual budgeting behavior	100 (100)	(1) Average value is important in setting budgets (H1) (2) Average value is no less important than marginal value (3) The majority of participants rate average value to be at least as important as marginal value
2	Study of real preferences over vacation activities	501 (451)	(1) Average value relates to budget allocations, above and beyond marginal value (H1)
3	Incentive-compatible consumption game	970 (821)	(1) Higher average value draws larger allocations, holding marginal value constant (H1) (2) Higher average value leads to more spending, holding marginal value constant (H2) (3) Budget allocation by budgeters is more sensitive to average value than is spending by purchasers (H3) (4) Spending by budgeters is more sensitive to average value than is spending by purchasers (H4)

Note—Summary of key findings. In column labelled *N*, top number indicates recruited sample size and bottom number, in parentheses, indicates final sample size.

STUDY 1: SURVEY OF BUDGETING BEHAVIOR

Do budgets reflect the value of downstream consumption? If so, is value captured by the category average or value at the margin? To begin to address these questions in a naturalistic fashion, we asked consumers to report and comment on how they set their actual budgets.

Method

We recruited 100 participants from Amazon Mechanical Turk (AMT) to take part in a survey about their own budgeting behaviors (M_{age} = 39; 41% female).[16] In an effort to capture an accurate description of actual behavior, we asked participants to consider their real financial situation when responding to all questions. We then provided participants with a series of questions to elicit their actual budgets, as well as their rationale for their budget structure.

First, participants indicated their total amount of monthly take-home pay (after taxes and deductions). Next, they indicated how much of the take-home pay went towards recurring, essential expenses. The difference of these two amounts reflects monthly discretionary funds, which was displayed to participants.[17] Subsequently, participants were asked to indicate their monthly budgets for discretionary funds across different categories of expenditures. For all participants, these categories included groceries, dining-out, entertainment, and clothing, which

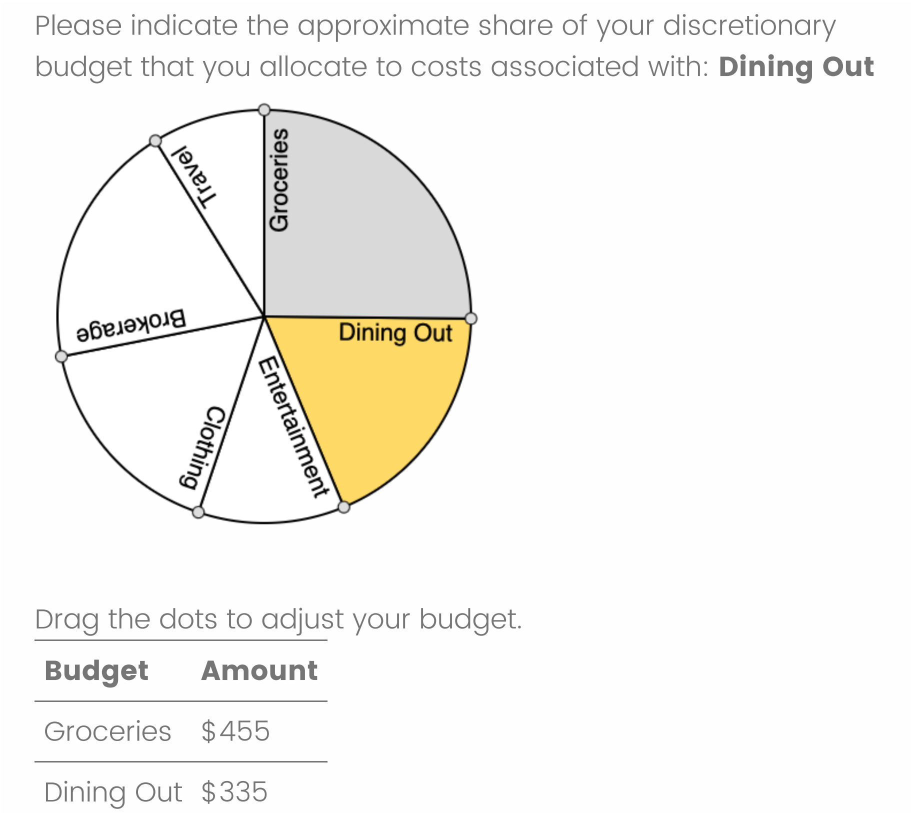
---

[16] There were originally 101 responses, though two responses were linked to the same identifier. Therefore, we removed the second, duplicate observation (i.e., the response that started after the prior one was completed).

[17] For the 18% (18/100) with less than or equal to $100 of monthly discretionary income, we asked them to imagine having $100 for the remainder of the survey. Excluding these participants does not meaningfully affect our findings.

are categories used by a majority of budgeters (Zhang et al., 2022). Participants also had the option to create up to three additional categories.

FIGURE 3.1 EXAMPLE OF BUDGETING PIE CHART (STUDY 1)



Note—This image depicts a participant indicating the size of their Dining Out budget, after previously setting the size of their budget for Groceries, but not yet setting the size of any other budgets. In addition to the area of each slice of the pie chart, participants also saw the dollar amount (rounded to the nearest $5) of each allocation. Groceries, Dining Out, Entertainment, and Clothing were presented to all participants. Brokerage and Travel are depicted here as two custom categories created by this participant.

Next, participants indicated the total dollars of discretionary money allocated to each budget category using a budgeting pie chart. Respondents worked one category at a time, setting their budget by adjusting the area of the pie slice (figure 3.1). Participants always encountered the four default categories (groceries, dining-out, entertainment, clothing) prior to any custom categories. For each of these default categories, we also asked participants on the same page to "briefly describe how you originally settled upon that amount for your budget." Responses were collected in an open-text form, which required at least 25 characters to proceed. The purpose of this exercise was to encourage reflection upon how funds were originally allocated (beyond merely recalling the allocations). After the four default categories, participants continued to set the budget sizes for any ad-hoc categories, though we did not collect open-ended protocols during this portion of the exercise.

After completing the budgeting pie chart, we showed budgeters their prior written responses and asked them to recall the extent to which they focused on average value and marginal value when originally setting their budgets (order counterbalanced). To reduce the likelihood of drawing conclusions based on specific ways of describing abstract concepts such as average value and marginal value, we sampled from a set of four question variants for each construct. All questions were measured on a 1-7 scale, anchored on 1 = "Not at all" to 7 = "Very much." The full set of question variant wordings is provided in table 3.2. (We additionally asked participants to self-report their focus on items and categories. We report those measures in appendix table 3.B1.)

TABLE 3.2: VARIANT WORDINGS (STUDY 1)

When thinking about setting your budget, to what extent did you find yourself…

Measure	Variant wording	Variant id
Average	…thinking about your overall impression of how much you like each category?	1
	…remembering your general liking of each category?	2
	…comparing your overall enthusiasm for each category?	3
	…relying on your general evaluation of each category?	4

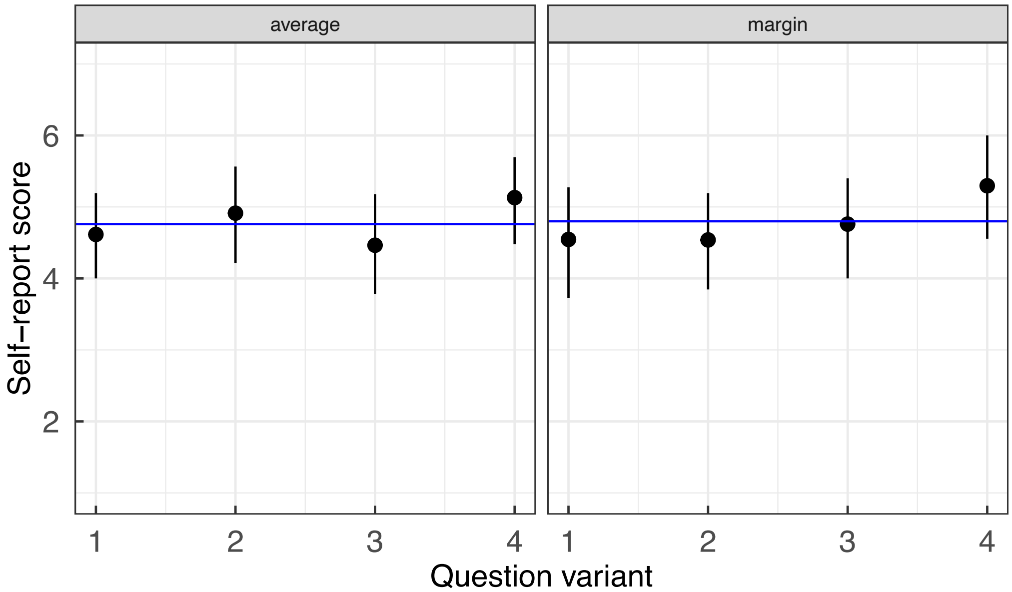
Margin	…thinking about what you could buy with just a little more money (or conversely, what you would lose if you spent a little less)?	1
	…imagining how a small adjustment to one of your budgets could change what you buy?	2
	…weighing the trade-offs between having enough money to buy one thing or the other, but not both?	3
	…considering how giving up one thing might allow you to buy something else?	4

Results

Budgeters reported considering both average value and marginal value when making allocation decisions. As depicted in figure 3.2, both means ($M_{\text{average}} = 4.76$, $SD_{\text{average}} = 1.70$; $M_{\text{margin}} = 4.80$, $SD_{\text{margin}} = 1.90$) were significantly above the midpoint of 4 on a 1-7 scale (both *t*s > 4.21; both *p*s < .001). Though normative principles suggest consumers should consider only value at the margin (Samuelson and Nordhaus 2009; Colander 2019), average value is not considered less than marginal value ($p = .88$). Furthermore, the majority of respondents (61/100) reported considering average value at least as much as marginal value.[18]

[18] 36/100 reported considering average value more than marginal value, 25/100 reported considering the two value metrics equally, and 39/100 reported considering average value less than marginal value.

FIGURE 3.2: SELF-REPORTED FOCUS ON AVERAGE AND MARGINAL VALUE (STUDY 1)



Note—Self-reported focus on the dimensions of average value and marginal value across question variants. Higher scores indicate an increased focus on the specific dimension of value. Error bars are 95% confidence intervals. The solid blue lines independently depict the mean of average value and marginal value. The question variant refers to the "variant id" in table 3.2.

Discussion

In study 1, participants described their budgets and reflected upon how they originally allocated their funds. Across a variety of question wordings, we find consistent evidence that budgeters considered both average value and marginal value when originally setting their budgets. Specifically, we identify three pieces of evidence for the important role of averages: (1) participants considered average value at levels above the scale midpoint; (2) average value is not significantly less considered than marginal value; and (3) the majority of respondents indicated they considered average at least as much as margin. This finding is an important departure from what would be expected if budgets strictly follow the marginal principle. Instead, it appears that budgeters draw upon multiple dimensions of value, including the average value of a budget category (H1).

STUDY 2: SETTING VACATION BUDGETS

Are budgeters sensitive to the average value of consumption options, above and beyond their sensitivity to marginal value? Whereas study 1 presented evidence that budgeters report sensitivity to average value (as well as marginal value), study 2 more strictly tests whether budgets are sensitive to average value, even after accounting for value at the margin. In other words: Study 2 provides a stronger test of sensitivity to average value, above and beyond what can be explained by marginal value (H1). We test this using a paradigm carefully designed to leverage participants' own preferences for a variety of vacation activities.

Method

Participants. 501 participants from AMT completed this study (M_{age} = 43; 37% female).

Design and Stimuli. Participants planned activities for a 3-day vacation for which flights, ground transportation, and an all-inclusive hotel reservation were already booked. They were instructed to budget for additional experiences during the trip and saw a travel brochure of 15 activities popular among tourists. The activities were presented in a random order, and each option cost $30. See figure 3.3. Participants set their budget using a slider snapped to $30 increments, ranging from $0 to $450.

FIGURE 3.3: TRAVEL BROCHURE OF ACTIVITIES (STUDY 2)

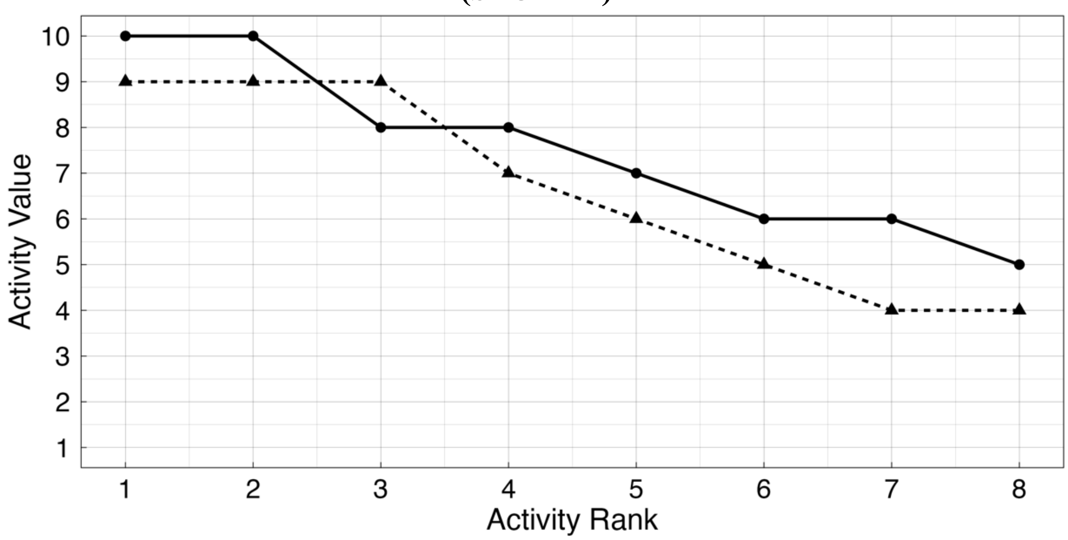
Activities ($30 each)
Running a half marathon
Dancing at a nightclub
Taking a yoga class
Riding the country's fastest roller coaster
Attending a science lecture at the nearby university
Bungee jumping
Visiting the museum of local history
Playing 9 holes of golf
Relaxing in local hot springs
Snorkeling in the reef
Unwinding at late-night whiskey bar
Hiking up a scenic trail
Attending a fancy afternoon tea
Renting a car to sightsee
Playing games in an arcade

After setting their budget, participants sequentially rated the value of each of the 15 activities on a 1-10 scale anchored on 1 = "Very little value" to 10 = "A lot of value." Participants were instructed "if you would never even consider paying for an activity, then click the following button," which was labelled "No Value – Would Not Even Consider." This feature is useful for identifying the considered set of activities—as opposed to the entire set. After *rating* the value of all 15 activities, participants then *ranked* only the considered items by sequentially identifying the best activity until no considered activities remained. Therefore, our key measures are: (1) the budget allocation; (2) the value ratings of each activity; and (3) the rankings of each considered activity, from best to worst. We can use these measures to assess how budget allocation (as the dependent measure) relates to the average of the value ratings, controlling for marginal value (H1). In addition to these key variables, we also observed which activities participants purchased, given their budget allocation. This measure is useful to confirm whether participants' preferences are consistent for both allocation and purchase decisions (they are).

Analysis Plan. Our primary interest is in whether average value guides budget allocation decisions, above and beyond value at the margin. Therefore, we analyze whether allocations are related to the average rated value of activities after controlling for marginal value. We limit our analysis to the values of considered activities (i.e., the rated values of all activities except those participants said they would not consider). Our main analysis considers whether a participant allocates enough money for a particular activity based on its rank, k. For example, we can observe whether a participant budgets for their $k = 4^{\text{th}}$ ranked activity by noting whether the budget is at least \$120 (because activities cost \$30 each). For each participant, for each rank, we consider whether they budgeted enough to purchase that activity. As predictors, we consider marginal value (the rated value of the k^{th} activity) as well as the average value of all other considered options. For additional precision, we distinguish between the average value of all considered options ranked better than the k^{th} activity and the average value of all considered options ranked worse than the k^{th} activity.

As an example, consider the two sets of hypothetical activity ranks and values depicted in figure 3.4. Imagine we are interested in whether a participant budgets enough for the $k = 3^{\text{rd}}$ ranked activity (at least \$90) as a function of (i) the *marginal* value of the 3^{rd} ranked activity and (ii) the *average* value of all other considered activities. At the margin, the value of the $k = 3^{\text{rd}}$ ranked activity is 9 (triangles) or 8 (circles). Normative theories of decision making suggest consumers will be sensitive *only* to this value when deciding whether to budget for the 3^{rd} ranked good. In our example, the prediction based on the marginal principle is that the hypothetical consumer depicted by triangles is more likely to budget for the 3^{rd} ranked activity than the consumer depicted by circles. This is because the incremental benefit of budgeting for 3 activities—compared to 2—only depends on how much additional value is offered by the 3rd ranked option. Any value conferred by the better-ranked options is not relevant in the decision to budget for an additional activity, as these values are already guaranteed with a budget for 2 activities. Similarly, the values of worse-ranked activities (e.g., ranked 4th and worse) do not affect the additional value offered by the 3rd ranked option, as these values will not be realized if the budget does not accommodate the 3rd ranked good.

FIGURE 3.4: ILLUSTRATION OF HYPOTHETICAL VALUE-RANK RELATIONSHIPS (STUDY 2)



Note—Rank-value pairings for two hypothetical respondents (circles and triangles). Both consider 8 items (thus indicating 7 other items would not be considered). When predicting whether each respondent allocates enough for 3 activities, our model considers each participant's value of the 3rd ranked activity (the marginal value) and the average value of activities ranked above the margin (ranks 1, 2) and below the margin (ranks 4-8).

Departing from this normative prediction, we suggest consumers might also be sensitive to the average value of the other considered items (H1). Following the previously described approach, we calculate the average value for activities ranked better than 3 ([9+9]/2 = 9 for triangles; [10+10]/2 = 10 for circles), and the average value for activities ranked worse than 3 ([7+6+5+4+4]/5 = 5.2 for triangles; [8+7+6+6+5]/5 = 6.4 for circles). In this example, marginal value (triangles > circles) and average value (triangles < circles) diverge. We predict participants will be sensitive to average value (H1).

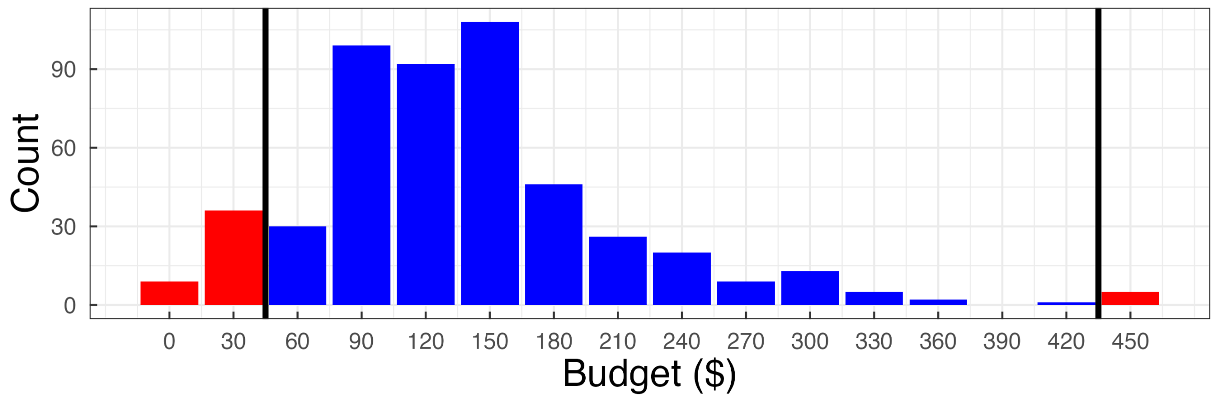
We implement this analysis by constructing a dataset where each observation corresponds to one considered item by one participant. We regress whether a participant has allocated for the k^{th} ranked activity (1 = yes, 0 = no) on the marginal value (the rated value of the k^{th} activity), the average value for activities ranked above *k* (more preferred), and the average value for activities ranked below *k* (less preferred). We include rank-level fixed effects (to account for differences in values across activity ranks) with cluster-robust standard errors (to account for non-independence).[19]

Results

Following our preregistered plan, we excluded 50 respondents who set extreme activity budgets of $0, $30, or $450, as depicted in figure 3.5. Following this exclusion results in a remaining sample of 451 participants.

[19] This represents a deviation from our preregistered analysis. Our preregistered analysis effectively separates this single model into a set of independent regression equations, one for each rank. We present the previously discussed model for ease of explication, while noting the results align closely with those of the preregistered analysis plan and support identical substantive conclusions (see appendix C.1).

FIGURE 3.5: DOLLARS ALLOCATED TO ACTIVITY BUDGET (STUDY 2)



Note—Histogram of budget allocations (in dollars) for vacation activities in study 2. Participants who set extremely small budget ($0, $30) or large budgets ($450) were excluded, following our preregistration. Excluded participants are depicted in red and fall beyond the preregistered cutoffs (vertical lines).

We then used a linear regression (with cluster-robust standard errors) to estimate how marginal value, average value of the above-margin options, and average value of the below-margin options relate to the decision to budget for the k^{th} ranked activity. Consistent with the marginal principle, the value of the marginal good (the value of the k^{th} good) was positively associated with the choice to budget for the k^{th} activity while controlling for rank and average value (b_{margin} = 0.020, *se* = 0.003, $t(280) = 5.93$, $p < .001$).[20] Consistent with H1, average value was also positively associated with the choice to budget for the k^{th} activity (after controlling for marginal value). Specifically, the average value below the marginal k^{th} ranked activity drove this effect (b_{below} = 0.035 *se* = 0.006, $t(202) = 5.56$, $p < .001$). There was no relationship between budget and the average value above the marginal k^{th} ranked activity (b_{above} = 0.001, *se* = 0.009,

[20] Given cluster-robust standard errors, all degrees of freedom are estimated.

$t(148) = 0.09$, $p = .92$).[21] Though the action appears to be driven by sensitivity to average value below the margin, participants are sensitive to the overall average value of all considered activities, after controlling for the value of the marginal k^{th} activity ($b = 0.03$, $se = 0.008$, $t(236) = 3.91$, $p < .001$; appendix 3.C4).

Robustness. One concern involves our interpretation of the coefficient "below" (the average value of activities ranked worse than the marginal activity). We claimed that normatively, the decision of whether to allocate for k activities should not be influenced by the average value below the marginal activity, conditional on the marginal value. This claim holds if the relationship between activity ranks and values is at least weakly monotonically decreasing for all options ranked worse than the marginal option. This is because if any options ranked worse than the k^{th} option had a higher value, then participants should be considering that lower-ranked-but-higher-valued option when decided how to set their budget. In other words, if rated values are not at least weakly monotonically decreasing over activity rankings, then we will misidentify the value of the marginal good. Furthermore, the true marginal value would be inadvertently captured by the average value below the margin.

To address this potential concern, we reconsider our analysis after removing instances of nonmonotonicity. Specifically, we remove any case in which the k^{th} activity (the activity we identify as the marginal good) is rated as less valuable than any worse-ranked option. This ensures we do not consider cases in which the value at the margin is clearly misestimated (and

[21] We preregistered a particular interest in the coefficient on the "above" variable. Unexpectedly, the action was in the coefficient on the "below" variable. We further discuss how average value is perceived across the distribution of values (relative to average value) in study 3 and in the general discussion.

incorporated into the average value of options below the margin). Reconsidering our analysis on this constrained subset of the data (1993/3864 observations) again finds a significant relationship with value at the margin (b_{margin} = 0.031, *se* = 0.006, *t*(177) = 5.13, *p* < .001) and average value below the margin (b_{below} = 0.024, *se* = 0.008, *t*(193) = 2.89, *p* = .004), but not average value above the margin (*p* = .20; appendix C.2).

As additional robustness checks, we consider various alternative model specifications. We include non-considered options in the construction of averages (appendix C.3); we consider the overall average of all non-margin considered items (appendix C.4); and we consider alternative constructions of average value for better- and worse-ranked items (appendix C.5). Results from these additional models do not substantively or statistically change the interpretation of the primary results. Furthermore, we confirm both value and rank are strong predictors of hypothetical purchase decisions, suggesting participant preferences and decisions within the paradigm are internally consistent (appendix C.6).

Discussion

In study 2, we consider how budget allocation decisions relate to both marginal value and average value. This design explores these relationships using only participants' stated preferences and valuations alongside a budget allocation decision. While the decision was hypothetical, the activities were intended to be familiar to participants; and the scenario was intended to resemble the types of contexts budgeters frequently encounter when making allocation decisions in their real lives. Beyond the naturalistic appeal, the design allows us to specifically focus on the values of options within the considered set (i.e., activities that are sufficiently valuable to be considered).

We find evidence that allocations are guided by average value in addition to marginal value (H1). Specifically, participants were more likely to set a budget to accommodate a given activity (of rank *k*) when they perceived greater average value among the considered options, after controlling for marginal value. This sensitivity to average value was driven by the average value among the set of considered items ranked below the marginal purchase. These values—which drive allocation decisions—also matter for spending. We observed a high degree of internal consistency among participants who consistently indicated they would purchase the most-valued and best-ranked activities, given their budgets.

One set of concerns around study 2 is the potential for measurement error. While our design benefits from participants' ability to express their preferences for the various vacation activities, there may be deviations between rated values, ranks, and true underlying preferences. Though our set of robustness checks lends confidence that our interpretation of the results is not driven by specific analysis decisions (appendix C), we acknowledge the potential role of measurement error underlying the valuations used across our analyses. This concern is addressed in study 3, where we use imputed values (rather than elicited values) to test the relationship between category average value and budget allocation.

BUDGETING VERSUS PURCHASING

Studies 1 and 2 provide initial evidence for budgeters' sensitivity to a category's average value—above and beyond marginal value—when making allocation decisions (H1). The implication is that allocations (and downstream consumption) will be sensitive to conditions that affect how a category's value is perceived *on average*. Does this sensitivity to average reflect a unique aspect of budgeting, or is it common to all purchase decisions?

Evaluation Mode. To better understand what makes budgeting unique, consider how allocation differs from purchasing. The task of budget setting is necessarily a decision involving one or more *categories*. Budgeting reorganizes individual purchase opportunities into sets that we call an "entertainment budget," a "vacation budget," or a "discretionary spending budget." Attention to categories naturally facilitates the extraction of averages (Ariely 2001; Haberman and Whitney 2009; Whitney and Yamanashi Leib 2018; Yamanashi Leib et al. 2020), so we expect budgeters will evaluate a category's average value. In contrast, purchasing requires evaluating individual options, which is better-suited to marginal analysis. In the case of purchasing, the consumer need only identify the most valuable option(s), given their constraints. Therefore, budget allocation and purchasing should differ in the mode of evaluation precisely because they are different tasks. The act of budgeting encourages *category-level evaluations*, whereas the act of purchasing encourages *item-level evaluations*. This distinction between category-level and item-level evaluation modes relates to the distinction between direct and derived evaluations (Sood, Rottenstreich and Brenner 2004).

To test whether budgeting and purchasing differ in evaluation mode, we conducted a supplementary experiment using stimuli from Amazon.com (see appendix). Participants (N = 200) identified an Amazon department they were likely to shop from and subsequently saw a matrix of the 30 best-selling items as though they were on Amazon's "Best Sellers" page. We randomly assigned participants to either indicate how much they would budget for these types of products or click on the types of products they would purchase. After engaging in the exercise to either budget or purchase, participants self-reported the extent to which they focused more on category-level or item-level evaluations during the task. Consistent with our theorizing, budgeters reported a relatively stronger focus on category-level evaluations than did purchases ($t(198) = 3.21$, $p = .002$; appendix).

The supplementary experiment provides evidence that allocating and purchasing engage distinct modes of evaluation. This is perhaps unsurprising, given allocating and purchasing are very different tasks. Yet, the very goal of setting a budget is to ultimately guide purchasing. If allocating and purchasing involve different modes of evaluations—as we suggest—then consumers with the same end goal (i.e., making the best purchases) may nevertheless form discrepant evaluations. It is precisely because of the unique evaluations associated with budgeting versus spending that we expect budgets to be sensitive to category average values, more so than purchases. Whereas the nature of budgeting draws attention to categories (therefore encouraging the use of average value), the nature of purchasing does not. This is why the assumption that decision makers rely on marginal value may be poorly suited for budgeting, compared to purchasing. Instead, because budgeters are more likely to engage in category-level evaluation than purchasers, we expect budget-setting to be more sensitive to category averages than purchasing.

H3: Budget allocation decisions are more sensitive to a category's average value than are purchase decisions (in the absence of a budget).

An implication of H3 is that allocations and purchases will differ, even among consumers with identical preferences and financial constraints. This is important, because 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, budgeters first allocate more to categories with higher average value and later adhere to those allocations when spending, making their eventual spending also sensitive to average values. Thus, although H3 implies a comparison of seemingly non-alignable tasks (budget allocations and purchase decisions), it has important downstream consequences. The prediction is that two otherwise identical consumers with the same level of spending will differ in the composition of spending if one allocates prior to purchasing and the other does not.

H4: 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.

In summary, we propose the sensitivity to average value (above and beyond marginal value) documented in studies 1 and 2 reflects a unique aspect of budgeting. Compared to purchasing, budgeting encourages greater category-level evaluations (supplementary study; Appendix). This difference in evaluation mode explains why allocation decisions are especially sensitive to the average value of budget categories. And because budgets are sticky, budgeters consume relatively more from high-value categories compared to purchasers, even at equivalent levels of overall spending. We test these accounts using an incentive-compatible experimental design in which we exogenously manipulate the average value of different budget categories, holding marginal value constant, and observe both budgeting and spending decisions.

STUDY 3: INCENTIVIZED CONSUMPTION GAME

Study 3 has three main goals. First, we test the complete set of hypotheses within a single experimental paradigm: Are budget allocations sensitive to average value (H1)? Is the downstream spending of budgeters sensitive to average value (H2)? Are budget allocations more sensitive to average value than non-budgeted purchase decisions (H3)? Is spending more sensitive to average value when consumers previously set budgets (H4)? Second, we use a highly controlled and incentive-compatible game design that allows us to (i) precisely induce and manipulate average values; (ii) hold marginal value constant by design; (iii) incentivize value-maximization; (iv) observe both allocation and spending decisions; and (v) collect data in an environment with ample opportunity for learning. Third, we consider a different context of budgeting: Dividing funds between distinct budget categories. Whereas the prior studies involved setting a single budget relative to the outside option, participants in the current study explicitly divide spending between two different accounts.

Method

Participants. 970 participants recruited from AMT completed this study (M_{age} = 41; 52%

female).[22]

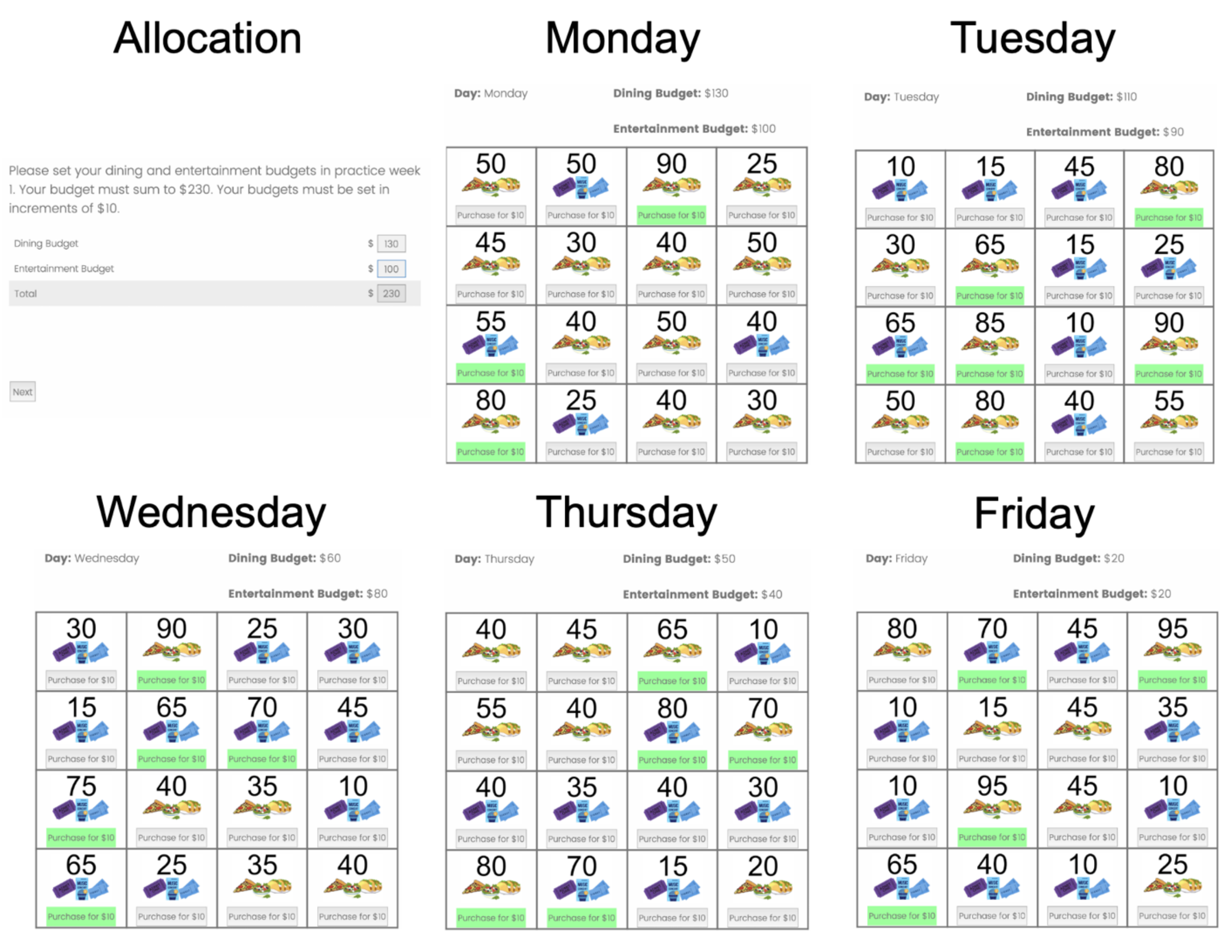
Design Overview. We developed an incentivized, multi-period consumption game in which participants spent money on items that varied in value. All items had the same cost but differed in the points (value) they awarded. Half of the items belonged to the dining category and half belonged to the entertainment category. The goal of the game was to accumulate as many points as possible, which was directly incentivized with a modest additional payout. Therefore, all participants were aligned in their goal to maximize value by purchasing the items with the highest point values, regardless of whether those items belonged to the dining or entertainment category. As our key manipulation of average value, we randomized whether the dining or entertainment category had a higher average value. Additionally, we randomly assigned participants to either play as budgeters (who allocated and then made purchases) or non-budgeters (who never allocated funds).

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

[22] 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. We excluded 35 observations 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.

(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 was re-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 3.6.

FIGURE 3.6: OVERVIEW OF THE CONSUMPTION GAME STAGES (STUDY 3)



Note—A depiction of a simulated week in the consumption game. Each of five practice weeks and five game weeks followed these steps. First, participants in the budgeting condition (as depicted) allocated $230 across a dining budget and an entertainment budget. Next, all participants made purchase decisions on 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.

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 budgeting or non-budgeting condition. Budgeters 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 (see figure 6); 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. Non-budgeters did not allocate weekly funds prior to encountering the five days in which they could spend. Just as it was for budgeters, the only spending constraint was that non-budgeters could not exceed \$230 of weekly spending. Therefore, participants in both conditions faced an identical spending task (spending up to $230 each week) and differed only in that budgeters previously allocated their $230 between the dining and entertainment account, while non-budgeters did not.

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). The threshold of 60 points was deliberate, as this was precisely the value of the marginal purchase for both categories, as detailed below.

Separately manipulating the average value in the high-point region and the low-point region allows us to better analyze where average value plays a role. Recall that we expect consumers to extract average values from the assortment of goods they consider (meaning unconsidered goods should not play a role). In study 2, the design asked participants to identify which vacation activities were considered and which were not. In the present design, we cannot precisely identify the set of considered items from the 80 different options that are presented each week. But prior research suggests high-value options are more likely to be considered and evaluated (Bear et al. 2020; Bettman and Park 1980; Payne 1976). Therefore, by independently manipulating average value among the high-value items and the low-value items, we can assess whether average value plays a role among the set of items most likely to be considered (point values ≥ 60) and the set of items least likely to be considered (point values < 60).

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). Given the expectation that participants would be most likely to consider high-point items (possibly paying little attention to low-point items), we preregistered a clear interest in the effect of the high-point distribution condition. 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 appendix D.

Disentangling Average and Marginal Values. We designed the distributions with one additional key feature in mind. Specifically, we aimed to hold the marginal value constant across the two categories in both conditions, such that participants could do no better by consuming an additional item in either category. We achieved this by ensuring there were always exactly 23 items worth 60 points or more, consisting of exactly 14 dining and 9 entertainment items. 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, corresponding to 14 dining and 9 entertainment items. This optimal split is unaffected by whether dining is manipulated to have a higher average value (relatively more items with point values of 80, 85, 90, or 95) or a lower average value (relatively more items with point values of 60, 65, 70, or 75). Our design also ensures minor deviations from the optimal split (e.g., 16 dining and 7 entertainment; or 12 dining and 11 entertainment) lead to symmetrically-lower payoffs, reducing any incentive to hedge in favor of the category with a higher average value. (The details regarding how we ensure symmetric implications, even as participants slightly deviate from the 14/9 split, are outlined in appendix D).

Holding the marginal value constant by design allows us to isolate the effect of manipulating a category's average value to be high or low. This experimental approach provides a high degree of control, enabling us to cleanly test all hypotheses in an engaging and incentive-compatible paradigm. Specifically, we can ask whether budgeters are sensitive to average value in their allocation decisions (H1) and whether this sensitivity carries through to their subsequent spending (H2). By comparing budgeters' allocations to non-budgeters' spending, we can compare whether allocations are more sensitive to average value than are purchases (H3). And by comparing budgeters' final spending after setting their allocations to non-budgeters' spending, we can observe whether consumers who engage in two-step budgeting are more sensitive to average value than consumers who merely make purchases (H4).

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 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 versus purchasers respond differently to category average values when marginal values are held constant.

Results

Because of the potential for noise and extreme responses, we preregistered to 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 the 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 appendix D).

Dining Share Measure. The fact that study 3 includes two budget categories allows us to consider outcomes in terms of preference for dining or entertainment. For ease of reporting, we discuss our outcomes (e.g., allocation, spending) in terms of the "dining share." Specifically, 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%.[23] For example, $150 to dining and $80 to entertainment equates to a dining share of 65%. For those in the budgeting condition, we can examine the dining share of allocation and the dining share of spending. For those in the non-budgeting condition, we only consider the dining share of spending (because these participants never set budget allocations).

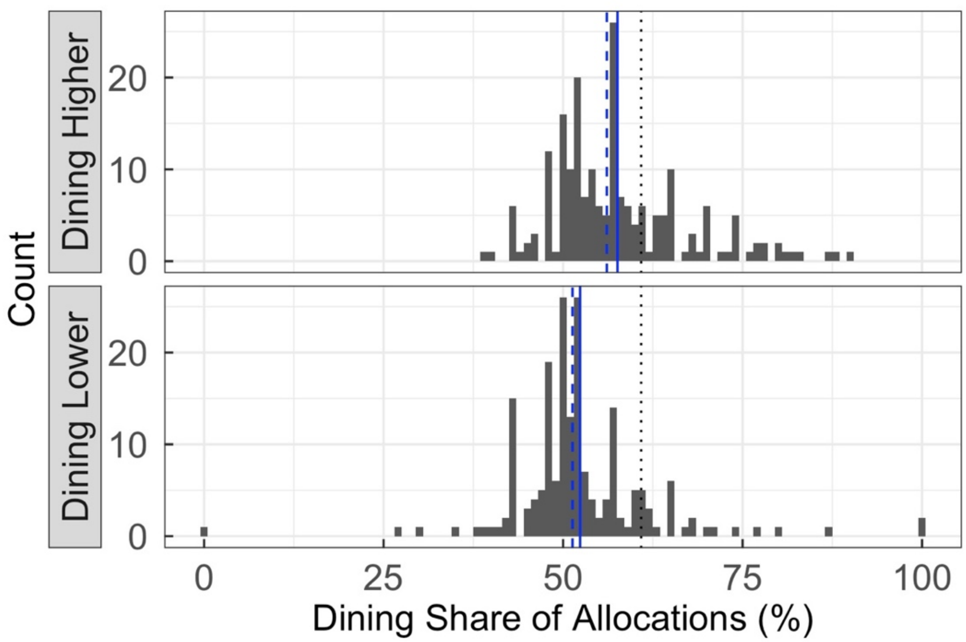
[23] Our preregistration specified (Dining – Entertainment) rather than (Dining / (Dining + Entertainment) x 100%). Because some participants did not exhaust their budget, these two measures are not perfectly deterministic transformations of one another. They are, however, extremely highly correlated (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 for ease of interpretation. Complete preregistered results using the difference measure are presented in appendix D.

Analysis Plan. We preregistered two levels of analysis. First, we consider only participants in the budgeting 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 of allocation in an environment in which marginal value is held constant across categories (H1). Extending this model to the dining share of spending provides a test for H2. Second, we consider all participants and examine the dining share of spending. Here, we are specifically interested in whether budgeters and purchasers differ in their sensitivity to the dining average condition (H3). We regress the dining share of spending on dining average, the low-point dining average, budget versus purchase (+1 = budget, -1 = purchase), and all two- and three-way interactions (H4).

Budget Allocations are Sensitive to Average Value (H1). The first analysis considers only participants in the budgeting 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 dining was manipulated to have a higher average (M = 57.58, SD = 9.47), compared to a lower average (M = 52.37, SD = 9.90; b = 2.61, se = 0.49, $t(390)$ = 5.35, $p < .001$, Cohen's d = 0.54; figure 3.7). This indicates allocations are sensitive to the average value of budget categories, even when the marginal value of each category is held constant (H1).

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 assigned to see lower average values in this region (*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). Similar to the robustness check in study 2, we construct an overall measure of average value from condition assignment in both the high- and low-point regions[24] and observe that being assigned to a higher overall average value increases budget allocations (*t*(392) = 4.96, *p* < .001; appendix D).

FIGURE 3.7: BUDGETS BY DINING AVERAGE CONDITION (STUDY 3)



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).

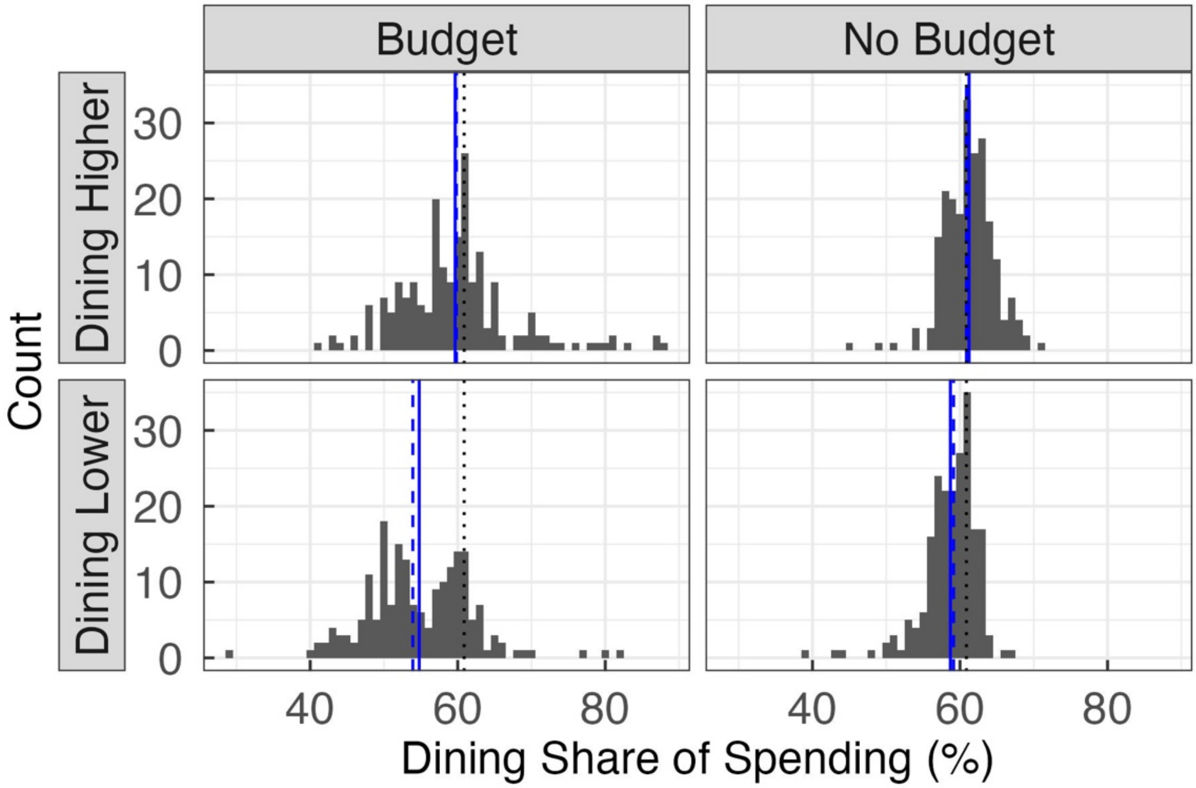
[24] We can approximate an overall average by collapsing the 2 (dining average: high, low) x 2 (low-point dining average: high, low) conditions into 3 conditions: both high (coded as +1), mixed (coded as 0), and both low (coded as -1).

Budgeters Spend More in Categories with a Higher Average Value (H2). We can test whether budgeters spend more in high-average categories by regressing the dining share of spending on the dining average (as well as the low-point region dining average, and their interaction). Even though participants could deviate from their allocations (so long as they adhered to the total constraint of $230 each week), we observe budgets are sticky (see supporting analyses in appendix D). As a result, budgeters in the higher dining average condition spent more on dining (M = 59.65, SD = 7.87) than those in the lower dining average condition (M = 54.78, SD = 7.13; $t(390)$ = 6.49, $p < .001$, Cohen's d = 0.65).

Budget Allocations are More Sensitive to Average than are Purchases by Non-Budgeters (H3). This study permits a non-preregistered test of H3. Specifically, we can compare the dining share of allocation 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 budgeters' allocations (M_{higher} = 57.78, SD_{higher} = 9.47 vs. M_{lower} = 52.37, SD_{higher} = 9.90) than non-budgeters' spending (M_{higher} = 61.23, SD_{higher} = 3.40 vs. M_{lower} = 58.70, SD_{higher} = 3.61; $t(813)$ = 2.56, p = .011, Cohen's f = 0.09). The main effect of budgeting may represent naïve diversification for budgeting (Benartzi and Thaler 2001; Bardolet, Fox, and Lovallo 2011), pushing allocations toward an even split.

Budgeters' Spending is More Sensitive to Average Value than Purchasers' Spending (H4). Recall, our preregistered plan was to regress dining spending on the dining average condition, the low-point dining average condition, budget condition, and all two- and three-way interactions. The goal of this analysis was to test whether the effect of dining average had a different effect on spending for budgeters versus 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 H4, which predicts spending will be more sensitive to average value for budgeters than non-budgeters. Specifically, spending was more sensitive to the dining average among participants who previously set budgets (M_{higher} = 59.65, SD_{higher} = 7.87 vs. M_{lower} = 54.78, SD_{lower} = 7.13) compared to those who never set budgets (M_{higher} = 61.23, SD_{higher} = 3.40 vs. M_{lower} = 58.70, SD_{lower} = 3.61). Given differences in variance across conditions, we repeated this analysis with robust standard errors. No substantive nor statistical conclusions changed. See figure 3.8. Full results are reported in appendix table 3.D4.

FIGURE 3.8: CONDITIONAL SPENDING DISTRIBUTIONS (STUDY 3)



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).

Discussion

Study 3 permits tests of all four hypotheses within a controlled and incentivized experimental paradigm, in which the marginal value, given value-maximizing decisions, is held constant by design. Consistent with H1, we find budgeters allocate more funds to categories with higher average values. Consistent with H2, budgeters spend more on categories with higher average values (because prior budget allocations guide subsequent spending, even though these allocations are non-binding). As predicted by H3, the allocations of budgeters are more sensitive to average value than the spending of non-budgeters. As a result, and consistent with H4, budgeters' spending is more sensitive to average value than non-budgeters'. Taken together, this experiment suggests consumers with identical preferences are differentially sensitive to category average values depending on whether or not they budget, thus leading to differences in spending.

Beyond the ability to test all four hypotheses within a single setting, we extend our prior findings to a budgeting context in which consumers allocate funds between multiple distinct accounts. Whereas prior studies examined allocation into a single budget category (relative to an outside option), the present study extends our findings to additional budgeting contexts. As a concluding note, the game design is quite distinct from the scenarios of other studies (including the supplementary study). This experiment provided an engaging and incentivized repeated decision task in which there was ample opportunity to learn with well-defined item values. Consistent findings across such varied designs lends support to the generalizability of the findings across contexts.

GENERAL DISCUSSION

Consumers' budget allocations matter because they affect spending. The current research explores how consumers set budgets and identifies average value as an important driver of budget allocations. Uncovering the role of average value in budget allocations introduces a number of possibilities that would not occur if budgeters followed normative allocation principles (i.e., setting budgets based on equating value at the margin). For example, we suggest allocations will be sensitive to factors that change the perception of average value, even when marginal value is unaffected (e.g., distributions that pull the average up or down, as in study 3).

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 (versus purchasing without prior allocation) changes the composition of spending, even when it does not change the amount of spending.

Future Directions

Measuring Central Tendency. In study 1, consumers reported their own assessment of average value; in study 2, we analyzed the mean of rated values; in study 3, we manipulated a distribution of values. Across these studies, the underlying construct of "average" does not distinguish among multiple forms 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 multiple possible measures of average.

Recent work suggests there may be meaningful distinctions between such metrics in certain contexts (Howard et al. 2022). We suspect such distinctions are unlikely to qualitatively impact our findings. First, our core interest is not in distinguishing between related measures of central tendency (Howard and Shiri 2022), but rather in examining 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 study 3. Across studies, we demonstrate the robustness of average value (as a proxy for central tendency) by measuring it in different ways. The same key pattern of results exists across these 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.

What Gets Considered When Evaluating Budget Categories. A second 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 both studies 2 and 3, participants are sensitive to the overall average value, though we separately consider average values in more localized regions (e.g., above or below the margin). Yet, in study 3, we find participants are more sensitive to average value for high-point items than low-point items, whereas in study 2 we found greater sensitivity to the average value of low-value considered items. This raises the possibility that consumers put differential weight on values across the distribution depending on context (Bear et al. 2020) or edit out non-considered options from consideration (Kahneman and Tversky 1979). This latter explanation may provide a way to reconcile the study 2 and study 3 results, as study 3 participants with ample learning opportunity may have grown accustomed to disregarding the low-value, infrequently purchased items. In study 2, on the contrary, even the low-value items were still identified by participants as part of the consideration set. This potential role for editing or differential attention may be especially important when consumers have budget categories spanning a wide array of possible values (as in study 3). Valuable alternatives which are highly accessible may encourage consumers to allocate more money to a budget, even if the value of that alternative 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 these are sometimes in conflict with one another (e.g., short-run value, like taste, versus 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. In the supplementary experiment detailed in the Appendix, we find allocating (versus purchasing) induces greater relative focus on category-level evaluations than on item-level evaluations. However, there are likely additional predictors of evaluation mode. 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.

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 to thinking on the margin. 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 study 3, in which we held constant the set of items that would earn the most points and the largest real bonus payment. Allocating (and downstream spending) in line with a category's average value dragged some participants away from this value-maximizing bundle. If 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 rather than the marginal purchases.

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 study 3. As budgeting tools in the fintech space like EveryDollar, YNAB (You Need A Budget), Rocket Money, and Simplifi 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 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 highest-valued 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. 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.

Chapter 3: Appendix

for

Average Value Affects Consumer Budgets

APPENDIX SUPPLEMENTAL STUDY: AMAZON

Do allocation decisions engage a different mode of evaluation than purchase decisions?

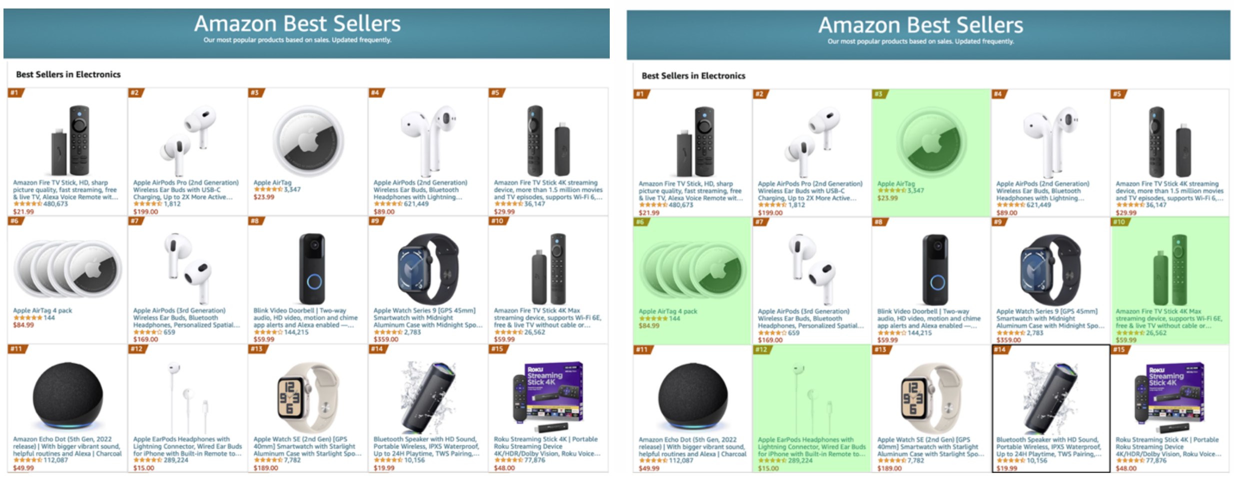
Method

We recruited 200 participants from AMT in the last week of July (2024) to take part in a brief experiment (M_{age} = 43; 41% female). All participants were asked to select the Amazon.com department they were most likely to shop from the list including: "Grocery & Gourmet Food," "Tools & Home Improvement," "Electronics," "Beauty & Personal Care," and "Kitchen & Dining." Subsequently, participants viewed an image of the "Amazon Best Sellers" for their selected department. These images were captured directly from Amazon.com and presented the top 30 best-selling items in late July 2024 (figure 3.S1).

Participants were randomly assigned to either a budget or purchase condition. In the budget condition, participants were asked to indicate "what budget/s would you set for these types of products in the near future?" and responded using an open-ended text box. In the purchase condition, participants were instructed to "think about the kinds of products you would purchase in this department if you had a $50 Amazon gift card." We asked participants to "click on the types of products you would buy" and reinforced "you do not need to use the entire gift card; and you can also spend additional money by going into your own pocket." In the purchase condition, each item from the Amazon Best Seller list was clickable using the Qualtrics heat map feature (figure 3.S1).

After the task of thinking through budget allocation or purchase decisions, participants provided a self-report measure of their evaluation mode. Specifically, respondents indicated whether they had been focusing more on individual items or more on the collection of items as a whole. This was measured on a 1-7 scale anchored on 1 = "Entirely focused on individual items" and 7 = "Entirely focused on the collection as a whole."

FIGURE 3.S1: EXAMPLE STIMULI FROM SUPPLEMENTAL STUDY



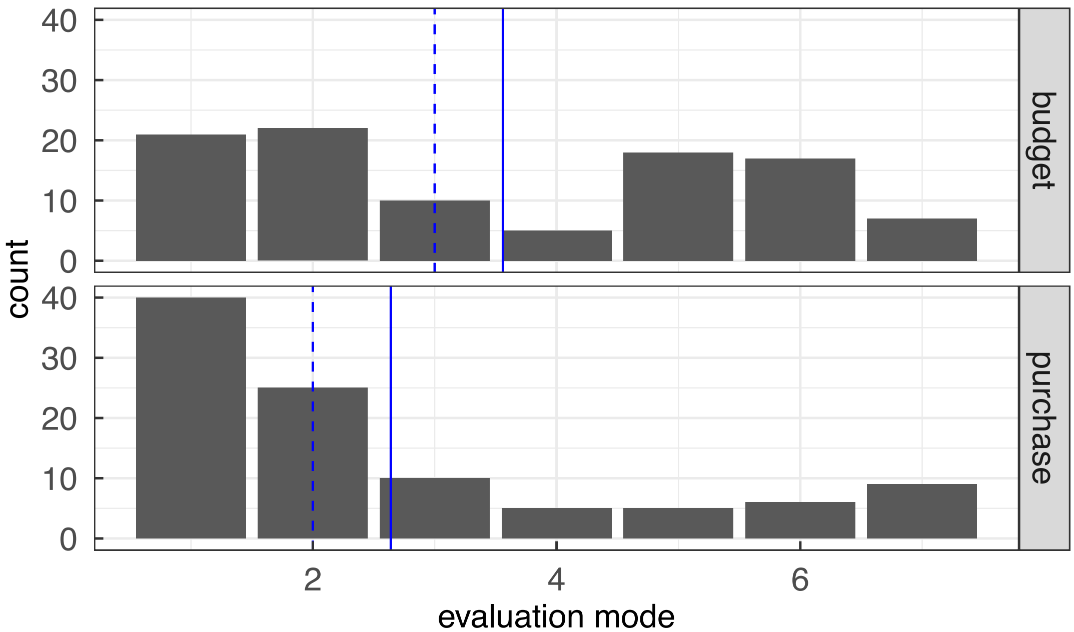
Note—(*Left*) Example of Amazon Best Sellers in the electronics department. Truncated screenshot shows 15/30 products presented to participants. (*Right*) In the purchase condition, participants clicked on the types of products they would buy within the Amazon department they previously selected.

Results

All participants were included for analysis. The preregistered t-test indicates budgeters' self-reported evaluation mode was more consistent with category focus (M = 3.56, SD = 2.05) than purchasers' was (M = 2.64, SD = 2.00; $t(198) = 3.21$, p = .002, Cohen's d = 0.45; figure

3.S2). 42% of budgeters—but only 20% of purchasers—indicated a greater relative focus on categories than items (5, 6, or 7).

FIGURE 3.S2: DISTRIBUTIONS OF SELF-REPORTED EVALUATION MODE



Note—Distributions of self-reported evaluation mode. Lower scores indicate greater relative focus on items, and higher scores indicate greater relative focus on categories. Solid lines are marginal means, and dashed lines are marginal medians.

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; as measured in the manuscript study 1), budgets for discretionary purchases are more likely to be set based on consumer preferences, in which case the value ofbudget categories may play an important role.

Method

We surveyed 200 participants from a gender-balanced sample on Amazon Mechanical Turk (AMT). 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%) usesome 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 3.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 3.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. Participants could select multiple responses.

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 3.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 studies in the main text, we consider "Dining" (a label encompassing discretionary elements of "food" and "groceries") and "Entertainment."

TABLE 3.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.

APPENDIX B: ADDITIONAL MATERIALS FOR STUDY 1

As this was a descriptive survey of consumers' own behaviors and experiences (and not an experiment with clear *a priori* statistical inferences), we did not preregister this study.

Additional measures: item focus and category focus. The manuscript details the key findings for average value and marginal value and indicates that the results for item focus and category focus are reported here. We decided to concentrate our manuscript discussion on average value and marginal value to enhance clarity to the reader, as we discuss (and test) the role of item-level and category-level evaluations later in the main text. The findings here are compatible with—and certainty not contradictory towards—the preregistered, experimental findings of the supplementary Amazon study (manuscript Appendix).

Design

Following the budgeting exercise using the pie chart, participants self-reported four measures pertaining to their focus on (i) average value, (ii) marginal value, (iii) categories, and (iv) items. For all four measures (average value, marginal value, category focus, item focus), we sampled from four different question variants. As discussed in the manuscript, we took this approach to reduce the likelihood that any conclusions are tied to a unique question wording. The full set of variant wordings is provided in table 3.B1.

TABLE 3.B1: VARIANT WORDINGS USED IN STUDY 1

When thinking about setting your budget, to what extent did you find yourself…

Measure	Variant wording	Variant id
Average	…thinking about your overall impression of how much you like each category?	1
	…remembering your general liking of each category?	2
	…comparing your overall enthusiasm for each category?	3
	…relying on your general evaluation of each category?	4
Margin	…thinking about what you could buy with just a little more money (or conversely, what you would lose if you spent a little less)?	1
	…imagining how a small adjustment to one of your budgets could change what you buy?	2
	…weighing the trade-offs between having enough money to buy one thing or the other, but not both?	3
	…considering how giving up one thing might allow you to buy something else?	4
Category	…thinking about categories of expenditures?	1
	…considering trade-offs between categories of purchases?	2
	…imagining a collection of purchases with a shared meaning?	3
	…focusing on the big-picture sense of what each category represents?	4
Item	…thinking about and imagining specific, individual expenses?	1
	…considering particular ways to spend money within a given category?	2
	…paying attention to specific things you could purchase?	3
	…visualizing a concrete thing that you will spend your money on?	4

Question ordering was counterbalanced across two factors. The first factor was whether the measure corresponded to value (average, margin) vs. evaluation mode (category, item). We counterbalanced which set of questions (value vs. evaluation mode) appeared 1st and 2nd, as opposed to 3rd and 4th. The second factor was whether the measure aligned more closely with ensemble perception (average value, category evaluations) or normative principles of decision making (marginal value, item evaluations). Thus, this second factor determined whether the ensemble vs. normative questions were presented first or second, conditional upon the first factor. These two counterbalancing factors were not significant predictors of measure responses, which we assess by regressing each of the 16 measure variants on the contrast-coded counterbalancing variables and their interaction. There are no significant main effects of either counterbalancing factor or the interaction after adjusting for multiple comparison testing (to account for the 48 comparisons: 16 main effects from the first counterbalanced factor, 16 main effects from the second counterbalanced factor, and 16 interactions).

Results

In the manuscript, we report our findings using the complete dataset of 100 participants. As footnoted, there were originally 101 complete observations; however, two observations were tied to the same participant identifier. To preserve independent and naïve responses, we eliminated the second observation linked to this participant.

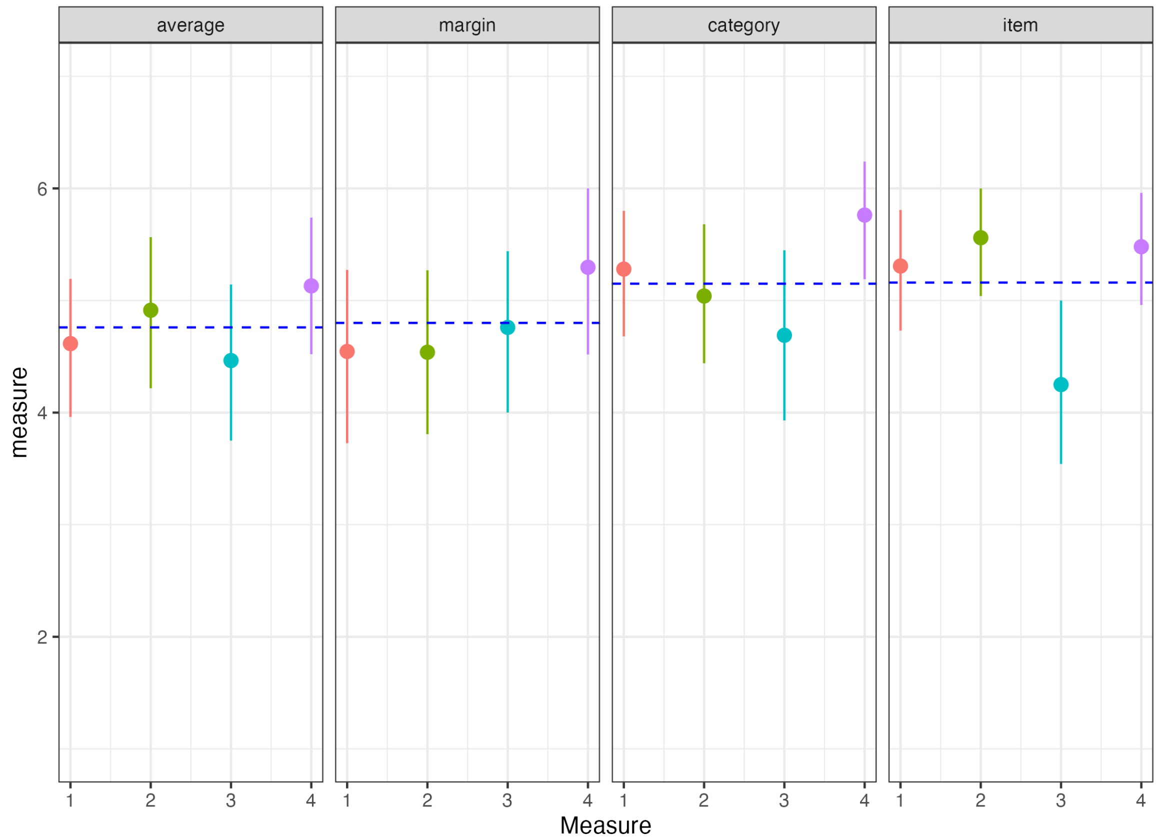
Monthly Finances. Participants reported a median monthly take-home pay of \$2500 (*M* = \$4588, *SD* = \$9225). Of this, the median amount dedicated to essential expenses was \$1550 (*M* = \$6120, *SD* = \$35559). Both distributions were quite skewed, as suggested by the comparison of median and mean, alongside the reported standard deviations. For checks of robustness, we identify any observations farther than two standard deviations from the mean (+/- 2 SD). We subsequently reconsider our analyses with these observations removed.

Recall, participants allocated discretionary funds using the pie chart budgeting tool. To calculate discretionary funds, we subtracted expenses from take-home pay. At the individual level, the median discretionary amount was $650. A total of 15/100 participants had discretionary levels below $100. We asked participants with sub-$100 discretionary funds to continue the exercise as though they had $100 available. Therefore, the mean amount of discretionary funds ($1566; *SD* = $2850) is dependent on this specific approach. A total of 3 participants had actual discretionary funds equal to $100, yielding the reported 18/100 participants who allocated $100 of funds, as reported in the manuscript.

Allocations. Though not central to the question of how budgeters perceive value (in terms of either average value or marginal value), we observe the self-reported discretionary budgets. On average, participants allocated 39% of their discretionary funds to groceries, 14% to dining out, 21% to entertainment, and 16% to clothing. The remaining 10% was allocated across self-generated categories (used by 23/100 participants). Additional descriptions and analyses of budgeters' allocations are provided in our ResearchBox materials.

Self-Reported Focus Measures. The complete descriptive results pertaining to the self-reported measures for average value, marginal value, category focus, and item focus are presented in figure 3.B1. As previously discussed, there was no effect of the question ordering (the two counterbalanced factors) on these measures. Furthermore, these measures (as well as the main reported findings in the manuscript) are robust to excluding participants with extreme incomes or expenses (+/- 2 SD) and the participants with imputed $100 discretionary spending. The associated plots and tests are available in our online materials.

FIGURE 3.B1: SELF-REPORTED FOCUS MEASURES ACROSS ALL QUESTION VARIANTS



Note—Self-reported focus on the dimensions of average value , marginal value, category focus, and item focus across all four question variants. Higher scores indicate an increased focus on the specific dimension of value. Error bars are 95% confidence intervals. The solid blue lines independently depict the mean average value and marginal value. The question variant refers to the “variant id” in table 3.B1.

APPENDIX C: ADDITIONAL MATERIALS FOR STUDY 2

We preregistered a set of independent regression models with the intent of presenting our findings graphically.

$$ALLOCATE_{k,i} = b_{0_k} + b_{1_k}ABOVE_{k,i} + b_{2_k}MARGIN_{k,i} + b_{3_k}BELOW_{k,i} \qquad \text{(EQ 1)}$$

For the range of data (k = [3, 13]), this amounts to the following independent regression equations, which are estimated in table C.1B and plotted in figure C.1.

$$ALLOCATE_{3,i} = b_{0_3} + b_{1_3}ABOVE_{3,i} + b_{2_3}MARGIN_{3,i} + b_{3_3}BELOW_{3,i}$$

$$ALLOCATE_{4,i} = b_{0_4} + b_{1_4}ABOVE_{4,i} + b_{2_4}MARGIN_{4,i} + b_{3_4}BELOW_{4,i}$$

$$\ldots$$

$$ALLOCATE_{13,i} = b_{0_{13}} + b_{1_{13}}ABOVE_{13,i} + b_{2_{13}}MARGIN_{13,i} + b_{3_{13}}BELOW_{13,i}$$

We note that the preregistered approach (EQ 1) can be equivalently presented as a set of separate equations or a single a nested model, in which there are no main effects, but rather a set of simple effects (for each level of k). The coefficient estimates are equivalent, though pooling the data and accounting for non-independence of observations results in similar but not identical standard errors.

For ease of explication, we deviate from our preregistered plan by unnesting the estimates of ABOVE, MARGIN, and BELOW in order to produce estimated main effects for each measure of value, where FE_k represents fixed effects for each rank. Specifically, we consider the model presented as EQ 2 (with cluster-robust standard errors).

$$ALLOCATE_{k,i} = b_0 + b_1 ABOVE_{k,i} + b_2 MARGIN_{k,i} + b_3 BELOW_{k,i} + FE_k \qquad \text{(EQ 2)}$$

These results are presented in table 3.C0 and in the main text. As a technical note, the main effects in the alternate model (EQ 2) may be roughly considered as the weighted average of the simple slopes across nested levels in the original model (EQ 1). Therefore, we present the alternate model (EQ 2) in the manuscript for the simplicity of reporting a single set of main effects (ABOVE, MARGIN, BELOW).

TABLE 3.C0: ESTIMATES FROM UNNESTED MODEL IN MANUSCRIPT

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df
a	0.0008	0.0087	0.0946	0.9248	-0.0165	0.0181	147.8891
m	0.0197	0.0033	5.9286	0.0000	0.0132	0.0263	280.2366
b	0.0351	0.0063	5.5580	0.0000	0.0227	0.0476	201.5317

Note—Main effects from a single unnested model for ABOVE (“a”), MARGIN (“m”), and BELOW (“b”), estimated using *lm_robust()* with cluster-robust standard errors (at the participant level) and rank fixed effects (estimatr package in R).

C.1: Preregistered Analysis

Nested Model

$$ALLOCATE_{k,i} = b_{0_k} + b_{1_k}ABOVE_{k,i} + b_{2_k}MARGIN_{k,i} + b_{3_k}BELOW_{k,i}$$

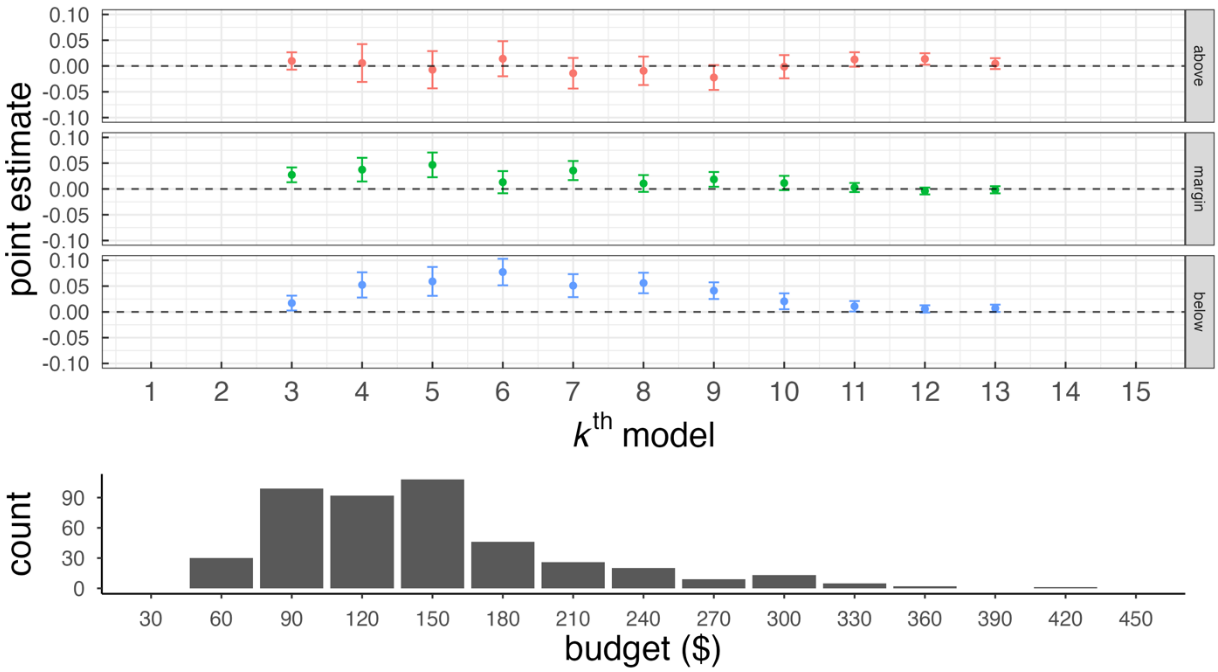
Additional details on the coding and construction of these variables are provided in table 3.C1.A.

TABLE 3.C1.A: VARIABLE NAME AND CONSTRUCTION

Variable Name	Variable Construction
ALLOCATE	Indicator that participant *i* allocates enough funds for at least *k* activities [0/1]
ABOVE	Mean rated value of all considered options ranked better than *k* for participant *i*
MARGIN	Rated value of the k^{th} option for participant *i*
BELOW	Mean rated value of all considered options ranked worse than *k* for participant *i*

The preregistered plan was to have ABOVE, MARGIN, and BELOW are nested under each level of *k*, in which the coefficients of interest can be presented as either (i) the set of simple slopes (in a single nested model with ABOVE, MARGIN, and BELOW interacting with each level of *k*) or (ii) independent *k*-level regressions. For clarity, we present the results following this latter approach. The point estimates for ABOVE, MARGIN, and BELOW are plotted in figure 3.C1 and the corresponding regression results are in table 3.C1.B.

FIGURE 3.C1: PRERESGISTERED RESULTS AND BUDGET SIZE



Note—The top panel plots the point estimates for ABOVE, MARGIN, and BELOW across independent regressions (at various levels of *k*). Error bars are 95% confidence intervals. The bottom panel plots the frequencies of participants budgeting a given amount, in dollars.

TABLE 3.C1.B: PREREGISTERED RESULTS

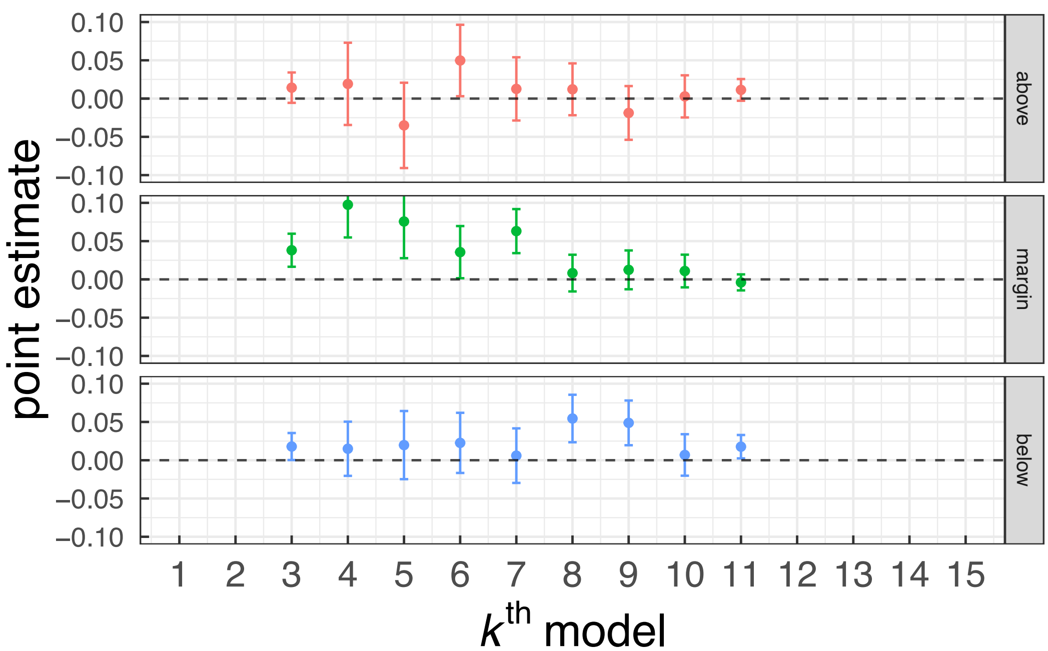
	Dependent variable:										
	dv3	dv4	dv5	dv6	dv7	dv8	dv9	dv10	dv11	dv12	dv13
ABOVE	0.01	0.01	−0.01	0.01	−0.01	−0.01	-0.02^{+}	−0.001	0.01^{+}	0.01^{*}	0.005
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
MARGIN	0.03^{***}	0.04^{**}	0.05^{***}	0.01	0.04^{***}	0.01	0.02^{*}	0.01	0.003	−0.004	−0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.004)	(0.003)	(0.004)
BELOW	0.02^{*}	0.05^{***}	0.06^{***}	0.08^{***}	0.05^{***}	0.06^{***}	0.04^{***}	0.02^{**}	0.01^{*}	0.01	0.01^{+}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.004)	(0.004)
Constant	0.54^{***}	0.14	−0.01	-0.25^{+}	−0.13	−0.08	−0.002	−0.06	-0.12^{*}	-0.08^{*}	−0.04
	(0.08)	(0.14)	(0.15)	(0.13)	(0.11)	(0.10)	(0.08)	(0.07)	(0.05)	(0.03)	(0.03)
Observations	446	438	434	424	409	385	349	305	272	227	175
Note:											+p<0.10;*p<0.05;**p<0.01;***p<0.001

C.2: Robustness: Removing monotonicity violations

Nested Model

The model in analysis 3.C2 is identical to the primary model, except we exclude any observations in which the marginal value (the k^{th} ranked option) is rated lower than any worse-ranked alternative. We provide both the nested (figure 3.C2, table 3.C2A) and unnested results (table 3.C2.B). As a technical note, as a result of excluding observations that are not at least weakly monotonically decreasing (when considering the rated value-to-rank relationship), there is no remaining variation in the dependent variable (allocating sufficient funds for a given level of k) when $k = 12$, 13. For this reason, we cannot create estimates for ABOVE, MARGIN, and BELOW when $k > 11$.

FIGURE 3.C2: ESTIMATES FROM NESTED MODELS, REMOVING MONOTONICITY VIOLATIONS



Note—This plot portrays point estimates of the independent regressions in a manner identical to figure C.2, except: (1) regressions are estimated after removing any monotonicity violations and (2) models for k = 12, 13 are no longer estimable due to data loss.

TABLE 3.C2.A: ESTIMATES FROM NESTED MODELS AFTER REMOVING MONOTONICITY VIOLATIONS

	Dependent variable:								
	dv3	dv4	dv5	dv6	dv7	dv8	dv9	dv10	dv11
ABOVE	0.01	0.02	−0.04	0.05**	0.01	0.01	−0.02	0.003	0.01
	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
MARGIN	0.04***	0.10***	0.08***	0.04**	0.06***	0.01	0.01	0.01	−0.004
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
BELOW	0.02**	0.01	0.02	0.02	0.01	0.05***	0.05***	0.01	0.02**
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Constant	0.39***	−0.34*	0.15	−0.52***	−0.37**	−0.21*	−0.002	−0.05	−0.09**
	(0.09)	(0.21)	(0.21)	(0.19)	(0.15)	(0.11)	(0.11)	(0.09)	(0.05)
Observations	300	250	225	237	229	225	198	166	163

Note: +p<0.10;*p<0.05;**p<0.01;***p<0.001

Unnested Model

Unnesting the model to consider main effects:

TABLE 3.C2.B: ESTIMATES FROM UNNESTED MODEL AFTER REMOVING MONOTONICITY VIOLATIONS

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df
a	0.0117	0.0092	1.2746	0.2048	-0.0065	0.0299	128.9565
m	0.0313	0.0061	5.1270	0.0000	0.0193	0.0434	176.8822
b	0.0244	0.0085	2.8871	0.0043	0.0077	0.0411	193.0599

Note—Main effects from a single unnested model for ABOVE ("a"), MARGIN ("m"), and BELOW ("b"), estimated using *lm_robust()* with cluster-robust standard errors (at the participant level) and rank fixed effects (estimatr package in R).

C.3: Including non-considered activities (with zero values)

Nested Model

$$ALLOCATE_{k,i} = b_{0_k} + b_{1_k}ABOVE_{k,i} + b_{2_k}MARGIN_{k,i} + b_{3_k}BELOW_{k,i}$$

A key feature of the design of study 2 was to allow participants to identify the considered set (vs. non-considered set). Our preregistered approach, manuscript approach, and the previously discussed robustness checks are all constrained to the values of considered activities. As an additional check, we can also incorporate the non-considered activities into the model with an imputed value of 0.[25] As can be seen in table 3.C3.A, this imputation affects only the construction of BELOW. Specifically, the average value of activities ranked worse than *k* (what is captured by BELOW) will decrease for participants who considered < 15 activities, as this measure will be dragged down by the inclusion of 0 values. We present results using the nested model (figure 3.C3, table 3.C3.B) and unnested model (table 3.C3.C).

TABLE 3.C3.A: VARIABLE NAME AND CONSTRUCTION

Variable Name	Variable Construction
ALLOCATE	Indicator that participant *i* allocates enough funds for at least *k* activities [0/1]
ABOVE	Mean rated value of all considered options ranked better than *k* for participant *i*
MARGIN	Rated value of the k^{th} option for participant *i*
BELOW	Mean rated value of all options ranked worse than *k* for participant *i*, including imputed 0 values for non-considered options

[25] Of course, there may be different, non-linear functional forms to assign imputed values to non-considered items; however, this is beyond the scope of our analysis.

FIGURE 3.C3: ESTIMATES FROM NESTED MODELS AFTER IMPUTING 0 VALUES FOR NON-CONSIDERED ACTIVITIES

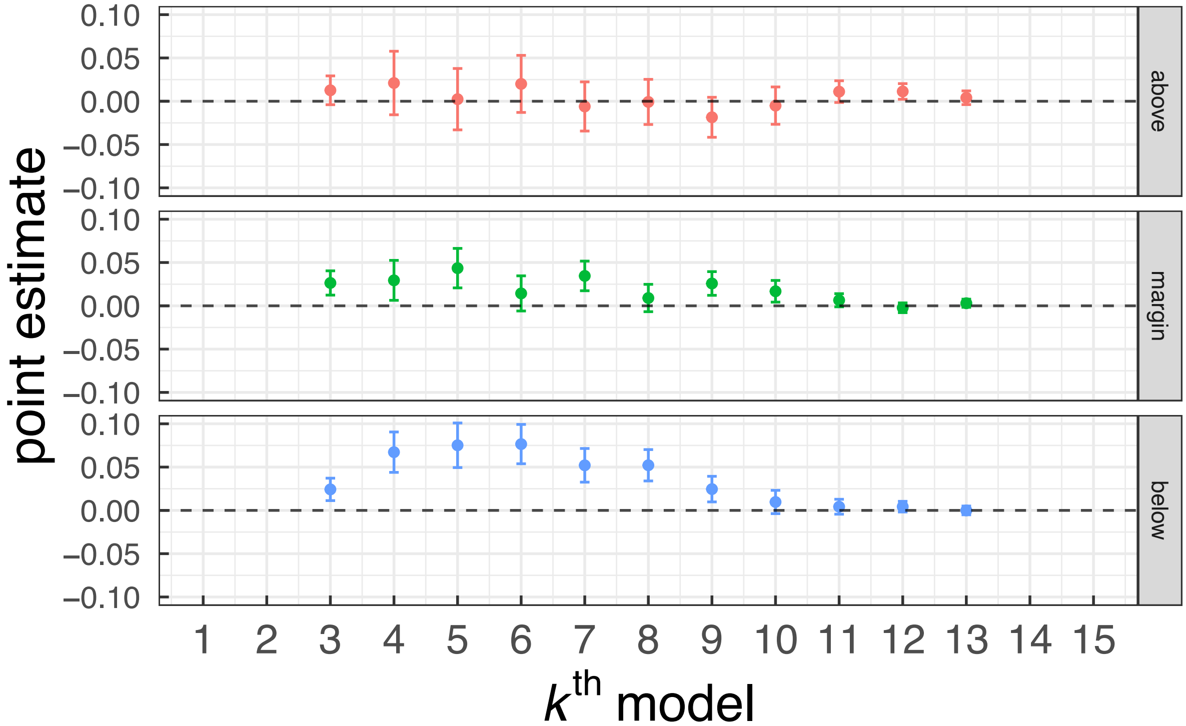


TABLE 3.C3.B: ESTIMATES FROM NESTED MODELS AFTER IMPUTING 0 VALUES FOR NON-CONSIDERED ACTIVITIES

	Dependent variable:										
	dv3	dv4	dv5	dv6	dv7	dv8	dv9	dv10	dv11	dv12	dv13
ABOVE	0.01	0.02	0.002	0.02	−0.01	−0.001	−0.02	−0.01	0.01^{+}	0.01^{*}	0.004
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.005)	(0.004)
MARGIN	0.03^{***}	0.03^{*}	0.04^{***}	0.01	0.03^{***}	0.01	0.03^{***}	0.02^{**}	0.01^{+}	−0.002	0.003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.004)	(0.003)	(0.002)
BELOW	0.02^{***}	0.07^{***}	0.08^{***}	0.08^{***}	0.05^{***}	0.05^{***}	0.02^{**}	0.01	0.004	0.004	−0.0000
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.004)	(0.003)	(0.003)
Constant	0.52^{***}	0.08	−0.04	−0.20	−0.12	−0.05	0.03	−0.004	-0.09^{*}	-0.07^{*}	−0.03
	(0.08)	(0.14)	(0.14)	(0.13)	(0.11)	(0.09)	(0.08)	(0.07)	(0.04)	(0.03)	(0.02)
Observations	449	446	438	434	424	409	385	349	305	272	227

Note: +p<0.10;*p<0.05;**p<0.01;***p<0.001

Unnested Model

Unnesting the model to consider main effects:

TABLE 3.C3.C: ESTIMATES FROM UNNESTED MODEL AFTER IMPUTING 0 VALUES FOR NON-CONSIDERED ACTIVITIES

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df
a	0.0043	0.0082	0.5222	0.6023	-0.0119	0.0205	158.7981
m	0.0188	0.0032	5.7725	0.0000	0.0124	0.0251	290.9755
b	0.0364	0.0056	6.4431	0.0000	0.0252	0.0475	160.7004

Note—Main effects from a single unnested model for ABOVE (“a”), MARGIN (“m”), and BELOW (“b”), estimated using *lm_robust()* with cluster-robust standard errors (at the participant level) and rank fixed effects (estimatr package in R).

C.4: Average of all considered options

Nested Model

$$ALLOCATE_{k,i} = b_{0_k} + b_{1_k}AVERAGE_{k,i} + b_{2_k}MARGIN_{k,i}$$

The primary hypothesis we test is whether budget allocations are sensitive to a category's average value, above and beyond marginal value (H1). Study 2 benefits from our ability to separate the average value of activities ranked better and worse than the marginal good (corresponding to ABOVE and BELOW) to understand what region of the value distribution may be driving the sensitivity to average value. However, it is still important to understand whether participants are sensitive to the simple average value, calculated from all considered options, aside from the marginal activity. Results from the nested model are presented in figure 3.C4 and table 3.C4.B, and the main effects from the unnested model are given by table 3.C4.C.

TABLE 3.C4.A: VARIABLE NAME AND CONSTRUCTION

Variable Name	Variable Construction
ALLOCATE	Indicator that participant *i* allocates enough funds for at least *k* activities [0/1]
AVERAGE	Mean rated value of all considered options except the k^{th} option for participant *i*
MARGIN	Rated value of the k^{th} option for participant *i*

FIGURE 3.C4: ESTIMATES FROM NESTED MODELS USING A SINGLE MEASURE OF AVERAGE

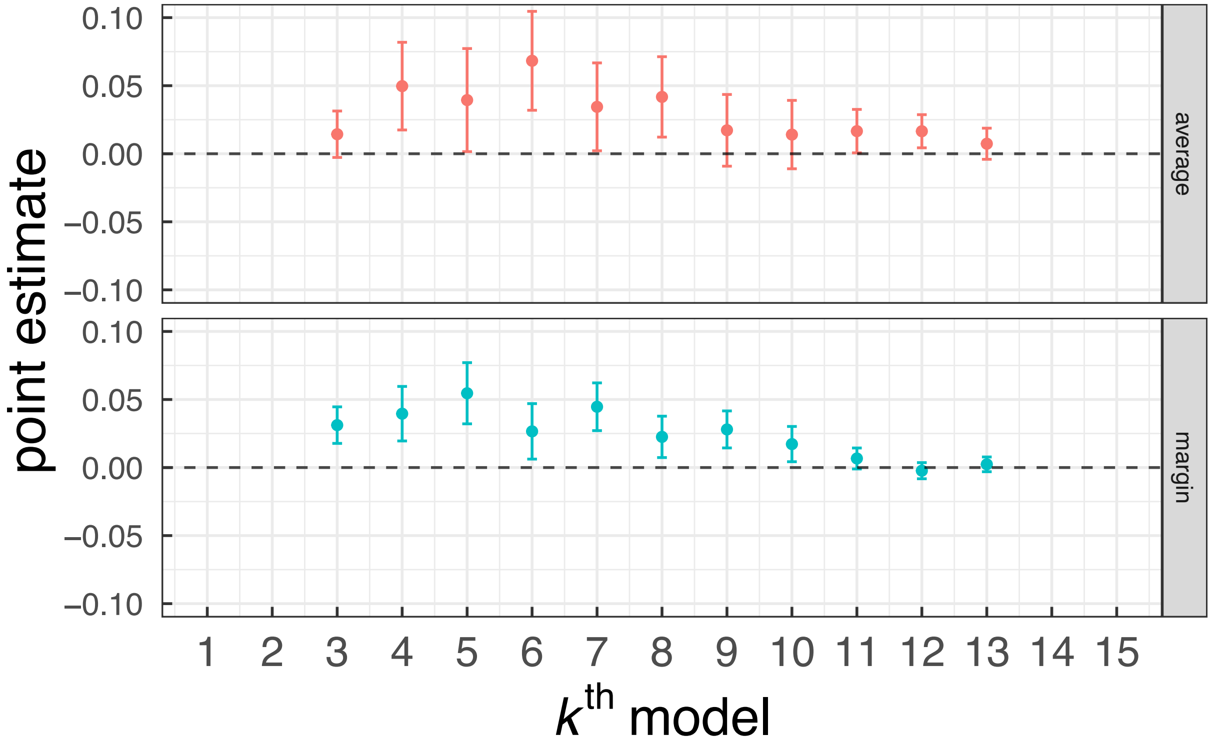


TABLE 3.C4.B: ESTIMATES FROM NESTED MODELS USING A SINGLE MEASURE OF AVERAGE

	Dependent variable:										
	dv3	dv4	dv5	dv6	dv7	dv8	dv9	dv10	dv11	dv12	dv13
AVERAGE	0.01	0.05***	0.04**	0.07***	0.03**	0.04***	0.02	0.01	0.02**	0.02***	0.01
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
MARGIN	0.03***	0.04***	0.05***	0.03**	0.04***	0.02***	0.03***	0.02***	0.01*	−0.002	0.002
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.004)	(0.003)	(0.003)
Constant	0.60***	0.13	−0.08	−0.30***	−0.28***	−0.25***	−0.16**	−0.11	−0.11***	−0.08**	−0.05
	(0.07)	(0.10)	(0.11)	(0.10)	(0.09)	(0.08)	(0.07)	(0.07)	(0.04)	(0.03)	(0.03)
Observations	446	438	434	424	409	385	349	305	272	227	175

Note: +p<0.10;*p<0.05;**p<0.01;***p<0.001

Unnested Model

Unnesting the model to consider main effects:

TABLE 3.C4.C: ESTIMATES FROM UNNESTED MODEL USING A SINGLE MEASURE OF AVERAGE

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df
m	0.0264	0.0033	7.8954	0.0000	0.0199	0.0330	283.6958
avg	0.0319	0.0082	3.9059	0.0001	0.0158	0.0480	236.3242

Note—Main effects from a single unnested model for MARGIN ("m") and the overall average of considered activities ("avg"), estimated using *lm_robust()* with cluster-robust standard errors (at the participant level) and rank fixed effects (estimatr package in R).

C.5: Substituting Adjacent Values for Averages

Nested Model

$$ALLOCATE_{k,i} = b_{0_k} + b_{1_k}ABOVE_{k,i} + b_{2_k}MARGIN_{k,i} + b_{3_k}BELOW_{k,i}$$

While our discussion of the literature of ensemble perception suggests people extract average representations from the considered set, we additionally test whether this sensitivity is captured by the valuations of activities with the most similar values to the marginal item. Specifically, we construct ABOVE as the single value of the activity ranked directly better and BELOW as the single value of the activity ranked directly worse. Results for the nested model are presented in figure 3.C5 and table 3.C5.B, and the main effects from the unnested model are given in table 3.C5.C.

TABLE 3.C5.A: VARIABLE NAME AND CONSTRUCTION

Variable Name	Variable Construction
ALLOCATE	Indicator that participant *i* allocates enough funds for at least *k* activities [0/1]
ABOVE	Rated value of option ranked just better than *k* for participant *i*
MARGIN	Rated value of the k^{th} option for participant *i*
BELOW	Rated value of option ranked just worse than *k* for participant *i*

FIGURE 3.C5: ESTIMATES FROM NESTED MODELS USING ADJACENT VALUES FOR AVERAGES

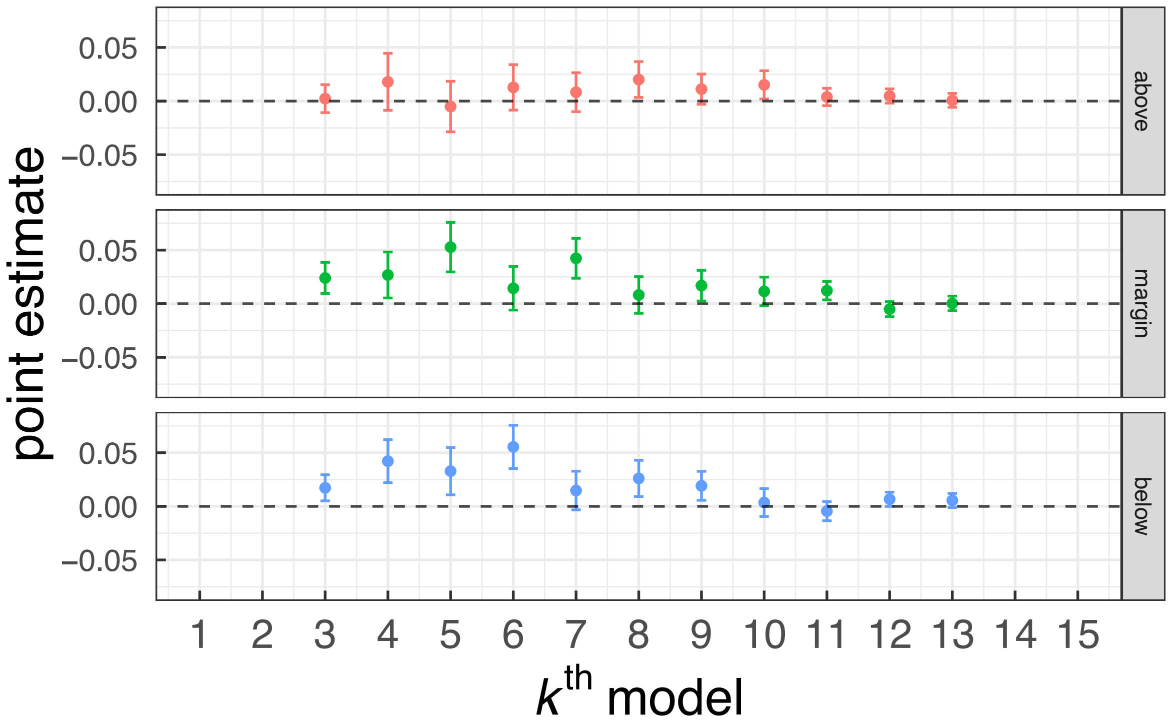


TABLE 3.C5.B: ESTIMATES FROM NESTED MODELS USING ADJACENT VALUES FOR AVERAGES

	Dependent variable:										
	dv3	dv4	dv5	dv6	dv7	dv8	dv9	dv10	dv11	dv12	dv13
ABOVE	0.002	0.02	−0.01	0.01	0.01	0.02**	0.01	0.02**	0.004	0.005	0.001
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.004)	(0.003)	(0.003)
MARGIN	0.02***	0.03**	0.05***	0.01	0.04***	0.01	0.02**	0.01*	0.01***	−0.01	0.0003
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.004)	(0.004)	(0.003)
BELOW	0.02***	0.04***	0.03***	0.06***	0.01	0.03***	0.02***	0.004	−0.004	0.01*	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.005)	(0.003)	(0.003)
Constant	0.59***	0.09	−0.003	−0.21***	−0.19***	−0.16***	−0.15***	−0.09***	−0.04*	−0.02	−0.02
	(0.06)	(0.11)	(0.10)	(0.08)	(0.06)	(0.05)	(0.04)	(0.03)	(0.02)	(0.01)	(0.01)
Observations	446	438	434	424	409	385	349	305	272	227	175

Note: +p<0.10;*p<0.05;**p<0.01;***p<0.001

Unnested Model

Unnesting the model to consider main effects:

TABLE 3.C5.C: ESTIMATES FROM UNNESTED MODEL USING ADJACENT VALUES FOR AVERAGES

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df
a	0.0056	0.0029	1.9293	0.0547	-0.0001	0.0112	278.3154
m	0.0200	0.0027	7.3482	0.0000	0.0146	0.0253	258.0804
b	0.0229	0.0031	7.3432	0.0000	0.0167	0.0290	288.1172

Note—Main effects from a single unnested model for the single value of the activity just ABOVE the margin (“a”), MARGIN (“m”), and the single value of the activity just BELOW the margin (“b”), estimated using *lm_robust()* with cluster-robust standard errors (at the participant level) and rank fixed effects (estimatr package in R).

C.6: Allocation Predicts Spending

Allocation Predicts Spending. We examine the hypothetical spending decisions by considering the 435/451 participants who considered more activities than their budget would allow (e.g., considered and provided value ratings to 12 activities, and set a $150 budget to accommodate up to 5 activities).[26] For these participants, activity value is a strong predictor of whether a given activity is selected for purchase, controlling for budget level and clustering standard errors at the participant level ($t(6764) = 56.31$, $p < .001$). Considering individual participants, 48% (208/435) purchased the optimal set of items (the highest ranked items, given their budget constraint) and the average proportion of optimal purchases across subjects was 84%. We take these findings as evidence that participants had stable and meaningful preferences (in terms of both values and ranks) that informed hypothetical purchase decisions in our paradigm.

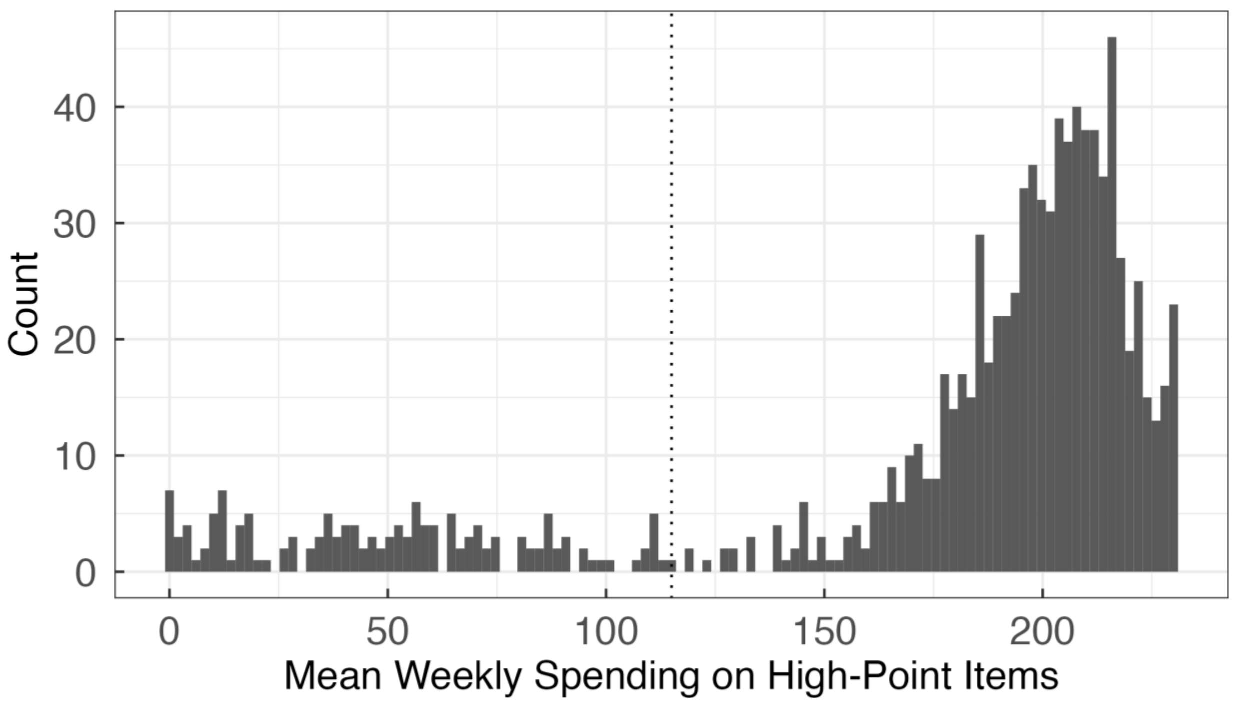
[26] For participants who considered *fewer* activities than allocated for, the purchased set is mechanically guaranteed to be the considered set.

APPENDIX D: ADDITIONAL MATERIALS FOR STUDY 3

Exclusions and Comprehension Checks

Exclusions. Figure 3.D1 depicts the distribution of spending on high-point purchases used for exclusion in study 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 3.D1: DISTRIBUTION OF MEAN WEEKLY HIGH-POINT SPENDING, USED FOR EXCLUSION



Note— Distribution of spending on high-point items in study 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 ("points," "entertainment purchases," or "dining purchases").

- *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" rather than "budgets must be followed exactly." (This question was only asked of 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 answered "True."

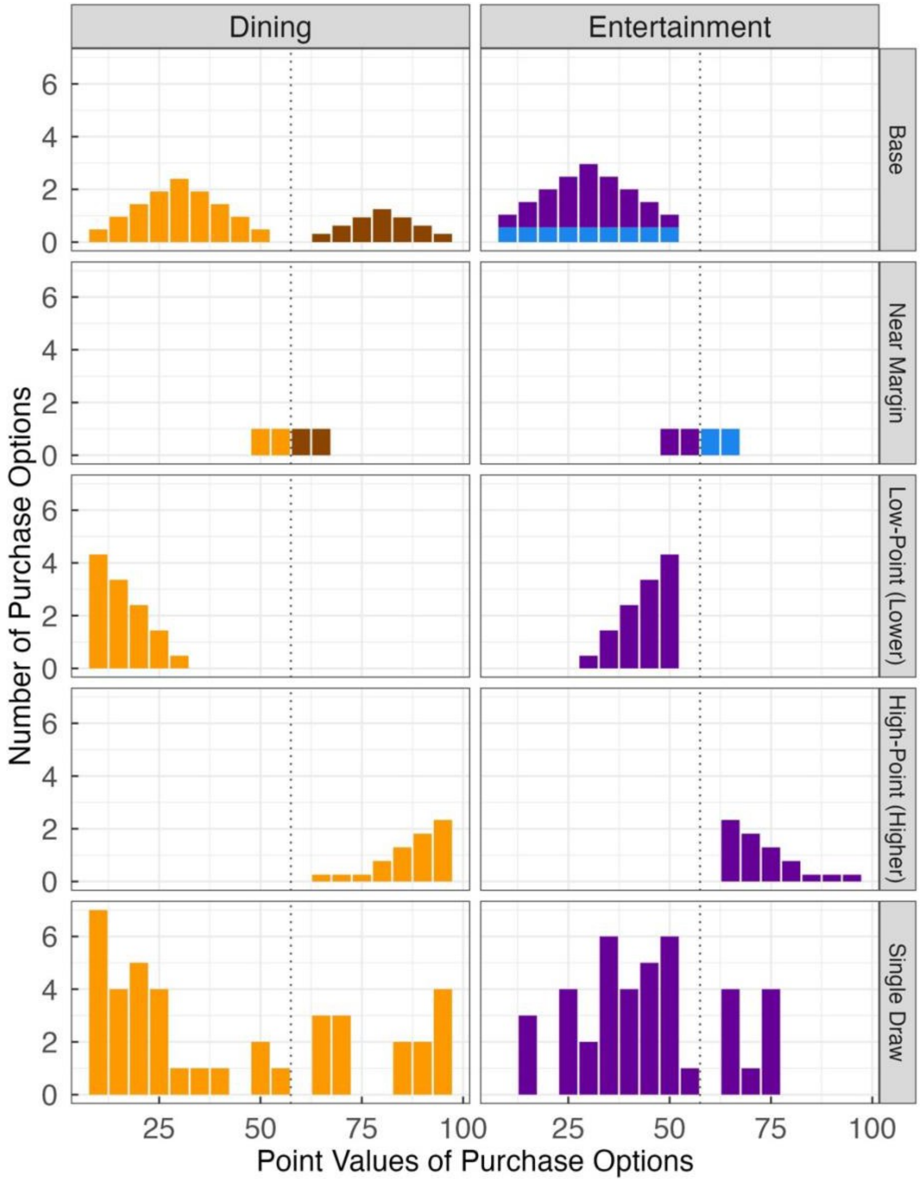
- *Q3:* "Any unspent money will carry over to the following week": 98% correctly answered "False."

- *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 3.D2 depicts the theoretical distributions from which items were drawn in study 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 3.D2: THEORETICAL POINT DISTRIBUTIONS IN STUDY 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 the analyses accompanying each hypothesis. Furthermore, we include both the dining share measure reported in the manuscript (for ease of explication), as well as the difference measure (as preregistered)

First Analysis (H1). We present the regression output for the full preregistered model in table 3.D1.A. As discussed in the manuscript, this model was constrained to those in the budgeting condition and regressed the dining share of allocation on both the dining average, the low-point dining average, and the interaction. 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 3.D1.A: REGRESSION RESULTS TO ACCOMPANY TEST OF H1

	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.

Additional Exploratory Analysis: Sensitivity to Overall Average Value. While our experimental design allows for the careful estimation of sensitive to average value in the high-point and low-point regions of the distribution, we can also approximate the sensitivity to overall average value. A simple way to conduct this approximation is to reconceptualize the original 2 (dining average: high, low) x 2 (low-point dining average: high, low) as a 3-condition, between-subjects design (high averages in both regions, mixed high and low, low averages in both regions). We reconsider the test of H1 after constructing a new variable for overall average, reflecting this condition assignment (+1 = both high averages, 0 = mixed high and low averages, -1 both low averages). Results are presented in Table 3.D1.B.

TABLE 3.D1.B: REGRESSION RESULTS TO ACCOMPANY TEST OF H1

	Dependent variable:
	Dining share
Overall average	3.439***
	p = 0.00001
Constant	55.015***
	p = 0.000
Observations	394
R^2	0.059
Adjusted R^2	0.057
Residual Std. Error	9.731 (df = 392)
Note:	+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

Note—Exploratory analysis suggests budgeters are sensitive to the overall average value, constructed from original condition assignment across the 2x2 design.

Second Analysis (H2). As discussed in the manuscript, the test of H2 uses the same model as that to test H1 (including being constrained to only those in the budget condition), with one difference. Whereas H1 considers the dining share of *allocation* as the dependent variable, H2 considers the dining share of *spending*. For completeness, we present results with and without preregistered exclusions, as well as using both the proportional dining share measure (from the manuscript) as well as the preregistered difference measure (table 3.D2).

TABLE 3.D2: REGRESSION RESULTS TO ACCOMPANY TEST OF H2

	Dependent variable:			
	Dining Share of Spending		Difference Measure	
	With Exclusions	Without Exclusions	With Exclusions	Without Exclusions
	(1)	(2)	(3)	(4)
Dining average	2.431***	2.232***	10.863***	9.710***
	p = 0.000	p = 0.00000	p = 0.000	p = 0.000
Low-point dining average	1.241**	0.948*	5.617***	4.912**
	p = 0.002	p = 0.022	p = 0.001	p = 0.004
Din. avg. x low din. avg.	−0.182	−0.381	−0.712	−1.830
	p = 0.627	p = 0.354	p = 0.668	p = 0.271
Constant	57.225***	57.272***	32.512***	30.502***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	394	478	394	478
R^2	0.120	0.071	0.123	0.086
Adjusted R^2	0.114	0.065	0.116	0.080
Residual Std. Error	7.429 (df = 390)	8.975 (df = 474)	32.858 (df = 390)	36.256 (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.

Third Analysis (H3). Though not preregistered, we can consider the dining share of *allocation* among budgeters and the dining share of *spending* among non-budgeters as a single dependent measure. Specifically, the dining share reflects [dining dollars / total allocated dollars] x 100% for budgeters and [dining dollars / total spent dollars] x 100% for non-budgeters. This allows us to include all participants (those assigned to both conditions) in the analysis. We regress this dining share on the dining average (+1 = high, -1 = low), the low-point dining average (+1 = high, -1 = low), the budget condition assignment (+1 = budgeting, -1 = non-budgeting), and all two- and three-way interactions. Full results are presented in table 3.D3. The variable of interest is the dining average-by-budget interaction. Columns 3 and 4 include the difference measure.

TABLE 3.D3: REGRESSION RESULTS TO ACCOMPANY TEST OF H3

	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 H3. 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 non-budget condition.

Fourth Analysis (H4). The full regression output for our preregistered test of H4 is presented in table 3.D4. The dependent variable is the dining share of spending, and the focal preregistered variable is the dining average-by-budget interaction. Aside from the differences in the dependent variable, this model is identical to the model to test H3 (table 3.D3).

TABLE 3.D4
REGRESSION RESULTS TO ACCOMPANY TEST OF H4

	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—Columns 1 and 3 apply the preregistered exclusions and columns 2 and 4 considers the full data, without exclusions.

Budgets Predict Spending in Study 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.

Points and Bonuses: Exploratory Analysis. As additional exploratory analyses, we consider both total points (accumulated throughout the incentivized five-week game) and total bonus payments as predicted by condition assignment. Bonuses were earned each week for scores of at least 1,560 points. Specifically, a participant earned an additional $0.01 bonus for every 10 points they scored above 1,550 points in a given week. Therefore, participants had five opportunities (five weeks) to earn non-zero bonuses, which accumulated to the final bonus payout (mean = \$0.75, median = \$0.80). The dining average condition and the low-point dining average condition did not affect points or bonuses (table 3.D5). Note that the distributions were designed for bonuses to be equivalent across conditions, in expectation (given optimal spending).

TABLE 3.D5: ANALYSIS OF POINTS AND BONUSES

	Dependent variable:	
	Points	Bonus
	(1)	(2)
Dining average	−0.772	−0.0003
	p = 0.965	p = 0.980
Low-point dining average	−1.233	−0.003
	p = 0.943	p = 0.789
Budget	−86.393***	−0.077***
	p = 0.00000	p = 0.000
Din. avg. x low din. avg.	10.187	−0.001
	p = 0.552	p = 0.916
Din. avg. x budget	−0.503	0.0004
	p = 0.977	p = 0.971
Low din. avg. x budget	21.489	0.011
	p = 0.210	p = 0.271
Din. x low x budget	−16.603	−0.014
	p = 0.333	p = 0.170
Constant	8,443.077***	0.753***
	p = 0.000	p = 0.000
Observations	821	821
R^2	0.033	0.070
Adjusted R^2	0.025	0.062
Residual Std. Error (df = 813)	489.650	0.286
Note:	+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001	

Note—Points (left) and bonuses (right) as a function of condition assignment.

There was an effect of being assigned to the budgeting condition, such that budgeters earned fewer points and smaller bonuses. We suspect this relates to the observed tendency towards naïve diversification (the preference to evenly split funds between dining and entertainment budgets). Specifically, if budgeters express a preference for naïve diversification in allocation (allocating near the 50%-50% split), then because budgets predict spending (as previously discussed), then budgeters' spending will be pulled away from the 61%-39% optimal split (corresponding to $140 to dining and $90 to entertainment; see manuscript figure 7 and 8).

Time Trends. We designed study 3 as a multiperiod incentivized game in which participants had ample opportunity to learn. One potential concern is whether participants had not fully learned or developed a decision strategy during the analyzed game weeks. Most concerning would be a pattern reflecting diminishing sensitivity to average value, over time, as participants learned the game dynamics. In exploratory analyses, we do not find evidence of such a pattern. Table 3.D6 presents five independent analyses of the five incentivized game weeks.

TABLE 3.D6: WEEKLY REGRESSION RESULTS

	Dependent variable:				
	Dining Share	Dining Share	Dining Share	Dining Share	Dining Share
	(1)	(2)	(3)	(4)	(5)
Dining average	2.319***	2.431***	2.685***	2.121***	3.476***
	p = 0.00005	p = 0.00005	p = 0.00001	p = 0.0003	p = 0.00000
Low-point dining average	0.958+	1.474*	0.537	1.054+	0.139
	p = 0.091	p = 0.013	p = 0.323	p = 0.065	p = 0.820
Din. avg. x low din. avg.	−0.266	0.065	0.312	0.094	0.120
	p = 0.639	p = 0.913	p = 0.566	p = 0.869	p = 0.844
Constant	54.597***	55.118***	54.771***	55.165***	55.265***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Observations	394	394	394	394	394
R^2	0.049	0.057	0.062	0.043	0.078
Adjusted R^2	0.041	0.050	0.055	0.035	0.070
Residual Std. Error (df = 390)	11.196	11.612	10.749	11.278	12.061
Note:	+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001				

Note—The dining share of allocation regressed on the contrast-coded condition variables for dining average (high = 1; low = -1), low-point dining average (high = 1; low = -1), and their interactions. Columns 1-5 correspond to incentivized game weeks 1-5. Conceptually, this corresponds to the model to test H1 (table 3.D1.A) with data disaggregated to the weekly level.

Each analysis regresses the dining share of allocation on the dining average and low-point dining average conditions (and their interaction). Across each of these models (corresponding to weeks 1-5), we observe greater allocations to the dining category in the higher dining average condition, and this effect is no smaller in the last week than the first incentivized week.

Building upon this approach, we construct a time-trend outcome variable that reflects the dining share across the five game weeks, each weighted by a linear contrast code to capture time trends. Specifically, we used the contrast weights -2, -1, 0, +1, +2 to correspond to weeks 1, 2, 3, 4, 5, respectively. We considered participants in the budgeting condition, and their dining share in each week was multiplied by the corresponding contrast weights. We then summed this variable. Negative values reflect greater dining allocations *earlier* in the five-week period, and positive values reflect greater dining allocations *later*. We regressed this time-trend on the condition variables (and their interaction). Thus, this approach allows us to test whether the dining share grew or shrank over time based on condition assignment. A positive coefficient on dining average indicates dining allocations grew throughout the five-week game in the high-average condition more than in the low-average condition. A negative coefficient on the low-point dining average indicates dining allocations shrank throughout the five-week game in the high-average condition relative to the low-average condition (in the low-point distribution). There was no tendency for the sensitivity to dining average to decrease across the five incentivized weeks; if anything, there was a marginally significant tendency for sensitivity to dining average to increase (table 3.D7).

TABLE 3.D7: TIME-TREND IN DINING SHARE

	Dependent variable:
	Time-trend outcome
Dining average	0.020^{+}
	p = 0.089
Low-point dining average	-0.021^{+}
	p = 0.080
Din. avg. x low din. avg.	0.008
	p = 0.495
Constant	0.014
	p = 0.239
Observations	394
R^2	0.016
Adjusted R^2	0.009
Residual Std. Error	0.233 (df = 390)
Note:	+ p<0.1; * p<0.05; ** p<0.01; *** p<0.001

Note—The time-trend outcome variable reflects the extent to which a participant's dining share was growing over time (positive value) or shrinking over time (negative value). This was constructed as the sum of the five weekly dining shares, weighted by the contrast weights, corresponding to weeks 1-5: -2, -1, 0, 1, 2. This measure was then regressed on the same set of predictors as in table 3.D1.A for participants in the budgeting condition (used to test H1).

CHAPTER 3 REFERENCES

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(13), 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.

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

Benartzi, Shlomo and Richard H. Thaler (2001), "Naive diversification strategies in defined contribution saving plans," *American economic review*, 91(1), 79-98.

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.

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(4), 393–404.

Colander, David. (2019), *Economics. 11th Edition*, New York: McGraw-Hill.

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

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

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(6), 109–131.

Fernbach, Philip M., Christina Kan, and John G. Lynch (2015), "Squeezed: Coping with Constraint through Efficiency and Prioritization," *Journal of Consumer Research*, 41(5), 1204–27.

Gathergood, John, Neale Mahoney, Neil Stewart, and Jörg 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), R751-R753.

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

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.

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.

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*, 104(2), 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 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.

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.

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

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., J. Jeffrey 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.

Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance*, 16(2), 366–87.

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 University Press.

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

Samuelson, Paul and Nordhaus, William (2009), *Economics. 19th Edition*, New York: McGraw-Hill.

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 (2024), "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(1), 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, 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*, 53(5), 790–803.

Thaler, Richard H. (1980), “Toward a Positive Theory of Consumer Choice,” *Journal of Economic Behavior & 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 Research*, 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(1), 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 K. Lyu (2022), “How Consumers Budget,” *Journal of Economic Behavior & Organization*, 204, 69–88.