# A PROTOTYPE THEORY OF CONSUMER EXPENSE MISPREDICTION

by

RAY CHARLES “CHUCK” HOWARD

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

A Prototype Theory of Consumer Expense Misprediction

submitted by Ray Charles "Chuck" Howard in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Business Administration

**Examining Committee:**

Dale Griffin, Marketing and Behavioural Science

Supervisor

David Hardisty, Marketing and Behavioural Science

Supervisory Committee Member

Elizabeth Dunn, Psychology

Supervisory Committee Member

Jiaying Zhao, Psychology

University Examiner

Hasan Cavusoglu, Accounting and Information Systems

University Examiner

## Abstract

The present research develops a prototype theory of consumer expense misprediction that helps explain why consumers display an *expense prediction bias* in which they *under*-predict their future spending, and how expense prediction accuracy can be improved. The logic of the prototype theory is that expense predictions are based on prototype attributes that come to mind easily when predictions are being constructed. These attributes represent a consumer's average spending, where "average" refers to the mode of their expense distribution. This leads consumers to under-predict their expenses because, generally speaking, the distribution of expenses is positively skewed with mode < mean. Accordingly, it is proposed that prompting consumers to consider reasons why their expenses might be different than usual will increase prediction accuracy by making atypical expenses cognitively easier to retrieve. A series of studies that includes a longitudinal field experiment, data from a personal finance app, and a variety of lab paradigms provide evidence for this prototype account of the bias and the effectiveness of the proposed intervention. Evidence is also provided that consumers with variable income (e.g., Uber drivers) display a corresponding *income prediction bias* in which they *over*-predict future income.

**Lay Summary**

The primary goal of this dissertation is to better understand why consumers under-predict their future expenses and how their expense predictions can be made more accurate. When consumers predict their future expenses (e.g., for the next week or month) they behave as though they are answering the question "what expenses do I typically incur?" This leads to under-prediction because in reality consumers very often encounter both typical and atypical expenses. However, prediction accuracy can easily be improved by prompting consumers to consider reasons why their expenses for the next week or month might be different from a typical week or month.

## Preface

This dissertation is original, unpublished work by Ray Charles "Chuck" Howard, conducted under the supervision of Professors Dale Griffin and David Hardisty at the University of British Columbia, and with guidance from Professor Abigail Sussman at the University of Chicago. Data collection was covered by UBC Ethics Certificates H16-00689 and H17-00241.

**Table of Contents**

**List of Tables**

**List of Figures**

**Acknowledgements**

I am deeply grateful to Dale Griffin and Dave Hardisty for their mentorship, support, and unending patience. I thank Liz Dunn, Abby Sussman, John Lynch, and Leaf Van Boven for their very helpful feedback on this research program at various stages of its development. I thank Darren Dahl for his tough love and for helping me understand the long game. I thank Joey Hoegg for her kindness and generosity. I thank Karl Aquino and the Immorality Lab for pushing my thinking in weird new ways. I thank Marcel Lukas for his friendship and laughter as we've learned together what it takes to do this job. I thank all the UBC students who helped me grow as a researcher, especially Johannes Boergershausen. I would also like to thank Emanuel, Judith, Anais, Lucia, Anouk, Claire, and Gizem for making the office a little more lively during their visits to UBC. I thank Mom, Dad, Tee, James, and Emma for their love and support; the only thing I'd change about my PhD experience is being so far away from home. Finally, I thank my loving wife Meghan for helping me stop to smell the roses.

## Dedication

This work is dedicated to my niece Emma, and the loving memory of her sister Hope.

## 1. Introduction

Consumers frequently try to predict their future expenses (Peetz et al. 2016; Canadian Task Force on Financial Literacy 2019), and the accuracy of these predictions can be highly consequential. In particular, the available evidence suggests that *under*-predicting future expenses can be costly. For example, over 25% of Americans with a 401(k) savings account end up breaching their account and withdrawing money before retirement, very often for the sole purpose of covering unpredicted expenses (Fellowes and Willemin 2013). These breaches cost consumers approximately $7 billion a year in penalties. Similarly, each year almost 2 million Americans use a payday loan to cover an unpredicted expense (Pew 2012). The annual interest rate on these loans frequently exceeds 400% (Consumer Federation of America 2018). Many consumers also hold the expectation that they will be able to pay off their credit card balance each month (Yang et al. 2007). Yet American consumers collectively hold over one trillion dollars in credit card debt and pay associated interest costs (Federal Reserve Bank of New York 2018).

These examples suggest that increasing expense prediction accuracy can help consumers spend, save, and/or borrow money in a more efficient manner. An accurate assessment of future expenses can help consumers better allocate funds between their checking and 401(k) accounts to avoid penalties for early withdrawal. And if consumers had a clearer idea of how much they would spend in the future, they might choose to spend less in the present to avoid the costs associated with borrowing or using credit to cover expenses down the road. The prosocial value of helping consumers avoid these costs is self-evident. The rush by venture capitalists to fund Financial Tech start-ups developing apps that help consumers manage their expenses (CB Insights 2018) indicates there is also firm value in improving expense prediction accuracy.

Echoing the real-world examples offered above, academic research also suggests that consumers tend to under-predict their future expenses (e.g., Ulkumen et al. 2008; Peetz and Buehler 2009), a phenomenon I label the *expense prediction bias*. The goal of the present research is to identify a key psychological driver of this bias, then leverage that theoretical insight to develop, test, and validate a practical intervention that improves consumers' expense prediction accuracy. To do so, I first theorize that expense predictions are based on prototype attributes that approximate the mode of a consumer's expense distribution. I then propose that "prototypical prediction" causes consumers to under-predict their expenses because, generally speaking, the distribution of consumer expenses is positively skewed with mode < mean. Finally, I show that prompting consumers to consider reasons why their upcoming expenses might be different than usual increases prediction accuracy by making atypical expenses cognitively easier to retrieve.

By developing and testing a prototype theory of consumer expense misprediction, this dissertation makes the following contributions. First, the prototype theory offers a parsimonious explanation for why the expense prediction bias occurs. Second, this work introduces what is to the best of my knowledge the first practical, field-tested intervention capable of reducing or eliminating the expense prediction bias. Third, the present research provides the first step toward a comprehensive understanding of the bias itself. For example, this research is the first to identify the magnitude, prevalence, and persistence of the bias in non-student samples. It is also the first work to study the bias longitudinally and in the field, and to measure monthly expense predictions against actual expenses incurred during the target month.

This research program also makes two contributions to the broader literature on consumer misprediction. First, it highlights the important role that distributional skew can have with respect to prediction accuracy. Second, where past research demonstrates that predictions do not sufficiently incorporate distributional information (e.g., Buehler, Griffin, and Ross 1994; Kruger and Evans 2004), the present research helps elucidate *what* distributional information is neglected. Similarly, the present research contributes to the literature on the use of prototype attributes in judgment and decision making by providing evidence that these attributes represent mode (rather than mean) outcomes in the context of expense prediction. Finally, the present research contributes to a developing literature that demonstrates a temporal asymmetry in which people mentally represent the future in more prototypical terms than the past (Kane, Van Boven, and McGraw 2012; Van Boven, Kane, and McGraw 2008). By comparing the nature of predicted versus recalled expenses, I extend this work to the domain of money.

The remainder of this dissertation is organized as follows: In Section 2 I develop my theoretical framework and introduce my hypotheses. In Section 3 I present a multi-method series of studies that test my hypotheses. In Section 4 I discuss the implications of this research for theory and practice, as well as its limitations and directions for future research.

## 2. Developing a Prototype Theory of Consumer Expense Misprediction

The prototype theory of consumer expense misprediction developed in the present research is organized around four propositions. First, I propose that consumers' expense predictions are based on *prototype attributes* that represent average expenses, because these prototypic attributes come to mind most easily when predictions are being made. Second, I propose that the "average" represented by prototype attributes approximates the mode of a consumer's expense distribution rather than the mean.[1] Third, I propose that the process of "prototypical prediction" described by propositions one and two leads consumers to under-predict their future expenses because, generally speaking, the distribution of consumer expenses is positively skewed with mode < mean. Fourth, I propose that prompting consumers to consider reasons why their expenses might be different than usual will increase prediction accuracy because most values in a positively skewed distribution that are different than the mode are larger than the mode.

To understand the intuition behind these propositions consider a consumer named Meghan who incurs the expenses summarized as a calendar in Table 1, and as a stylized density curve (i.e., a smoothed histogram) in Figure 1. The prototype theory proposes that when Meghan predicts her weekly expenses her groceries, fuel, coffee, and lunch will come to mind first because these are the expenses she incurs most regularly. Moreover, the prototype theory proposes that when Meghan predicts how much she's going to spend on each of these expenses she will predict about $100 for groceries, $45 for fuel, $35 for coffee, and $20 for lunch because these values approximate the amount she most typically spends on these items. This process of

[1] Although the term average is most commonly used to refer to the mean of a distribution, its formal definition also pertains to the mode (Oxford English Dictionary 2019).

"prototypical prediction" means that Meghan's baseline expense prediction will be $200/week, which approximates the mode of her density curve in Figure 1.[2]

## Table 1

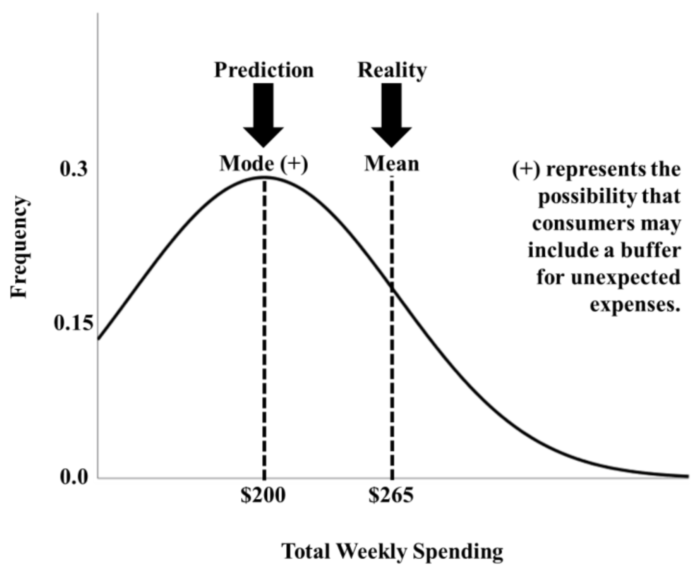
## Calendar Summary of "Meghan's" Expenses

| **Week 1** | | **Week 2** | | **Week 3** | | **Week 4** | | **Week 5** | | **Week 6** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Groceries | $102.31 | Groceries | $99.43 | Groceries | $101.98 | Groceries | $105.61 | Groceries | $104.29 | Groceries | $100.38 |
| Fuel | $45.75 | Fuel | $47.61 | Fuel | $43.21 | Starbucks | $34.16 | Fuel | $43.82 | Fuel | $48.52 |
| Starbucks | $34.02 | Starbucks | $35.17 | Starbucks | $33.84 | Friday work lunch | $19.74 | Starbucks | $34.97 | Starbucks | $36.18 |
| Friday work lunch | $18.98 | Friday work lunch | $17.26 | Friday work lunch | $21.17 | | | Friday work lunch | $23.80 | Friday work lunch | $19.53 |
| | | Car repair | $254.32 | | | | | New clothes | $178.65 | | |
| **Total** | **$201.06** | **Total** | **$453.79** | **Total** | **$200.20** | **Total** | **$159.51** | **Total** | **$385.53** | **Total** | **$204.61** |

| **Week 7** | | **Week 8** | | **Week 9** | | **Week 10** | | **Week 11** | | **Week 12** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Groceries | $115.68 | Groceries | $107.29 | Groceries | $110.28 | Groceries | $106.30 | Groceries | $104.29 | Groceries | $101.83 |
| Fuel | $49.81 | Fuel | $51.17 | Fuel | $49.16 | Fuel | 45.96 | Fuel | $43.82 | Fuel | $48.63 |
| Starbucks | $37.50 | Starbucks | $36.86 | Starbucks | $35.73 | Starbucks | $34.18 | Starbucks | $34.97 | Starbucks | $37.83 |
| Friday work lunch | $24.83 | Friday work lunch | $23.70 | Friday work lunch | $19.02 | | | Friday work lunch | $23.80 | Friday work lunch | $19.35 |
| | | Birthday dinner | $45.23 | | | | | Concert tickets | $250.00 | | |
| **Total** | **$227.82** | **Total** | **$264.25** | **Total** | **$214.19** | **Total** | **$186.44** | **Total** | **$456.88** | **Total** | **$207.64** |

## Figure 1

## Stylized Density Curve of "Meghan's" Total Spending in Table 1




[2] This example assumes that Meghan takes a "bottom-up" approach to prediction whereby she arrives at her final prediction by aggregating individual expenses. Note that if we instead assume that Meghan takes a "top-down" approach to prediction whereby she predicts a total dollar value without consciously considering individual expenses, the prototype theory still implies her prediction will be ~ $200, because $200 approximates the total amount of money she spends most regularly. Thus, in contrast to earlier theoretical accounts of expense under-prediction (Ulkumen et al. 2008) the prototype theory does not need to make any assumptions about whether consumers adopt a bottom-up or top-down approach.

Now let's consider Meghan's prediction accuracy under two different scenarios that assume her future spending will resemble her past spending. In scenario one Meghan does not foresee any atypical expenses and therefore believes her prototypical prediction of $200/week is sufficient. Unfortunately for Meghan, this means she will under-predict her expenses by almost 25% on average, which in this example amounts to $65/week or $260/month. In scenario two let's imagine that Meghan is a more conscientious consumer, so although she uses the mode of her expense distribution as a baseline for her predictions, she also recognizes that she should include a buffer to account for unexpected expenses. The available evidence suggests that including a buffer like an "extra" $20 for miscellaneous expenses will indeed help Meghan increase her prediction accuracy. However, if Meghan is like most consumers, her buffer will not be enough to prevent under-prediction (Ulkumen et al. 2008), because it will not adequately account for the frequency of atypical expenses (Peetz and Buehler 2013; Peetz et al. 2015) and/or the true cost of these expenses (Sussman and Alter 2012).

Fortunately for Meghan, the prototype theory suggests a simple solution to improve prediction accuracy. If Meghan considers reasons why her expenses might be different than usual (i.e., different than the mode), then she will produce a more accurate prediction because most of a positively skewed distribution that is different than the mode is also higher than the mode. In other words, considering reasons why her expenses might be different than usual will move her prediction from the mode toward the mean so that her prediction more closely resembles reality.

In the following subsections I discuss each proposition of the prototype theory in relation to past research, then present my theoretical model and hypotheses.

## 2.1 Prototype Attributes and Predictions

The dominant theoretical perspective in the psychological literature on prediction is that prediction biases occur in large part because people fail to incorporate relevant past experience when predicting the future. This perspective is rooted in the seminal work of Daniel Kahneman and Amos Tversky who built their program of research on heuristics and biases by studying intuitive prediction and the systemetic errors that result. In this context, the term "prediction" was most commonly used to reference judgments related to category membership or outcome probability, such as predicting the likelihood that an individual belongs to a particular group. One foundational insight generated by their work in this area is that both sophisticated and naïve predictors tend to overweight "singular information" that represents evidence about the specific outcome under consideration, and underweight "distributional information" that represents knowledge about the distribution of past outcomes in similar situations (Kahneman 2003; Kahneman and Tversky 1973, 1977). For example, in their research on the representativeness heuristic Kahneman and Tversky observed that people answer questions like "will Chuck pass his dissertation defense?" by evaluating the degree to which Chuck is representative of (i.e., similar to) the typical student who passes, rather than considering the proportion of students who typically pass (Kahneman and Tversky 1972; Kahneman and Tversky 1973; Tversky and Kahneman 1974). In other words, predictions in this case do not include informative base rate information about the frequency with which a target outcome occurs in the population.

The idea that predictions largely neglect past behavior has since been extended to work on behavioral predictions. For example, the "inside-outside" account of the planning fallacy proposes that people under-predict their project completion times because they adopt an "inside-view" based on a narrow set of future plans rather than an "outside-view" based on past experience (Buehler, Griffin, and Ross 1994; Kahneman and Tversky 1979). As a result, people under-predict their project completion times even when they are well aware that similar projects have taken longer than planned in the past (see Buehler, Griffin, and Peetz 2010 for a review). Similarly, affective forecasting theory argues that individuals over-predict the duration of their affective reactions to future events in large part because they over-focus on the prospective event in question (Dunn, Gilbert, Wilson 2011; Gilbert and Wilson 2007; Gilbert and Wilson 2009; Wilson et al. 2000). Hence, this "durability bias" is observed even when the event is something forecasters have ample experience with, such as a favorite sports team winning or losing a big game (Wilson et al. 2000). More broadly, prediction biases across phenomena as diverse as prosocial behavior, voting, and relationship endurance have all been attributed in part to instantiations of base-rate neglect in which available information about typical outcomes is discounted or ignored (Dunning, Griffin, Milojkovic, and Ross 1990; Epley and Dunning 2000; Vallone, Griffin, Lin, and Ross 1990).

The perspective that predictions largely neglect past behavior has also been invoked in more nascent research on expense prediction accuracy (Peetz and Buehler 2009, 2012). In the present research I develop an alternative account that proposes expense predictions are in fact deeply rooted in past behavior. Specifically, I propose that expense predictions are based on *prototype attributes*. Consistent with past research, I define prototype attributes as representations of an average (Kahneman 2003; Kahneman and Frederick 2002). So, in the context of expense prediction, prototype attributes can be conceptualized as the response to internal queries like "what expenses do I *typically* incur each [week/month]?" or "how much do I *typically* spend each [week/month]?" In contrast, the target in expense prediction is the response to some variation of the query "how much will I *actually* spend next [week/month]?"

The proposition that expense predictions are based on prototype attributes (hereafter referred to as the "prototype proposition") is derived from the observation that the target in expense prediction is relatively low in accessibility, meaning it does not come to mind without deliberation or effort. To illustrate this point, consider that a comprehensive expense prediction requires anticipating all possible future expenses, estimating the amount of each expense and the probability it will occur, and adding these probability-weighted amounts together. In contrast, when an individual is repeatedly exposed to variations of a stimulus – as is the case with average, regularly occurring expenses – prototypes attributes that represent accurate impressions of an average can be formed with relative ease, and once formed, they are highly accessible (Ariely 2001; Kahneman 2003; Kahneman and Frederick 2002; Rosch and Mervis 1975). Thus, the prototype proposition is broadly consistent with the finding that accessibility often determines the content of judgments and decisions (e.g., Epley and Gilovich 2006; Gilovich, Griffin, and Kahneman 2003; Johnson, Häubl, and Keinan 2007; Kahneman 2003; Tversky and Kahneman 1973, 1974).

The prototype proposition also integrates – and has the potential to explain – three key results in the expense misprediction literature. For example, previous work in this area has demonstrated that consumers' expense predictions do not fully incorporate either their distribution of actual past expenses or their distribution of possible future expenses (Peetz and Buehler 2012; Peetz et al. 2015). This kind of *distribution neglect* is easily explained by the prototype proposition because prototype attributes represent a relatively thin slice of a consumer's expense distribution (i.e., the mode). Relatedly, it has been shown that consumers behave as if their atypical or exceptional expenses will not reoccur (Sussman and Alter 2012). This can be explained by the prototype proposition because the "thin slice" that prototype attributes represent does not include these expenses in the right tail of the distribution. Finally, it has been suggested that consumers under-predict their future expenses in part because they are overconfident in their prediction accuracy, and as a result they do not sufficiently adjust their initial predictions upward (Ulkumen et al. 2008). This possibility can be explained by the prototype proposition because confidence often reflects cognitive ease (Alter and Oppenheimer 2009), and prototypes are "easy on the mind" (Winkielman et al. 2006).

### 2.1.1 Think Aloud Protocol Pilot Study

To explore the prototype proposition and help illuminate the cognitive processes that underlie expense prediction, I initially undertook a think-aloud protocol study. Fifty-five undergraduate commerce students at the University of British Columbia were recruited to take part in a study about consumer financial decision making. Each participant was taken to a private room where they received written instructions to "Say aloud every thought that enters your mind as you think about the following question, decide on your answer, and even as you write down your answer" (Buehler, Griffin, and Ross 1994). Participants were then asked "How much do you estimate your total expenses will be for the next week (i.e., the next 7 days)?" With informed consent, each participant's verbal protocol was recorded and later transcribed and coded by two research assistants. Specifically, the research assistants independently coded each protocol for whether or not a participant referenced prototypical expenses (e.g., "average," "typical," "normal," or "usual" expenses), future-oriented expenses (i.e., expenses specific to the next week), or an adjustment or "buffer" for unexpected expenses. The research assistants also coded which type of thought came to mind first.

Table 2 presents the results of the think aloud study with verbatim examples. A significantly higher proportion of participants referenced prototypical expenses than future-oriented expenses (Mean difference = 29.09%, 95% CI = [11.88%, 44.12%], $\chi_{(1)} = 10.80$, $p$ = .001). In other words, when given the task of estimating their future expenses people were more likely to mention what they *usually* spend, rather than what they *would* spend. Prototypical expenses also came to mind first for a strong majority of participants (67.27%, $z = 2.56$, 95% CI = [53.29%, 79.32%], $p$ = .010). Finally, only about half of the sample made a conscious adjustment for unexpected expenses. Taken together, these results provide preliminary support for the prototype proposition by demonstrating that prototype attributes are both easily accessible and a foundational component of expense predictions.

**Table 2**

**Results of the Think Aloud Protocol Pilot Study**

| Classification | Proportion | First Thought | Examples |
|---|---|---|---|
| Typical | 83.64% | 67.27% | "**Typically** I buy groceries every week. Thats about $50 dollars or so." "On **average**, I would say I spend around $10 per day on food and drinks." "So on **average** I spend $500 per month." "On **average** I buy mainly a coffee or a lunch on campus so let's say I spend like up to $10-ish maybe." "On **average**, I spend around $45 to $60 with food." "**Normally** I will spend uh approximately $20 per day for food." "So groceries are **normally** $100." "So I **usually** buy my lunch, cuz I have to." "On Friday I **usually** get gas so that's **usually** thirty dollars a week." "I **usually** for grocery shopping I spend, umm, around $30 a week." |
| Future-Oriented | 54.55% | 32.73% | "Huh, I'm traveling **next week** too, traveling is…I'll say $400, yeah." "Um **the next week, the next seven days.** Let's see...so this weekend I'm going...where am I going this weekend? So I'm going to ****'s party so that's probably around let's say $15 of things." "**This Sunday**, I might go to the mall to get new work clothes for my co-op, so that might be dress shoes, that might be maybe $120." "Are there any birthdays **coming up**? Oh wait my brother's birthday. He is an expensive person so that's going to be about $300." |
| Adjustment | 50.91% | 0.00% | "I might have **other expenses**, little things, stuff from vending machines, probably adds up to about $10 over a week." "I'll put about $20 for like **miscellaneous items**." "And just for **miscellaneous** items I would put another $10." "And then, shopping...**miscellaneous**, we'll just budget $50 for that." |

## 2.2 Prototype Attributes and "Average" Expenses

If expense predictions are based on prototype attributes that represent average or typical expenses – as the results of the think aloud study suggest – then it is necessary to ask next *what is the "average" that prototype attributes represent?* The word "average" is most commonly used to refer to the mean of a distribution of outcomes (Oxford English Dictionary 2019), and there are reasons to believe that is what consumers may be doing when they reference their average expenses. Theoretical work regarding the use of prototype attributes in judgment and decision making has proposed that when people use an average as a heuristic it is the mean of a subset of relevant outcomes (Kahneman 2003; Kahneman and Frederick 2002). Consistent with this theoretical proposition, laboratory research has shown that people are remarkably adept at mean identification, not only with respect to numbers and prices (Andre, Reinholtz, and de Langhe 2017; Beach and Swenson 1966), but also with respect to faces that vary in emotional expression (Haberman and Whitney 2009) and sets of shapes that vary in size (Ariely 2001). Taken together, these findings suggest that mean identification might be a widely generalizable skill that can be deployed whenever people are presented with repeated stimuli, as is the case with many consumer expenses.

However, there are also compelling reasons to believe that in the case of expenses prototype attributes do *not* represent a consumer's mean spending. First, from a theoretical perspective, prototype formation requires a certain degree of homogeneity within the set of relevant stimuli (Kahneman 2003). This degree of homogeneity might hold true in simple sets of numbers, faces that are varying degrees of happy or sad, or differently sized circles. However, calculating mean expenses requires incorporating heterogeneous outcomes (i.e., typical and atypical expenses). Furthermore, from an intuitive perspective, it seems unlikely that the human mind is equipped to keep a running tally of mean expenses, because the sheer number of expenses most consumers encounter in a busy day – much less a busy week or month – would make this dynamic calculation so cognitively taxing the mind would be hard pressed to accomplish much else. Incidentally, this logic also suggests that prototype attributes do not represent median expenses, because this would require our minds to engage in the presumably very effortful process of automatically ordering our expenses from lowest to highest (or vice versa) and continually identifying the $50^{th}$ percentile.

Thus, the second proposition of the present research is that the "average" represented by prototype attributes is akin to the *mode* of a consumer's expense distribution rather than the mean. This proposition is most directly supported by the finding that consumers are fairly accurate when predicting their ordinary, recurring expenses (Sussman and Alter 2012), because if predictions are based on protoypte attributes that represent mode expenses, then it follows that consumers should be relatively adept at predicting ordinary expenses which are modal by definition. This proposition is also broadly consistent with memory models in cognitive psychology (e.g., MINVERA-DM; Dougherty, Gettys, and Ogden 1999) and behavioral economics (e.g., Utility Weighted Sampling; Lieder, Griffiths, and Hsu 2017) which imply that people will overweight mode outcomes in their decision making because, in essence, the mode is the maximum likelihood estimator for any single decision event.

### 2.3 Prototypical Predictions and the Expense Prediction Bias

Past research on the psychology of prediction has acknowledged the detrimental impact that underweighting or ignoring distributional information can have on prediction accuracy (e.g., Buehler, Griffin, and Peetz 2010; Kahneman and Tversky 1977). However, the potential influence of distributional *skew* on prediction accuracy has received very little attention in this literature. Propositions one and two of the present research describe a process of "prototypical prediction" in which expense predictions are based heavily on prototype attributes that represent mode expenses. Proposition three is that prototypical prediction leads consumers to under-predict their expenses in large part because the distribution of consumer expenses is, generally speaking, positively skewed with mode < mean.

The expectation that expenses are positively skewed is derived from two observations. First, from a purely mathematical perspective, a consumer's expense distribution is bounded by zero on the left but free to run as high as their credit will allow on the right. Second, from a behavioral perspective, consumers appear to overspend their budgets with ease (Sussman and Alter 2012), but they have a significantly harder time spending less than they budget (Peetz and Buehler 2009).

Consistent with the proposition that skew impacts prediction accuracy, Table 3 confirms with field data from Studies 3 and 6 of the present research that the within-subject distribution of consumer expenses is positively skewed. The implications of this observation for improving prediction accuracy are discussed next.

**Table 3**

**Expense Distribution Skewness for the Median Consumer in Studies 3 and 6**

| Study # | 3 | 6 | | |
|---|---|---|---|---|
| Data Source | Vancity† | Money Dashboard†† | | |
| Expense Category | Total Weekly Expenses | Monthly Dining & Drinking Expenses | Monthly Grocery Expenses | Monthly Fuel Expenses |
| Sample Size | 187 | 2,420 | 2,574 | 1,099 |
| Observation Period | 5 weeks | 12 months | 12 months | 12 months |
| Mean Expenses | $665.89 | £302.02 | £408.41 | £142.54 |
| SD | 407.84 | 202.61 | 254.93 | 73.12 |
| **Skew*** | **1.03** | **0.38** | **0.24** | **0.10** |

*Skew is reported for the median user in each sample, as are the Means and SDs. The formula used to calculate skew is Peason's second skewness coefficient: 3(Mean - Median)/SD.

†Vancity is Canada's largest community credit union. Participants were recruited through an online panel of members they use to conduct market research. Participants were asked to self-report their total spending at the end of each week from their online bank account(s). Study details are provided in section 3.3.

††Moneydashboard is the UK's largest financial aggregation app. De-identified user data was provided by the app that allowed me to observe each user's spending for the 12 months *before* they downloaded the app. Study details are provided in section 3.5.

### 2.4 An "Atypical Intervention" to Improve Expense Prediction Accuracy

If the expense prediction bias is caused in part by a prototypical prediction process that largely neglects atypical expenses, then it follows that helping consumers bring atypical expenses to mind when they are constructing their predictions will increase their prediction accuracy. I reasoned that having people list three reasons why their expenses might be *different* than usual would serve as a simple intervention that accomplishes this goal, because most of a positively skewed distribution that is different than the mode is also higher than the mode.

Mechanistically, this "atypical intervention" bears some resemblance to the unpacking intervention derived from support theory, in which people are asked to unpack their prediction into its component parts (e.g., individual expenses) to elicit greater consideration of the distribution of possible future outcomes (Kruger and Evans 2004; Peetz et al. 2015; Tversky and Koehler 1994). There is however an important distinction between the atypical intervention and unpacking: Where the unpacking intervention prompts people to consider the full distribution of possible outcomes, the atypical intervention prompts them to consider outcomes that exist predominately in the right tail of the distribution. This is important from a theoretical perspective because the unpacking intervention says only that distributional information is missing from predictions. In contrast, the atypical intervention deepens our understanding of *what* distributional information is missing. The atypical intervention also carries a practical advantage: It requires consideration of only a handful of reasons why expenses may be atypical (vs. trying to unpack all possible expenses), which makes it easier to employ. This is noteworthy given that many expense predictions are made spontaneously (Peetz et al. 2016), which suggests that a simpler tool will be more widely used in practice. Finally, it is worth noting that some evidence supports the caveat that although the unpacking intervention improves mean prediction accuracy, it can sacrifice correlational accuracy (Kruger and Evans 2004; Peetz et al. 2015). In other words, while unpacking can increase predictions to better match the group mean of actual outcomes, it can make some people's prediction accuracy substantially worse. I speculate that one reason this may be true is that consumers who invest the effort that unpacking requires tend to make higher predictions, while those who invest very little effort make a lower prediction than they would have otherwise. Therefore, I posit that the simplicity of the atypical intervention will help it avoid this fate.

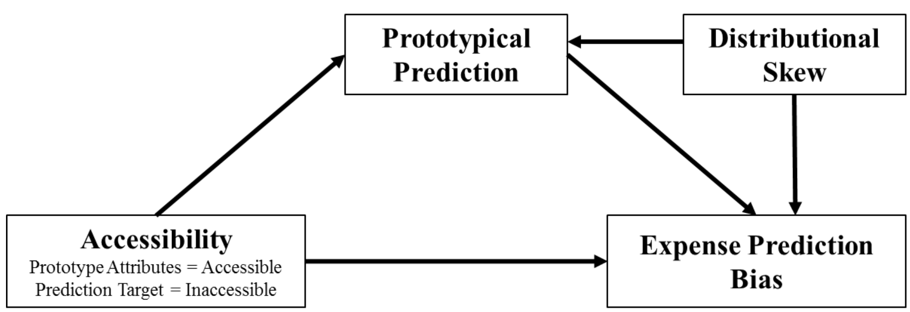
The logic underlying the atypical intervention also mirrors research demonstrating that "defocalizing" can improve affective forecasts. For example, Wilson et al. (2000) found that people induced to think about their post-event daily routines before they predicted how that event would make them feel were subsequently less likely to over-predict their affective reaction to the event. Because affective prototypes tend to be somewhat extreme (Gilbert et al. 2004; Schkade and Kahneman 1998; Wilson et al. 2000), a reasonable interpretation of this finding is that having people consider non-prototypical information lessened their reliance on prototype attributes and in doing so improved their prediction accuracy. The same logic applies to consumer expense predictions: while people naturally consider typical (i.e., mode) expenses, and therefore under-predict, an intervention that prompts them to consider reasons why their expenses might be different than usual should increase prediction accuracy by making atypical expenses more accessible.

### 2.5 Theoretical Model

Taken together, the propositions outlined above lead to the model illustrated in Figure 2. First, the model shows that the relative accessibility of prototype attributes leads to "prototypical predictions," which I adopt as shorthand for "predictions based on prototype attributes that approximate the mode of a consumer's expense distribution." Second, the model shows that prototypical prediction leads to an expense prediction bias (EPB) in which consumers under-predict their future expenses. Note that in isolation, the Prototypical Prediction → EPB pathway assumes the general case of positively skewed expenses. Similarly, the Accessibility → Prototypical Prediction → EPB pathway indicates that increased accessibility of atypical expenses will reduce the tendency to make prototypical predictions, and therefore improve prediction accuracy when expenses are positively skewed. Finally, the Distributional Skew → Prototypical Prediction → EPB pathway gives the model the flexibility it needs to acknowledge that both the statistical nature of prototypical predictions (i.e., the mode) and the relative prediction accuracy they produce will vary with the degree of positive skew. The logic here is plain: If prototypical predictions approximate mode expenses, then all else equal predictions should be more accurate when expenses approach a normal distribution with mode = mean versus when they are positively skewed with mode < mean.

**Figure 2**

**Theoretical Model**



## 2.6 Hypotheses

The central argument of the prototype theory is that consumers make prototypical expense predictions that resemble the mode of their expense distribution. A different way of saying this is that expense predictions largely neglect the atypical expenses represented in the long right tail of a consumer's expense distribution. In contrast, retrospection is more deeply grounded in reality (Johnson and Raye 1981; Kane, Van Boven, and McGraw 2012; Van Boven, Kane, and McGraw 2008), which implies that expense *recall* will include both typical and atypical expenses. Thus, one way to test the veracity of the prototype theory is to compare the *perceived typicality* of recalled and predicted expenses. If the prototype theory is correct and consumers do indeed engage in prototypical prediction, then it should be the case that they perceive their predicted future expenses to be more typical than their recalled past expenses. On the other hand, if under-prediction occurs predominantly because consumers overlook their typical expenses, then perceived typicality of past and future expenses should not differ. Using weekly expenses as a conservative unit of analysis that allows consumers to recall their past expenses with a fresh memory and predict their future expenses with minimal uncertainty, I hypothesize that:

> **H1:** On average, consumers perceive their predicted expenses for the next week to be more typical than their recalled expenses for the past week.

The prototype theory also implies that the more heavily a consumer relies on prototype attributes when predicting his or her expenses, the lower their expense prediction will be. This follows from the general case that expenses are positively skewed: If prototype attributes represent mode expenses, then a more prototypical prediction will generally be a lower prediction, because the mode of a positively skewed distribution is lower than most other points on the distribution. In contrast, if prototype attributes represent mean expenses, then a more prototypical prediction will generally be a higher prediction, because the mean of a positively skewed distribution is higher than most other points on the distribution. Therefore, a second descriptive test of the prototype theory is to use perceived typicality of future expenses as a measure of prototypical prediction and hypothesize:

> **H2:** Perceived typicality of future expenses is negatively correlated with expense predictions.

The third hypothesis of the present research is that consumers display an expense prediction bias in which they under-predict their future spending. Note that under-prediction can be defined in two ways. First, it can mean predicting that expenses will be lower in the future than they were in the past. This follows logically from H1 and H2: If consumers believe their future expenses will be more typical than their past expenses, and more typical expenses are lower expenses, it should be the case that predicted expenses are lower than recalled expenses. Second, under-prediction can mean predicting lower expenses for a given time period (e.g., the next week or month) as compared to the actual expenses a consumer ends up incurring during that time period. This follows directly from the prototype proposition and the observation that the distribution of expenses is positively skewed: If predictions are based on the mode then most consumers will under-predict to at least some extent. Formally, I hypothesize that:

> **H3a:** Consumers predict lower expenses for the future as compared to the expenses they recall for the past.
>
> **H3b:** Consumers under-predict their expenses for the future as compared to the expenses they actually incur during the target time period (e.g., the next week or month).

The fourth hypothesis of the present research is that making atypical expenses more cognitively accessible when consumers are constructing their predictions will reduce their tendency to make prototypical predictions and therefore increase their increase prediction accuracy. I reasoned that having people list three reasons why their expenses might be *different* than a typical week or month would serve as a simple intervention that accomplishes this goal. The logic underlying this “atypical intervention” follows that of H2 above—most of a positively skewed distribution of expenses that is different than the mode is also higher than the mode, and “different” should therefore make predictions less prototypical and closer to the mean. Using the number of atypical expenses that a consumer is able to bring to mind as a measure of prototypical prediction, I hypothesize that:

**H4a**: Prompting consumers to consider three reasons why their expenses might be different from a typical week increases expense prediction accuracy.

**H4b**: The effect of the atypical intervention on predictions (vs. control) is mediated by the number of atypical expenses that come to mind.

Finally, if consumers make prototypical predictions based on mode expenses, then all else equal, a higher mode should mean a higher and more accurate prediction. This leads to the expectation that predictions will be more accurate when the distribution of expenses is more normally distributed (with mode = mean) versus when it is positively skewed (with mode < mean). Moreover, if the prototype theory is correct, then the effect of skew on prediction accuracy should be mediated by what consumers perceive to be their average expenses. I therefore hypothesize:

**H5a**: Expense predictions will be more accurate when the distribution of expenses is more normally distributed versus more positively skewed.

**H5b**: The effect of skew on prediction accuracy is mediated by what consumers perceive to be their average expenses.

## 3. Studies

To test the hypotheses introduced above I conducted a multi-method series of ten studies. Just as the think aloud pilot study provided descriptive evidence of the link between accessibility and prototypical predictions, Study 1 (n = 485) provides descriptive evidence of the link between prototypical predictions and the expense prediction bias by testing H1–H3a. Study 2 directly replicates Study 1 with a nationally representative sample of Americans consumers (n = 1,048), and examines the Accessibility → Prototypical Prediction → EPB path by using the atypical intervention to test H4a and H4b. Study 3 (n = 187) tests H1–H4a in a longitudinal field study conducted with members of Canada's largest community credit union. Study 4 (n = 601) extends Study 3 by directly comparing the effect of the atypical intervention (H4a) on weekly versus monthly expense predictions. Study 5 (n = 593) extends the Accessibility → Prototypical Prediction → EPB model tested in Studies 1–4 by examining the effect of the atypical intervention on saving intentions. Study 6 (n = 6,224) uses budgeting and spending app data provided by a popular financial aggregation app in the UK to test H5a in a descriptive manner. Study 7 (n = 401) utilizes an experimental paradigm to manipulate skew and test H5a in a causal manner. Study 8 (n = 351) replicates and extends Study 7 by testing H5a and H5b. Finally, Studies 9a (n = 27) and 9b (n = 134) test the additional hypothesis that there exists an *income prediction bias* in which consumers who face variable income *over*-predict their future earnings.

As highlighted in the theoretical development section, the distribution of consumer expenses generally displays a strong positive skew, which presents serious challenges for inferential analysis based on Normal theory statistics. To address this situation I exclude the data of outlier participants whose reported expenses exceed their predictions by more than a factor of 10 (or vice versa), and LN-transform the distributions of reported and predicted expenses. For ease of interpretation, I then exponentiate the descriptive results so they can be easily understood as dollar values. Notably, this procedure ameliorates concerns related to homogeneity of variance assumption violations and the influence of outliers, but it does not change the pattern of results observed in the raw data. All data exclusions are reported in the main text. Studies 2, 4, 5, and 7 were pre-registered. Those pre-registration documents are available along with the data and syntax for studies 1, 2, 4, 5, 7, 8, 9a and 9b here: **https://tinyurl.com/t737pvo**

## 3.1 Study 1: Are Predicted Expenses More Typical Than Recalled Expenses?

The first goal of Study 1 is to test the prototype theory by comparing perceived typicality of predicted future expenses and recalled past expenses. Although memory is not perfect, retrospection does approximate reality (Kane, Van Boven, and McGraw 2012). This leads to the expectation that expense *recall* will reflect the full distribution of expenses, including atypical expenses. However, the prototype theory leads to the expectation that expense *prediction* will overweight typical expenses and neglect atypical expenses. Therefore, Study 1 tests the hypothesis that consumers perceive their predicted future expenses to be more typical than their recalled past expenses (H1).

The second goal of Study 1 is to examine the relationship between perceived typicality of future expenses and expense predictions. If typical expenses represent the mode of a consumer's expense distribution, then it follows that higher perceived typicality of future expenses will be associated with lower predictions, because the mode of a positively skewed distribution is lower than most other points on distribution. Therefore, Study 1 tests the hypothesis that perceived typicality of future expenses is negatively correlated with expense predictions (H2).

The third goal of Study 1 is to test the hypothesis that consumers predict lower expenses for the next week as compared to the expenses they recall for the past week (H3a), which follows logically from the two preceding hypotheses.

**Method**

*Participants*. Four hundred and ninety-nine US residents were recruited via Amazon Mechanical Turk to participate in a short consumer expense survey ($M_{\text{age}}$ = 33.5; 41.3% female). The reported expenses of 14 participants exceeded their predictions by more than a factor of 10 (or vice versa), leaving an effective sample size of 485 ($M_{\text{age}}$ = 33.7; 41.6% female).

*Procedure*. On the first page of the study participants were asked to report their expenses for the past week. Specifically, they received the following instructions:

> Please take some time to estimate your expenses for the past week (i.e., the past 7 days).
>
> Please enter your total estimated expenses (in dollars) for the past week. Your estimate should account for all the expenses you incurred except monthly expenses like rent that happen to be due in the past week.

On the second page of the study participants received the same instructions with respect to predicting their expenses for the next week. Weekly expenses were chosen as the unit of analysis in this study for two reasons. First, it allows participants to recall their past expenses while their memories are fresh, and predict their future expenses with minimal uncertainty. This should make weekly expenses a conservative test of H1. Second, pre-test survey data ($n$ = 1,514) collected from the same participant pool suggests that consumers in this sample most commonly think about their expenses in weekly terms.[3] Participants were asked to exclude monthly expenses like rent from their estimates to reduce the possibility that any observed difference between recall and prediction could be due to variation in the timing of these expenses.

On the third page of the study participants were asked to indicate how typical they perceived their recalled expenses to be using the following measure:

> "How *different* or *similar* do you think your expenses were for the **past** week, relative to a typical week?" (1 = Very different; 7 = Very similar).

On the fourth page of the study participants responded to the same measure modified to ask about their predicted expenses. Finally, participants were asked to report basic demographic information on the fifth and final page of the study.

**Results**

*Perceived Typicality*. Supporting H1, participants predicted their future expenses would be more typical than their past expenses, ($M_{pastweek}$ = 4.71, 95% $CI_{pastweek}$ = [4.56, 4.86], $M_{nextweek}$ = 5.02, 95% $CI_{nextweek}$ = [4.89, 5.16]), as confirmed by a within-subject t-test ($t(484) = 4.22$, $p < .001$, $d = .19$). Supporting H2, correlational analysis showed that higher perceived typicality of future expenses was associated with lower expense predictions ($r(483) = -.17$, $p < .001$). As a robustness check for the H2 result, I also regressed predicted expenses onto perceived typicality of future expenses and participant income to examine the possibility that the correlation between predictions and perceived typicality was due to wealthier participants with higher expenses also having more atypical expenses. Consistent with H2, the negative relationship between predicted

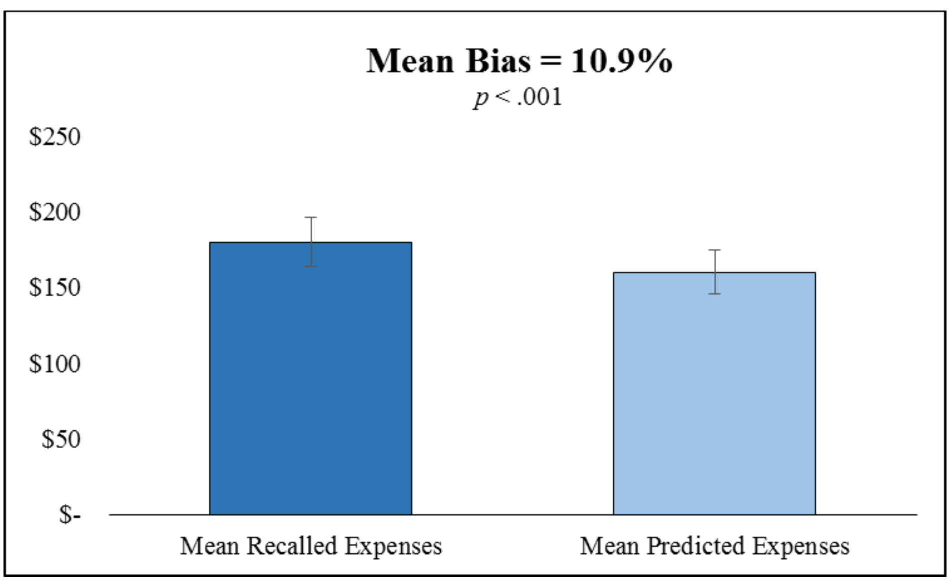
[3] 38.6% of survey respondents indicated thinking about their finances on a weekly basis, 27.1% indicated thinking about their finances on a biweekly basis, 28.5% indicated thinking about their finances on a monthly basis, and 5.7% indicated "other.

expenses and typicality remained significant after controlling for participant income ($B$ = -.16, $t(482) = -3.88$, $p < .001$, $R^2 = .13$).

*Expense Prediction Bias (Recalled – Predicted Expenses)*. Supporting H3a, predicted expenses for the next week were 10.90% lower than reported expenses for the past week ($M_{pastweek}$ = \$179.33, 95% $CI_{pastweek}$ = [\$163.76, \$196.37], $M_{nextweek}$ = \$159.78, 95% $CI_{nextweek}$ = [\$145.91, \$174.95]), as confirmed by a within-subject t-test ($t(484) = 3.90$, $p < .001$). With respect to the prevalence of the bias, 18.6% ($n = 90$) of the 485 participants in this study predicted that their expenses for the next week would be the same as their expenses for the past week. Of the remaining 395 participants, a significant majority (59.5%) predicted that their future expenses would be lower than their past expenses (95% CI = [54.48%, 64.38%], $z = 3.78$, $p < .001$).

**Figure 3**

**Mean Recalled and Predicted Weekly Expenses in Study 1**
Error Bars Represent 95% Confidence Intervals



**Discussion**

The results of Study 1 paint a picture of consumer expense prediction that supports the prototype theory in three ways: Consumers predict their next week's expenses will be more typical (H1) and lower (H3a) than their past week's expenses, and perceived typicality of future expenses is negatively correlated with expense predictions (H2). Study 1 also provides evidence that the H2 result was not merely driven by high income participants with wildly atypical expenses. Taken together, these results provide descriptive support for the Prototypical Prediction → EPB path in the theoretical model (Figure 2).

The results of Study 1 also suggest that the expense prediction bias is prevalent, and that its magnitude (10.9%) may be economically significant for many consumers, especially in light of the fact that the paradigm used in this study included only weekly expenses, and not larger, monthly expenses. Study 2 offers the opportunity to directly replicate these results in a nationally representative sample of Americans and it introduces the atypical intervention.

## 3.2 Study 2: Does the "Atypical" Intervention Increase Predictions?

The primary purpose of Study 2 is to test the effectiveness of the atypical intervention and provide insight into how it works. If the prototype theory is correct and expense predictions are prototypical, then prompting consumers to consider reasons why their expenses might be different than usual should increase predictions (H4a) by making atypical expenses more accessible (H4b). Therefore, if the intervention works as hypothesized, it will not only offer a simple way to improve consumers' expense predictions, it will provide further support for the prototype theory. The control condition of Study 2 also provides an opportunity to directly replicate the results of Study 1.

**Method**

*Participants.* A nationally representative sample of 1,091 adult Americans completed Study 2 via Time-Sharing Experiments for the Social Sciences. The recalled expenses of 43 participants exceeded their predictions by more than a factor of 10 (or vice versa), leaving an effective sample size of 1,048 ($M_{age}$ = 49.6; 53.0% female; 72.8% Caucasian, 9.4% Black, 10.7% Hispanic, 7.2% Other; Mode level of education = Bachelor's degree; Median household annual income = $50–59,999).

*Procedure.* Participants were randomly assigned to one of three conditions: control, typical, or atypical. In the control condition, they recalled and predicted their weekly expenses for the past and next week as in Study 1. Participants in the typical condition also recalled and predicted their weekly expenses, but they received the following instructions before making their prediction: "Now consider why your expenses for next week might be *similar* to that of any other week. In the spaces provided below, please type 3 reasons why your expenses for next week might be *similar* to that of any other week." My expectation was that this would *not* significantly impact perceived typicality of future expenses or expense prediction amount (vs. control), because if the prototype proposition is correct, then predictions in the control condition should already be based primarily on typical expenses. The atypical intervention condition paralleled the typical condition but instructed: "Now consider why your expenses for next week might be *different* from that of any other week. In the spaces provided below, please type 3 reasons why your expenses for next week might be *different* from that of any other week." My preregistered hypothesis was that this would increase expense predictions versus control (H4a). The order of prediction and recall was counterbalanced in all conditions.[4]

Participants were next presented with an atypical expense-listing task that asked, "Is there anything you believe you will spend money on in the next week that you did NOT spend money on during the past week?" and "Is there anything that you spent money on during the past week that you believe you will NOT spend money on in the next week?" Participants were then given the opportunity to list a description and corresponding dollar amount for up to five such expenses. My principal expectation was that the atypical intervention would make atypical expenses more accessible during prediction, which would result in a higher number of expenses being listed in the atypical condition as compared to the control and typical conditions. Furthermore, I expected that the number of atypical expenses listed for the next week would mediate the relationship between experimental condition and predicted expenses (H4b), because a higher number of atypical expenses listed would indicate a reduction in prototypical prediction.

Finally, participants completed the same measures of perceived typicality used in Study 1, and five measures designed to explore the relationship between the expense prediction bias and variables such as financial slack (Berman et al. 2016; Zauberman and Lynch 2005) and available resources. These exploratory measures yielded null results that I discuss in Appendix A.

**Results**

*Replicating Study 1*. The results observed in Study 1 were directly replicated in the control condition of Study 2. Supporting H1, participants in this nationally representative sample

[4] A set of 2 (order: predict first vs. recall first) × 3 (condition: control vs. typical vs. atypical) ANOVAs with predicted expenses, recalled expenses, and bias scores (recalled – predicted expenses) as the DVs revealed no order effect ($p$'s > .17).

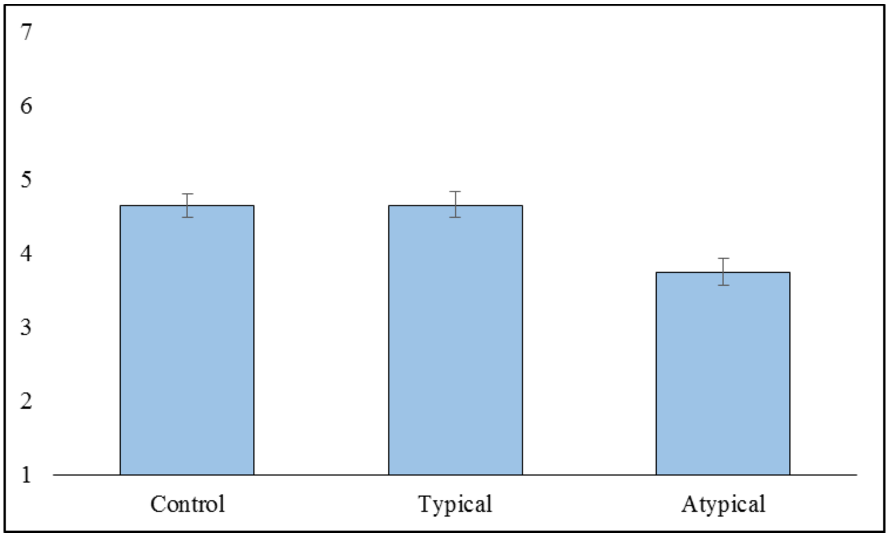
predicted their future expenses would be more typical than their past expenses ($M_{pastweek}$ = 4.40, 95% $CI_{pastweek}$ = [4.23, 4.57], $M_{nextweek}$ = 4.65, 95% $CI_{nextweek}$ = [4.48, 4.81]), as revealed by a within-subject t-test ($t(415) = -3.42$, $p = .001$, $d = .17$). Supporting H2, higher perceived typicality of future expenses was once again associated with lower expense predictions ($r(414) = -.21$, $p < .001$), and this association remained significant after controlling for participant income ($B = -.23$, $t(413) = -4.97$, $p < .001$, $R^2 = .10$). Supporting H3a, participants predicted their future expenses would be 9.0% lower than their past expenses ($M_{pastweek}$ = \$236.77, 95% $CI_{pastweek}$ = [\$214.52, \$261.36], $M_{nextweek}$ = \$215.47, 95% $CI_{nextweek}$ = [\$195.00, \$238.05]), as revealed by a within-subject t-test ($t(415) = 2.76$, $p = .006$). Furthermore, of the 294 participants in the control condition who predicted that their expenses would be either lower or higher in the next week as compared to the past week, a significant majority (57.1%) predicted lower expenses (95% CI = [51.22%, 62.83%], $z = 2.44$, , $p = .015$). I next test for differences in perceived typicality and EPB across all three conditions.

*Perceived Typicality of Future Expenses*. A one-way ANOVA with prediction condition (control vs. typical vs. atypical) as the independent variable and perceived typicality of future expenses as the dependent variable revealed a significant main effect of prediction condition ($F(2, 1044) = 32.27$, $p < .001$, partial eta squared = .06). Planned contrasts further revealed that perceived typicality was virtually identical in the control and typical conditions ($M_{control}$ = 4.65, 95% $CI_{control}$ = [4.48, 4.81], $M_{typical}$ = 4.65, 95% $CI_{typical}$ = [4.48, 4.83], $t(1044) = -.03$, $p = .97$), which is consistent with the prototype theory and may indicate that predictions in the control condition were based on typical expenses. Furthermore, perceived typicality in the atypical intervention condition was significantly lower than in the control and typical conditions ($M_{atypical}$

= 3.74, 95% $CI_{atypical}$ = [3.56, 3.93], $t(1044) = -8.02$, $p < .001$, $d = .55$), which supports the prototype theory by providing preliminary evidence that the intervention works as intended.

**Figure 4**

**Mean Perceived Typicality of Future Expenses in Each Condition of Study 2**
(1 = Very Different from a Typical Week; 7 = Very Similar to a Typical Week)
Error Bars Represent 95% Confidence Intervals




*Expense Prediction Bias (Recalled – Predicted Expenses)*. Supporting H3a, predicted expenses were 9.0% lower than recalled expenses in the control condition ($t(415) = 2.76$, $p = .006$), and 6.4% lower than recalled expenses in the typical condition ($t(331) = 2.07$, $p = .039$). However, supporting H4a, predicted expenses did not differ significantly from recalled expenses in the atypical intervention condition ($t(299) = -1.49$, $p = .14$). In other words, the expense prediction bias was neutralized (and even slightly reversed) by the atypical intervention. A 3 (condition: control vs. typical vs. atypical) × 2 (expenses: past week vs. next week) mixed-model ANOVA with condition as a between-subjects variable and expenses as a within-subject variable confirmed a significant main effect of condition ($F(2, 1045) = 4.61$, $p = .010$, partial eta squared = .01), no main effect of time period ($F(1, 1046) = 1.67$, $p = .20$), and a significant condition-by-expenses interaction ($F(2, 1045) = 5.21$, $p = .006$, partial eta squared = .01). Planned contrasts confirmed that predicted expenses in the atypical condition ($M_{\text{atypical}}$ = \$271.16, 95% $\text{CI}_{\text{atypical}}$ = [239.15, 307.42]) were 25.8% higher than in the control condition ($M_{\text{control}}$ = \$215.47, 95% $\text{CI}_{\text{control}}$ = [195.00, 238.05]; $t(1045) = 2.90$, $p = .004$), and 35.7% higher than in the typical condition ($M_{\text{typical}}$ = \$199.80, 95% $\text{CI}_{\text{typical}}$ = [179.18, 222.78]; $t(1045) = 3.66$, $p < .001$), but that predictions did not differ between the control and typical conditions ($t(1045) = .98$, $p = .33$). Planned contrasts also revealed that recalled expenses did not differ between the atypical condition ($M_{\text{atypical}}$ = \$251.84, 95% $\text{CI}_{\text{atypical}}$ = [222.49, 285.09]) and control condition ($M_{\text{control}}$ = \$236.77, 95% $\text{CI}_{\text{control}}$ = [214.52, 261.36]; $t(1045) = .78$, $p = .43$), nor did they differ between the typical condition ($M_{\text{typical}}$ = \$213.43, 95% $\text{CI}_{\text{typical}}$ = [191.41, 237.98]) and control condition ($t(1045) = 1.36$, $p = .18$). Recalled expenses were 15.3% lower in the typical condition than in the atypical condition ($t(1045) = 2.00$, $p = .046$), but this makes the test of EPB between these two conditions notably conservative, because lower (higher) recalled expenses decreases (increases) the size of the bias.[5] However, despite lower recalled expenses in the typical condition and higher recalled expenses in the atypical condition, a significant bias was observed in the former but not in the latter.

*Atypical Expense-Listing Task*. A one-way ANOVA with condition (control vs. typical vs. atypical) as the independent variable and number of atypical expenses listed for the next week as the dependent variable revealed a significant main effect of condition ($F(2, 1045) = 5.85$, $p = .003$, partial eta squared = .01). Planned contrasts further revealed that the number of expenses listed in the atypical condition ($M_{\text{atypical}} = 1.65$, $\text{SD}_{\text{atypical}} = 1.62$) was significantly higher than in

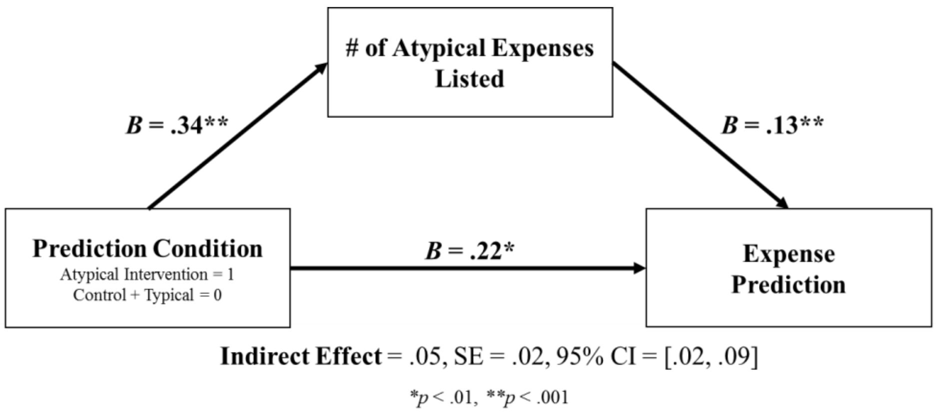
[5] As noted in the preceding footnote, the order of prediction and recall does not interact with condition, but adding order to the model does reveal directionally lower recall for participants in the typical condition who predicted first, and directionally higher recall for participants in the atypical condition who predicted first. Therefore, I believe the difference in recalled expenses between these two conditions may be the result of the prediction manipulation in each condition spilling over (somewhat) into recall.

the control condition ($M_{control}$ = 1.25, $SD_{control}$ = 1.46, $t(1045) = 3.41$, $p = .001$, $d = .26$) and typical condition ($M_{typical}$ = 1.39, $SD_{typical}$ = 1.59; $t(1045) = 3.13$, $p = .002$, $d = .16$). This supports H4b by providing evidence that the intervention makes atypical expenses more accessible during prediction. However, the number of expenses listed in the control and typical conditions did not differ ($t(1045) = 1.19$, $p = .24$). This supports the prototype theory by further suggesting that predictions in the control condition may have been based on typical expenses. Finally, a one-way ANOVA with condition (control vs. typical vs. atypical) as the independent variable and average dollar amount of atypical expenses as the dependent variable revealed no effect of condition ($F(2, 548) = .78$, $p = .46$). This indicates that although the intervention makes *more* atypical expenses accessible, it does not make *higher* atypical expenses accessible.

*Mediation Analysis.* The results above confirm that the atypical intervention lived up to its name by making atypical expenses more accessible during prediction, and that expense predictions were significantly higher in the atypical condition as well. To further investigate the relationship between the number of atypical expenses that participants listed in each condition and their expense predictions, I next tested a mediation model with condition (atypical = 1 vs. control and typical = 0) as the independent variable, expense prediction as the dependent variable, and the number of atypical expenses listed as the mediating variable. The indirect effect of condition on expense prediction via number of atypical expenses was significant (indirect effect = .05, SE = .02, 95% CI = [.02, .09]). Specifically, the model confirms that the atypical intervention succeeded in increasing the number of atypical expenses listed ($B = .34$, 95% CI = [.13, .55]; $t(1046) = 3.21$, $p < .001$), and that this number was associated with higher expense predictions, even while controlling for condition ($B$ = .13, 95% CI = [.09, .17]; $t(1045) = 6.28$, $p < .001$), supporting H5b.[6]

**Figure 5**

**Study 2 Mediation Model**




## Discussion

Study 2 provides support for the prototype theory in three ways. First, the control condition of Study 2 directly replicates the findings of Study 1 in a nationally representative sample, demonstrating that consumers predict their future expenses will be both more typical (H1) and lower than their past expenses (H3), and that perceived typicality of future expenses is negatively correlated with predictions (H2). Second, it demonstrates that the intervention can increase predicted expenses to the level of recalled expenses by making atypical expenses more accessible when consumers construct their predictions (H4a and H4b). This not only supports the prototype theory, it also offers a practical tool that can be easily applied to help consumers improve their prediction accuracy. Finally, while being careful not to over-emphasize the implications of a null result, the fact that perceived typicality and expense predictions do not

[6] The same results are obtained when using only the atypical and pure control conditions as levels of the IV (indirect effect = .05, SE = .02, 95% CI = [.02, .09]) and when running a categorical mediation model that includes all three conditions (indirect effect of atypical dummy = .05, SE = .02, 95% = [.02, .08]; indirect effect of typical dummy = .02, SE = .02, 95% CI = [-.01, .05]).

differ between the control and typical conditions supports the prototype theory in that it is consistent with the suggestion that predictions in the control were already based on typical expenses.

The principal limitation of Studies 1 and 2 is that the operationalization of EPB in these studies (recalled – predicted expenses) does not reveal true prediction accuracy. To address this, Study 3 operationalizes EPB as the difference between participants' predicted expenses and the expenses they actually incur during the target week or month, then measures prediction accuracy within-subject using a repeated-measures longitudinal design. Study 3 also allows the atypical intervention to be tested in the field.

### 3.3 Study 3: A Longitudinal Field Study of the Expense Prediction Bias

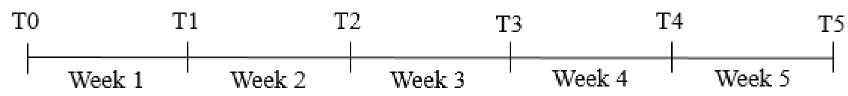
The first goal of Study 3 was to gain a more comprehensive understanding of expense prediction bias as a phenomenon by observing its magnitude and persistence in a longitudinal field conducted with a highly engaged group of participants. The second goal of Study 3 was to examine the effectiveness of the atypical intervention (H4a) in the field. To accomplish these goals I partnered with Canada's largest community credit union to run a five-week longitudinal field study with its members.

#### Method

*Participants and Procedure.* Participants for this study were members of a midsized financial cooperative (~500,000 members) recruited through an online panel of members (~5,000) that the cooperative maintains in order to conduct market research. I targeted a sample size of 200 based on the effect sizes observed in pilot studies. Each participant completed six surveys over the course of five weeks, as illustrated by each time period marked in Figure 6.

**Figure 6**

**Data Collection Schedule for Study 3**




Because I had no prior experience sampling from this population, data collection took place in two waves. In wave 1, I sent a survey at time zero (T0) to 400 randomly selected members at noon on Sunday, September $10^{th}$, 2017. Ninety-three people completed the survey before it was deactivated at 11:59pm on Monday, September $11^{th}$, 2017. I then monitored attrition for two weeks before calculating that the T0 survey should be sent to another 800 randomly selected members (the maximum number allowed by the Credit Union) in the second wave of data collection so that I could recruit as close to 200 total participants as possible. In wave 2, two hundred and nineteen members completed the T0 survey. At the end of both waves of data collection I had complete data from 187 participants (61 from wave 1 and 126 from wave 2, $M_{age}$ = 51.12, 57.8% female). Compensation for each participant included a personalized spending report (provided at the end of the study) that served as an incentive to predict and report expenses as accurately as possible. Participants also received a $10 Amazon gift certificate for each completed survey.

**Table 4**

**Observable Characteristics of Participants Who Completed vs. Dropped Out of Study 3**

| | **Participants who Completed Study** | **Participants who Dropped Out** | ***t*-test** |
|---|---|---|---|
| N | 187 | 126 | |
| Budgeter | 23.53% | 26.12% | $p = .27$ |
| Hours spent financial planning each week | 1.83 | 1.87 | $p = .83$ |
| Age | 51.02 | 48.87 | $p = .20$ |
| Female | 55.6% | 58.7% | $p = .29$ |
| Income | $50-60K | $60-70K | $p = .02$ |
| T0 weekly prediction | $474.14 | $441.82 | $p = .59$ |
| T0 monthly prediction | $2224.97 | $2377.50 | $p = .48$ |

All surveys were emailed to participants at noon on a Sunday and required completion before 11:59 pm the next day. The first survey asked participants to predict their expenses for the next week as follows: "Please take some time to estimate your total expenses for the **next week**. By "total expenses" we mean everything you will pay for during the next week. (Page Break). Please enter your estimated total expenses for the **next week**." Participants were then asked to indicate how typical they expected their expenses to be ("How different or similar do you think your expenses will be for the next week [month], relative to a typical week [month]?" 1 = very different, 7 = very similar), and how confident they were in their prediction accuracy ("How sure or confident are you that your estimate of your total expenses for the next week is accurate?" 1 = very unsure, 7 = very sure). Participants then answered the same questions with respect to their monthly expenses.

The remaining five surveys began by asking participants to log into their online bank account and report their expenses for the past week, then predict their expenses for the next week. Both expense reports and predictions were followed by the same measures of perceived typicality and confidence used in survey 1. In the second-to-last survey (i.e., at T4), half of the sample was randomly assigned to receive the atypical intervention, making the final week of the study a 2 (condition: control vs. atypical) × 2 (expenses: predicted vs. actual) between-within design.[7]

Across the six surveys used to measure prediction accuracy in this study I also measured a number of theoretically relevant individual differences that could be associated with expense prediction accuracy. The first was the presence of a savings goal, because the motivation to save could produce lower spending predictions but not lower spending behavior (Peetz and Buehler 2009). The second was trait optimism (Scheier, Carver, and Bridges, 1994). I predicted that trait optimism would actually *not* be correlated with EPB because research on the planning fallacy has demonstrated that optimistic predictions in that domain are not the result of an optimistic disposition (Buehler, Griffin, and Ross, 1994). Nonetheless, because the relationship between trait optimism and expense predictions seems intuitively compelling and has not yet been explored, I felt there was value in testing it.

The third individual difference I measured was short-term financial propensity to plan (PTP; Lynch et al., 2010). One prediction regarding the relationship between PTP and EPB is that consumers with a higher PTP will make more accurate predictions because they are more attuned to (or concerned with) future outcomes. However, it could be the case that greater plan-focus leads consumers to be less attentive to unplanned expenses and therefore more likely to under-predict their future expenses (cf. Buehler, Griffin, and Peetz, 2010). I therefore included the Lynch et al. (2010) measure of short-term financial PTP so that I could explore these competing hypotheses.

[7] Participant dropout over the last week of the study was minimal ($n$ = 4) and did not differ by condition.

The fourth individual difference measure I included was the Rick, Cryder, and Lowenstein (2008) spendthrift-tightwad scale. There are three hypotheses for how this measure could be correlated with EPB. The first is that tightwads display *higher* EPB because anticipatory pain of paying causes them to predict even lower expenses than they actually incur. The second hypothesis is that tightwads display *lower* EPB if they are more sensitive to expenses and therefore more accurate. Finally, spendthrifts may display higher EPB because they lack pain of paying during purchase and therefore may be more likely to overspend vs. prediction.

The fifth individual difference measure I included was numeracy (Schwartz et al., 1997) because consumers who are unable to perform the mental calculations required to make an accurate expense prediction may display higher EPB. The sixth measure was linear vs. cyclical time orientation (adapted from Tam and Dholakia, 2014). Consumers with a stronger cyclical orientation may display lower EPB because they see life as a series of recurring events (Tam and Dholakia, 2014), and therefore be more easily able to incorporate past atypical expenses into their predictions for the future. The seventh individual difference I included was openness to experience (John, Donahue, and Kentle, 1991), because being more likely to consider a wider range of outcomes when making predictions could be associated with lower bias. Finally, I also measured temporal discounting for both losses and gains (Kirby & Maraković 1996). Consumers with a relatively higher discount rate for losses (i.e., those with a stronger preference to pay more later vs. less now) may display higher EPB because they may want to postpone payment as much as possible, and our measure of EPB in this study was for the coming the week. The same logic applied in reverse led me to belief that temporal discounting of gains may be negatively correlated with EPB. Each individual difference measure is included in its entirety in Appendix B.

**Results**

*Perceived Typicality*. To test the hypothesis that consumers predict their expenses will be more typical in the future than in the past (H1), I compared reported and predicted expense typicality at T1, T2, T3, and T4. In other words, I tested whether or not participants predicted their expenses would be more typical in week 2 than week 1, week 3 than week 2, and so on. As illustrated in Figure 7 and Table 5, participants predicted their expenses would be more typical in the next (vs. past) week at all four points in time. Figure 7 also shows that the atypical intervention succeeded in neutralizing this tendency at T4. In sum, these results provide strong support for H1.

To test the hypothesis that perceived typicality of future expenses is negatively correlated with expense predictions (H2), I analyzed the Pearson correlation between perceived typicality of future expenses and weekly expense predictions for each week of the study, as well as for the month. Perceived typicality of future expenses was negatively correlated with weekly expense predictions at T0 ($r(185) = -.30$, $p < .001$), T2 ($r(185) = -.28$, $p < .001$), and T4 ($r(185) = -.25$, $p = .001$). The correlation at T1 was marginally significant ($r(185) = -.12$, $p = .09$), as was the correlation between perceived typicality of monthly expenses and expense predictions for the month ($r(185) = -.12$, $p = .09$). The correlation at T3 was directionally consistent, though not significant ($r(185) = -.03$, $p = .74$). I next regressed predictions onto perceived typicality of future expenses and participant income to rule out the possibility that the correlation between predictions and typicality was due to wealthier participants with higher expenses also having more atypical expenses. This analysis produced only one meaningful change in these results: The relationship between perceived typicality and prediction at T1 became significant ($p = .013$). In aggregate, these findings provide support for H2.

**Figure 7**

**Mean Reported (Past) and Predicted (Future) Typicality for Each Week of Study 3**

(1 = Very Different from a Typical Week; 7 = Very Similar to a Typical Week)

Error Bars Represent 95% Confidence Intervals

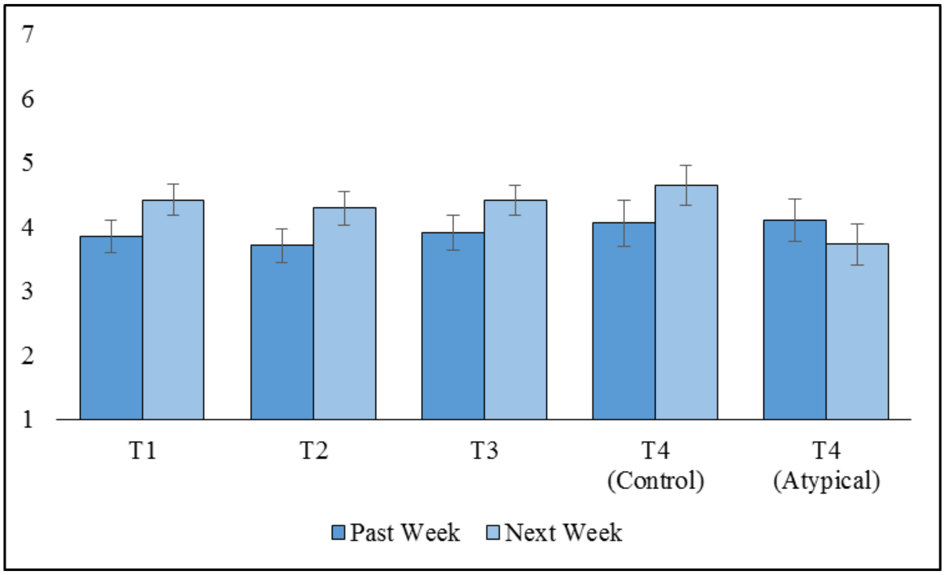



**Table 5**

**T-tests Comparing Reported and Predicted Typicality for Each Week of Study 3**

| | T1 | T2 | T3 | T4 (Control) | T4 (Atypical) |
|---|---|---|---|---|---|
| *t* | 3.45 | 4.10 | 3.25 | 2.88 | -1.78 |
| *df* | 174 | 173 | 177 | 89 | 88 |
| *p* | 0.001 | < 0.001 | 0.001 | 0.005 | 0.079 |
| *d* | 0.26 | 0.31 | 0.24 | 0.30 | 0.19 |

*Expense Prediction Bias (Recalled – Predicted Expenses).* To test the hypothesis that consumers tend to predict their future expenses will be lower than their past expenses (H3a), I compared reported expenses against predicted expenses at T1, T2, T3, and T4. That is, I tested whether participants predicted their expenses would be lower in week 2 than week 1, week 3 than week 2, and so on. As illustrated in Figure 8 and Table 6, mean predicted expenses were significantly lower than mean reported expenses at each stage of the study until the atypical intervention was deployed. This longitudinally replicates the findings of studies 1 and 2 and provides strong support for H3a.

**Figure 8**

**Mean Reported (Past) and Predicted (Future) Expenses for Each Week of Study 3**

Error Bars Represent 95% Confidence Intervals

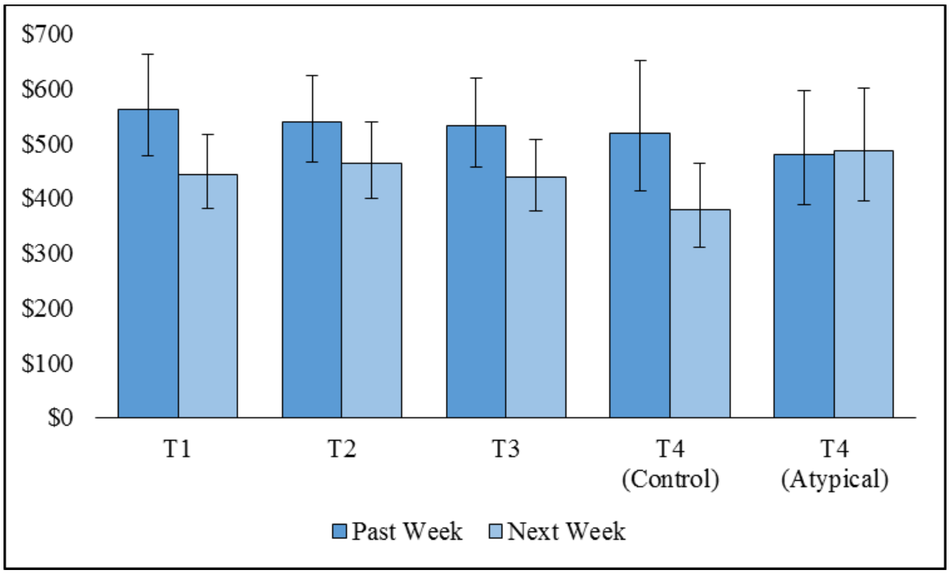



**Table 6**

**T-tests Comparing Reported and Predicted Expenses for Each Week of Study 3**

| | T1 | T2 | T3 | T4 (Control) | T4 (Atypical) |
|---|---|---|---|---|---|
| *t* | 3.45 | 2.40 | 3.25 | 3.88 | -0.16 |
| *df* | 174 | 173 | 177 | 89 | 88 |
| *p* | 0.001 | 0.018 | 0.001 | < .001 | 0.870 |
| *% Diff* | 21.0% | 13.9% | 17.6% | 26.9% | -1.4% |

*Expense Prediction Bias (Actual – Predicted Expenses).* To examine observed prediction accuracy I next compared participants' predicted expenses at the start of each week to the expenses they actually incurred during that week. As illustrated in Figure 9 and Table 7, mean predicted expenses were significantly lower than mean incurred expenses in each and every week of the study, except during week 5 in the atypical condition where the intervention completely neutralized the expense prediction bias. A 2 (condition: control vs. atypical) × 2 (expenses: predicted vs. actual) between-within ANOVA confirmed a significant condition-by-expenses interaction ($F(1, 181) = 5.08$, $p = .025$). Planned contrasts further confirmed that predicted expenses were 36.7% higher in the atypical (vs. control) condition ($F(1, 181) = 4.48$, $p = .036$), and that actual expenses did not differ by condition ($F(1, 181) = .44$, $p = .51$).

In dollar terms, EPB in the control condition was \$79.99 (different from zero, $t(91) = 3.19$, $p = .002$) versus –\$6.65 in the atypical condition (not different from zero, $t(90) = -.20$, $p = .85$). It is also notable that our intervention did not require sacrificing correlational accuracy in the association between predicted and actual expenses ($r_{\text{control}}(90) = .81$, $r_{\text{treatment}}(89) = .80$, z = .19, $p = .85$). These findings provide support for H3b and H4a.

**Figure 9**

**Weekly Expense Prediction Accuracy in Study 3**

Error Bars Represent 95% Confidence Intervals

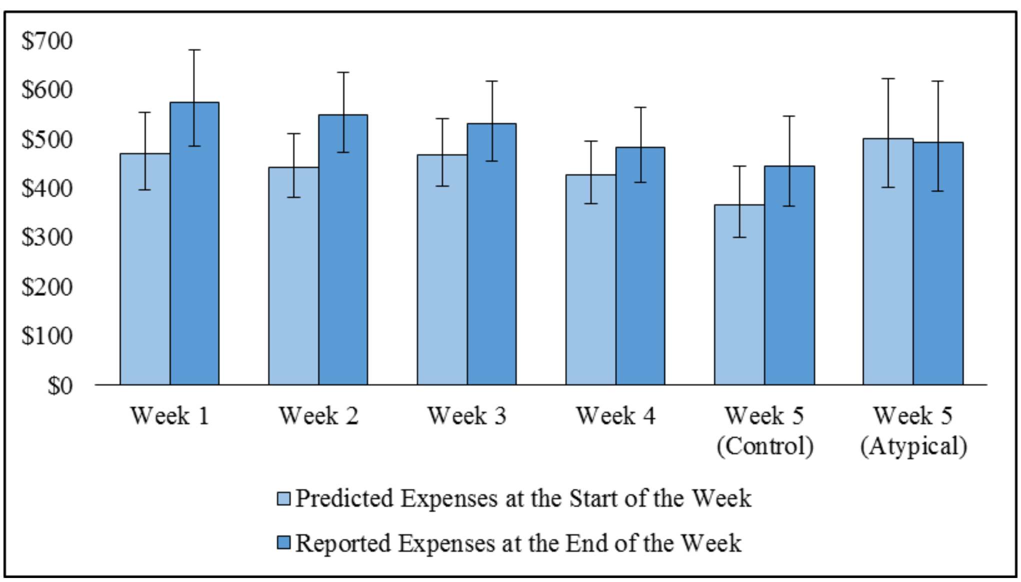


**Table 7**

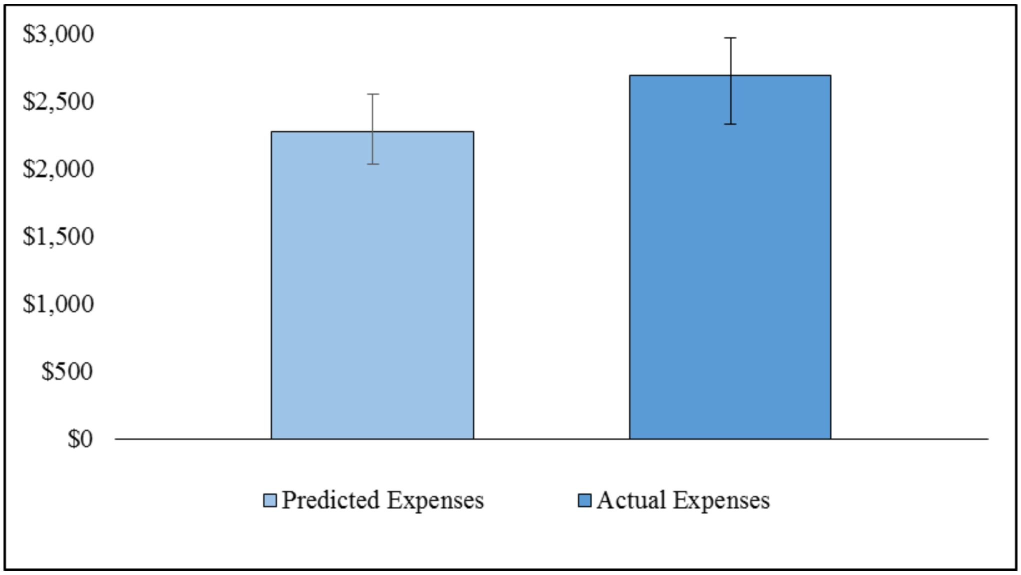
**T-tests for Prediction Accuracy (Predicted vs. Reported Expenses) in Study 3**

| | **Week 1** | **Week 2** | **Week 3** | **Week 4** | **Week 5** (Control) | **Week 5** (Atypical) |
|---|---|---|---|---|---|---|
| *t* | 3.71 | 4.35 | 2.86 | 2.73 | 3.19 | -0.19 |
| *df* | 179 | 179 | 180 | 182 | 91 | 90 |
| *p* | < .001 | < .001 | 0.005 | 0.007 | 0.002 | 0.85 |
| *% Diff* | 18.3% | 19.4% | 11.8% | 11.4% | 18.0% | -1.3% |

*Monthly Expense Prediction Accuracy.* I was also able to test the accuracy of participants' monthly expense predictions by comparing them to the expenses they actually incurred in weeks 1 through 4 of the study. As illustrated in Figure 10, mean predicted expenses for the target month ($M_{\text{pred}}$ = \$2276.74, 95% CI = [2031.64, 2551.15]) were \$416.77 (15.5%) lower than mean expenses actually incurred ($M_{\text{actual}}$ = \$2693.51, 95% CI = [2376.06, 3053.37]; $t(184) = 3.85$, $p < .001$). To put the size of this monthly EPB in perspective, consider that 46% of Americans report they do not have enough money to cover a \$400 emergency expense (US Federal Reserve 2016). To the best of my knowledge this provides the first evidence that consumers under-predict their monthly as well as weekly expenses.

**Figure 10**

**Monthly Expense Prediction Accuracy in Study 3**
**Error Bars Represent 95% Confidence Intervals**




*Intervention content analysis.* To gain more insight into the cognitive process underlying expense prediction I had two research assistants who were blind to the hypotheses independently code the content of participants' responses to the atypical intervention. Specifically, I had them code each response as either a reason expenses would be higher, a reason expenses would be lower, a reason expenses would not actually be different, or a reason that was too ambiguous to be coded as higher, lower or not different. Interrater reliability was 96.1% and disagreements were resolved through discussion between the raters.

The results of this coding exercise are presented in Table 8. The results in the top row ("Overall") are striking because they are highly consistent with a positively skewed distribution of expenses: most reasons given were reasons why expenses would be higher than typical, with a low to moderate percentage of responses representing reasons why expenses would be lower. The results in the third column of the table ("Not Different") are also notable: the percentage of responses indicating that expenses would not be different grew by almost a factor of five between reasons 1 and 3 ($\chi_{(1)}$ = 7.79, $p$ = .005). This finding complements the results of the think aloud study (which showed that prototypical expenses are relatively accessible) by providing evidence that atypical expenses are relatively inaccessible.

**Table 8**

**Summary of Results for the Study 3 Intervention Content Analysis**

| | Higher | Lower | Not Different | Ambiguous |
|---|---|---|---|---|
| **Overall** | 73.5% | 12.9% | 9.3% | 4.3% |
| **Reason 1** | 83.9% | 8.6% | 3.2% | 4.3% |
| **Reason 2** | 68.8% | 15.1% | 9.7% | 6.5% |
| **Reason 3** | 67.7% | 15.1% | 15.1% | 2.2% |
| **Examples** | "Buying some new duds"<br>"I've started a keto diet and some of the extra food might be expensive."<br>"Possible legal bill" | "No commute expenses"<br>"Doggy daycare paid until December"<br>"Recently stocked up on certain grocery items" | "No reason"<br>"N/A"<br>"None" | "Different routine"<br>"Beginning of the month"<br>"Jury duty" |

*Prediction Confidence.* I measured prediction confidence in this study because it has been shown that lower prediction confidence leads people to adjust their expense predictions upward (Ulkumen et al. 2008). Therefore, it could be the case that the atypical intervention is effective "merely" because it is decreasing prediction confidence. To test this possibility I first performed an independent samples t-test with condition as the independent variable and week 5 prediction confidence as the dependent variable. This analysis revealed that the atypical intervention did decrease prediction confidence ($M$ = 4.69, SD = 1.13, $M$ = 5.04, SD = 1.10, t(181) = -2.13, $p$ = .035, $d$ = .31). However, the effect of condition (atypical versus control) on expense predictions remained significant when controlling for prediction confidence ($F$(1, 180) = 4.20, $p$ = .042, partial eta squared = .02), demonstrating that the effect of the atypical intervention is operating above and beyond conditional differences in confidence. Moreover, prediction confidence did not mediate the effect of condition on expense predictions (indirect effect = .01, SE = (.03), 95% CI = [-.05, .06]), indicating that prediction confidence was not directly related to expense predictions in this study. Finally, I also examined the correlation between prediction confidence and expense predictions in week 1–4 of the study and found that the effect sizes were neither substantively meaningful ($r$'s < .07) nor statistically significant ($p$'s > .40).

*Individual Differences.* As summarized in Table 9, none of the individual differences measured in Study 3 were found to be significantly correlated with EPB or perceived typicality of predicted expenses on a consistent, significant, and/or substantive basis.

**Table 9**

**Pearson Correlation for Individual Differences Measured in Study 3 and EPB**

| | Correlation between Individual Difference Measures and EPB in Study 3 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Savings Goal | Trait Optimism | Propensity to Plan | Numeracy | Spendthrift - Tightwad | Openness to Experience | Temporal Discounting (Losses) | Temporal Discounting (Gains) | Risk Aversion (Losses) | Risk Aversion (Gains) | Cyclical vs. Linear Time Orientation |
| Week 1 EPB | 0.05 | -0.03 | 0.03 | 0.06 | 0.05 | -0.09 | -0.08 | 0.02 | -0.13 | 0.06 | -0.06 |
| Week 2 EPB | 0.06 | -0.04 | 0.01 | -0.03 | 0.09 | 0.09 | 0.09 | 0.07 | 0.00 | 0.07 | -0.01 |
| Week 3 EPB | -0.06 | -0.07 | -0.03 | 0.09 | 0.00 | -0.11 | 0.04 | -0.03 | 0.00 | -0.06 | -0.04 |
| Week 4 EPB | -0.06 | -0.05 | -0.15* | 0.06 | -0.04 | -0.11 | 0.06 | 0.06 | -0.04 | -0.02 | 0.03 |
| Monthly EPB | -0.04 | -0.09 | -0.08 | -0.02 | -0.07 | -0.10 | -0.14 | 0.09 | -0.01 | -0.04 | 0.00 |

*$p$ < .05

| | Correlation between Individual Difference Measures and Perceived Typicality of Predicted Expenses in Study 3 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Savings Goal | Trait Optimism | Propensity to Plan | Numeracy | Spendthrift - Tightwad | Openness to Experience | Temporal Discounting (Losses) | Temporal Discounting (Gains) | Risk Aversion (Losses) | Risk Aversion (Gains) | Cyclical vs. Linear Time Orientation |
| T0 (week) | 0.10 | 0.03 | -0.03 | 0.02 | -0.04 | -0.07 | 0.00 | 0.03 | 0.07 | -0.13 | 0.13 |
| T0 (month) | 0.12 | 0.04 | 0.14 | -0.07 | -0.03 | -0.08 | 0.07 | 0.01 | 0.05 | 0.04 | 0.04 |
| T1 | -0.03 | -0.04 | -0.01 | 0.08 | -0.01 | -0.06 | 0.02 | -0.01 | -0.07 | 0.00 | -0.02 |
| T2 | 0.02 | -0.16 | 0.02 | 0.00 | 0.05 | -0.16* | 0.04 | 0.04 | -0.02 | 0.00 | 0.02 |
| T3 | -0.10 | 0.01 | -0.01 | 0.05 | 0.03 | -0.13 | 0.04 | 0.05 | -0.03 | 0.00 | 0.21* |
| T4 | 0.06 | 0.00 | 0.00 | 0.10 | 0.05 | -0.03 | 0.11 | -0.06 | 0.11 | 0.05 | -0.04 |

*$p$ < .05

## Discussion

The results of study 3 offer longitudinal evidence from the field that provides compelling support for the prototype theory. First, consumers predicted that their expenses would be more typical (H1) and lower (H3a) than their past expenses in each and every week of the study. Second, perceived typicality was shown to be negatively correlated with expense predictions on a consistent basis (H2). Third, consumers persistently under-predicted their weekly expenses as well as under-predicted their expenses for the month (H3b). Fourth, it was shown that the atypical intervention is capable of virtually eliminating the expense prediction bias in a real-world setting (H4a). The null results regarding the individual differences measured in this study are also consistent with the assertion that EPB is the result of a general cognitive process like prototypical prediction rather than dispositional traits. Finally, the results of Study 3 demonstrate that the magnitude of the expense prediction bias—approximately $100/week or $400/month—is economically significant.

### 3.4 Study 4: Intervention Effectiveness of Weekly vs. Monthly Predictions

The primary goal of Study 4 was to more directly determine if perceived typicality of future expenses differs between weekly and monthly predictions. This is important from a theoretical perspective because it is possible that prototype attributes do not influence predictions for longer periods of time as much as they do for shorter periods of time. For example, a more distant prediction horizon may be associated with more uncertainty, which could cause consumers to build a buffer into their prediction that accounts for atypical expenses (Ulkumen et al. 2008). If true, we would expect to observe lower perceived typicality for monthly (vs. weekly) predictions in the absence of an intervention. It would also be reasonable to expect a weaker effect of the atypical intervention on perceived typicality for monthly (vs. weekly) predictions, because if monthly predictions are less influenced by prototype attributes there should be less room for perceived typicality to be shifted downward by the intervention. Study 4 tests these possibilities, as well as the preregistered hypothesis that the atypical intervention increases monthly as well as weekly expense predictions, which is a matter of practical importance given that Study 3 shows consumers under-predict for both time periods.

**Method**

*Participants and Procedure.* I recruited 601 participants (48.6% female, $M_{age}$ = 37.93) from Amazon Mechanical Turk to take part in a consumer expense survey. Participants were randomly assigned to predict their expenses in a 2 (prediction time frame: week vs. month) × 2 (intervention condition: control vs. atypical) between-subjects design that utilized the same prediction prompts as in Study 3, modified where necessary for monthly predictions. Participants also reported perceived typicality and prediction confidence as in Study 3.

**Results**

*Weekly vs. Monthly Prediction Typicality.* Perceived typicality did not differ as a function of time period, only as a function of intervention condition. Specifically, a 2 (prediction time period: week vs. month) × 2 (intervention condition: control vs. atypical) ANOVA with perceived typicality as the dependent variable revealed a main effect of intervention condition such that perceived typicality was higher in the control condition ($M$ = 5.20, 95% CI = [5.03, 5.38]) than in the atypical condition ($M$ = 3.12, 95% CI = [2.94, 3.30]; $F(1, 597) = 263.73$, $p < .001$, partial eta squared = .31). However, there was no main effect of time period ($F(1, 597) = .72$, $p = .40$) and no intervention condition by time period interaction ($F(1, 597) = 2.09$, $p = .15$).

*Monthly Expense Predictions*. The impact of the atypical intervention on monthly expense predictions was consistent with its impact on weekly expense predictions in the final week of Study 3: Monthly expense predictions were 24.6% higher in the atypical condition ($M$ = $1505.98, 95% CI = [$1314.88, $1724.86]) than in the control condition ($M$ = $1208.34, 95% CI = [$1049.32, $1391.59]), as revealed by an independent-samples t-test ($t(297) = 2.22$, $p = .027$).

Furthermore, perceived typicality was found to be significantly lower in the atypical condition ($M$ = 2.97, 95% CI = [2.74, 3.21]) than in the control condition ($M$ = 5.24, 95% CI = [4.99, 5.50]; $t(297) = 12.91$, $p < .001$, $d = 1.49$). Prediction confidence did differ between the two conditions ($M_{atypical} = 4.79$, 95% = [4.56, 5.01]; $M_{control} = 5.39$, 95% $CI_{control}$ = [5.19, 5.60]; $t(297) = 3.94$, $p < .001$, $d = .45$), but the focal comparison between predicted expenses in the atypical and control conditions remained significant after controlling for confidence ($F(1, 296) = 8.33$, $p = .004$, partial eta squared = .03).

*Weekly Expense Predictions.* The pattern of results observed in the final week of Study 3 was replicated in Study 4: Weekly expense predictions were 61.0% higher in the atypical condition ($M$ = \$335.26, 95% CI = [\$290.03, \$387.53]) than in the control condition ($M$ = \$208.20, 95% CI = [\$181.22, \$239.20]), as revealed by an independent-samples t-test ($t(300) = 4.67$, $p < .001$). Furthermore, perceived typicality was found to be significantly lower in the atypical condition ($M$ = 3.27, 95% CI = [2.98, 3.55]) than in the control condition ($M$ = 5.16, 95% CI = [4.93, 5.40]; $t(300) = 10.16$, $p < .001$, $d = 1.17$). Finally, it was found that prediction confidence did not differ by condition ($M_{atypical} = 5.20$, 95% = [4.99, 5.40]; $M_{control} = 5.35$, 95% $CI_{control}$ = [5.17, 5.54]; $t(300) = 1.13$, $p = .26$).

**Discussion**

The results of Study 4 reveal that perceived typicality of future expenses differs only as a function of intervention condition (control vs. atypical) and not as a function of prediction time period (week vs. month). Moreover, the impact of the atypical intervention on perceived typicality is roughly equal across time periods. In tandem, these findings suggest that the influence of prototype attributes on predictions—and the influence of the atypical intervention—is similar across these two common prediction time frames. The results of Study 4 also demonstrate that prompting consumers to consider atypical expenses can increase monthly expense predictions as well as weekly expense predictions. Taken together with the results of Study 3—which showed that consumers under-predict their monthly expenses as compared to their actual expenses during the target month—this suggests that the atypical intervention is capable of reducing monthly expense prediction bias as well as its weekly cousin.

## 3.5 Study 5: Does Increasing Expense Predictions Increase Saving Intentions?

The results of Studies 1–4 provide consistent support for the prototype theory. The purpose of Study 5 is to extend these results by examining the effect of the atypical intervention on downstream decision-making. Specifically, this study looks at whether or not the atypical intervention can increase saving intentions by increasing expense predictions.

### Method

*Participants and Procedure.* Five hundred and ninety-five US residents were recruited through Amazon Mechanical Turk to participate in a short study about financial decision making (53.4% female, $M_{age}$ = 37.0). Participants were randomly assigned to one of four conditions in a 2(Prediction Condition: Control vs. Atypical Intervention) x 2(Prediction Period: Next Week vs. Next Month) design. Participants predicted their expenses as in Study 4. To measure intention to save, participants were asked to imagine they had just received $1,000, then "How much of the $1,000 would you save to help cover your expenses for the next week/month?" I also measured the following preregistered control variable to minimize error variance in the analyses pertaining to savings: "If you think you are going to *spend* more than you typically do in the future, you should try to compensate by *saving* more than you typically do in the present." (Strongly

Disagree =1, Strongly Agree = 7).[8] My preregistered hypotheses for this study were that the atypical intervention would increase expense predictions versus control (H4a), and that higher expense predictions would be associated with higher savings.

**Results**

*Expense Predictions.* A 2(Prediction Condition: Control vs. Atypical Intervention) x 2(Prediction Period: Next Week vs. Next Month) ANOVA with expense predictions as the dependent variable replicated the effect of the atypical intervention: predictions in the atypical conditions were higher than in the control conditions ($M_{atypical}$ = \$596.45, 95% $CI_{atypical}$ = [522.70, 686.08], $M_{control}$ = \$484.44, 95% $CI_{control}$ = [424.54, 552.80], $F(1, 591) = 4.83$, $p = .028$, partial eta squared = .01). There was also an unsurprising main effect of time period, such that monthly predictions were significantly higher than weekly predictions ($M_{month}$ = \$1378.84, 95% $CI_{month}$ = [1,207.13, , 1,574.98], $M_{week}$ = \$210.61, 95% $CI_{week}$ = [184.01, 241.05], $F(1, 591) = 379.07$, $p < .001$, partial eta squared = .39). There was no interaction between prediction condition and time period ($F(1, 591) = 1.61$, $p = .21$).

*Savings*. A 2(Prediction Condition: Control vs. Atypical Intervention) x 2(Prediction Period: Next Week vs. Next Month) ANOVA with savings as the dependent variable and spend-save compensation as a covariate revealed a main effect of prediction condition such that savings in the atypical conditions were 16.6% higher than in the control conditions ($M_{atypical}$ = \$699.38, 95% $CI_{atypical}$ = [663.08, 735.69], $M_{control}$ = \$599.79, 95% $CI_{control}$ = [564.69, 634.89], $F(1, 590) = 15.00$, $p < .001$, partial eta squared = .03). The ANOVA also revealed a main effect of time period such that intended savings were higher in the monthly conditions than the weekly

[8] A 2(Prediction Condition: Control vs. Atypical Intervention) x 2(Prediction Period: Next Week vs. Next Month) ANOVA with this "compensation" variable as the DV revealed no significant effects ($p$'s > .36). However, it was found to be significantly correlated with the savings intention variable ($r(590) = .15$, $p < .001$).

conditions ($M_{month}$ = \$726.68, 95% $CI_{month}$ = [691.16, 762.20], $M_{week}$ = \$572.50, 95% $CI_{week}$ = [536.60, 608.39], $F(1, 590) = 35.95$, $p < .001$, partial eta squared = .06). There was no interaction between prediction condition and time period ($F(1, 590) = 1.54$, $p = .22$).

*Mediation Analysis*. To test the expectation that the atypical intervention increases intentions to save by increasing expense predictions I performed a mediation analysis with prediction condition as the independent variable (atypical = 1, control = 0), expense predictions as the mediator, and savings as the dependent variable. The spend-save compensation variable was again included as a control, as was prediction time period to account for the 2x2 nature of the study design. (Both of these control variables were included in my pre-registered mediation analysis plan.) The indirect effect of prediction condition on savings was significant (indirect effect = 5.28, SE = 3.90, 95% CI = [.17, 16.57]). Specifically, the model confirms that the atypical intervention succeeded in increasing expense predictions ($B = .21$, SE = .10, 95% CI = [.02, .40]; $t(591) = 2.16$, $p = .032$), and that higher predictions are associated with higher intended savings ($B = 25.30$, SE = 10.89, 95% CI = [3.92, 46.69]; $t(590) = 2.32$, $p = .021$).

**Discussion**

Studies 1–4 provide support for the prototype theory by documenting both descriptive and experimental evidence corroborating the Accessibility → Prototypical Prediction → EPB pathway in Figure 2. Study 5 extends those results by demonstrating that the atypical intervention not only increases predictions, but that higher predictions are also associated with higher intended savings. In Studies 6–8 I turn my attention to the relationship between skew, prototypical predictions, and EPB.

### 3.6 Study 6: Examining Budget Accuracy and Distributional Skew with App Data

The purpose of Study 6 is to examine the association between varying levels of skewness and budget forecast accuracy. To accomplish this, I use data provided by Money Dashboard (https://www.moneydashboard.com/), a financial aggregation app with approximately 70,000 active users in the UK. The primary function of Money Dashboard (MDB) is to provide users with a holistic overview of their financial situation. To do so, MDB collects and combines all transactional information across all financial accounts for each user. So, for example, if a user has two credit cards, a chequing account, and a savings account, MDB will aggregate all inflows and outflows across these cards and accounts and present the user with up-to-the-minute information on when, where, and how they are spending their money. Figure 11 shows the user interface for the mobile and web application, and Table 10 summarizes users' demographic, budgeting and account data.

**Figure 11**

**Money Dashboard Interface**

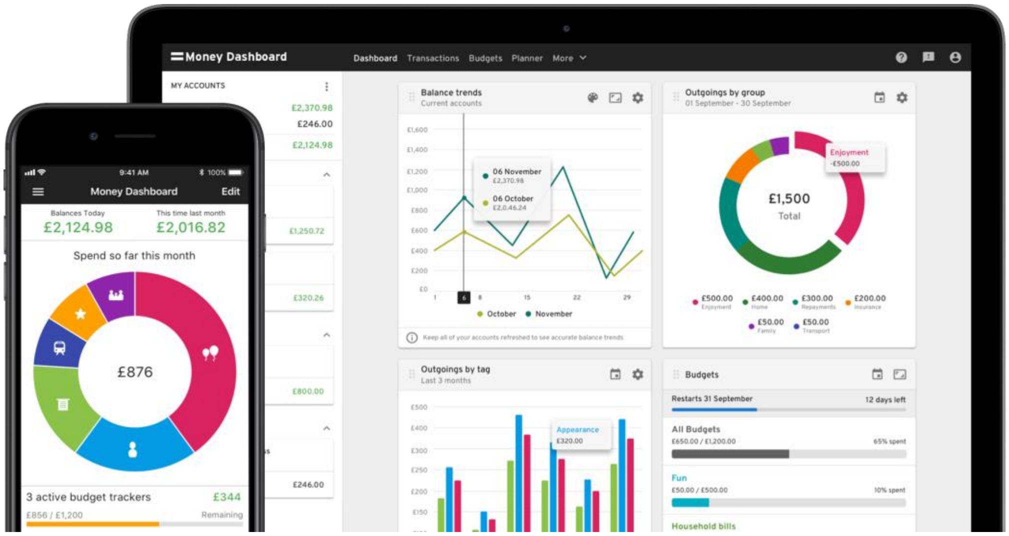



**Table 10**

**Money Dashboard User Characteristics**

| | Mean | Median | St. Dev. |
|---|---|---|---|
| Age | 36.2 | 34 | 9.67 |
| Annual Salary (£ 000's) | 28.4 | 25 | 14.10 |
| Logins per month | 5.8 | 4 | 6.41 |
| # of Budget Categories per User | 1.58 | 1 | 0.709 |
| # of Accounts Linked per User | 4.69 | 4 | 3.54 |
| % Male | 68.2% | N/A | 0.47 |
| % England | 85.6% | N/A | 0.35 |
| % Scotland | 9.6% | N/A | 0.30 |
| % Wales | 4.8% | N/A | 0.21 |

## Spending Data

The data set includes all user transactions – more than 350 million – between January 2014 and December 2016. Each transaction is automatically assigned to a spending category by MDB (e.g., "Groceries"), and includes a merchant tag (e.g., "Tesco"), and a time stamp. A particularly novel feature of the data is that we are able to observe each user's transaction history from *before* they download the app because MDB's terms of service allow it to access each users' transactional data for the twelve months prior to download.

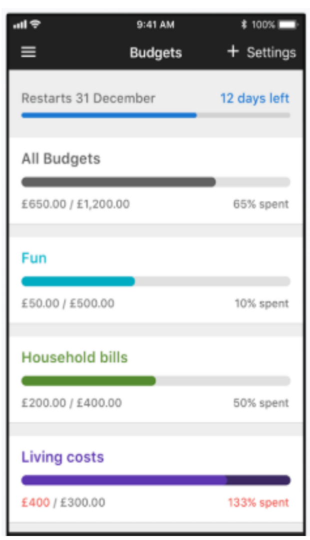
Cash spending represents 2.90% of total spending in our sample, and it appears in the dataset as ATM withdrawals. So, although we do not observe exactly what users spend their cash on, we do observe exactly how much they withdraw. This allows us to estimate cash spending in each budget category using spending data from the UK Office for National Statistics (ONS).

**Budgeting Data**

The MDB budgeting function allows users to set budgets for expenses in multiple categories. For example, a user may set a monthly budget of £150 for dining and drinking, £350 for groceries, and £100 for fuel. MDB then automatically tracks transactions in these categories and allows users to observe their spending against their budget. Users do not receive any push notifications about their budget compliance and they have to manually login to track their expenses. However, when they do login, they are presented with a very salient illustration of their budget and remaining funds in each category, as shown in Figure 12.

**Figure 12**

**Money Dashboard Budget Interface**



In total, the dataset includes 9,403 monthly budgets. In my analysis I focus on the three most popular budget categories: Dining and Drinking (n = 2,479), Groceries (n = 2,618), and Fuel (n = 1,127). As illustrated in Table 11, there is substantial variation in the amount of pre-budget skew displayed in each category. Therefore, if the prototype theory is a correct description of expense prediction behavior "in the wild", we should observe that relative budget accuracy (i.e., the difference between budgeted and actual spending) is substantial worse (i.e, more negative) for dining and drinking than for groceries, and for groceries than for fuel (H5a).

**Table 11**

**Pre-Budget Expense Skew by Budget Category in Study 6**

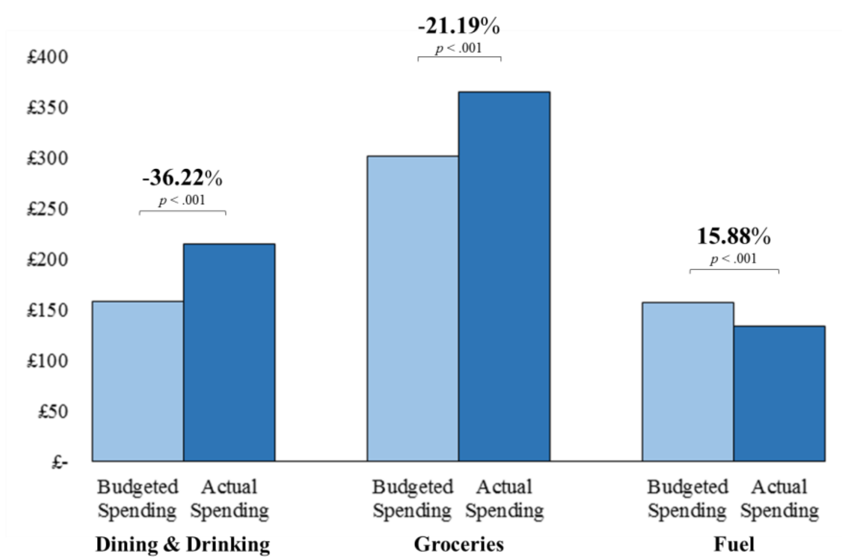
| | Skew |
|---|---|
| Dining and Drinking | 0.377 |
| Groceries | 0.239 |
| Fuel | 0.103 |

## Results

Figure 13 illustrates the relative difference between budgeted and actual spending across the three budget categories. Consistent with H5a, relative budget accuracy is lower for dining and drinking than it is for groceries, and for groceries as compared to fuel. Interestingly, where consumers under-budgeted for both dining and drinking (mean difference = £56.50, SD = 202.00, $t(2,474) = 13.92$, $p < .001$) and groceries (mean difference = £63.48, SD = 259.05, $t(2,638) = 12.58$, $p < .001$), they actually appear to have *over*-budgeted for fuel (mean difference = -£23.84, SD = 118.26, $t(1,132) = 6.79$, $p < .001$). To try and understand why, I examined the percentage change in the price of goods in each budget category during the observation period (2014 to 2016), as reported by the UK Office of National Statistics. The change in the price of goods related to dining and drinking (2.0%) and groceries (2.53%) was small, but the price of fuel fell dramatically (-15.59%), which is remarkably consistent with the difference between the amount consumers budgeted and spent on fuel (+15.88%). This suggests that if consumers had known fuel prices were going to fall and adjusted their budgets accordingly, their budgets would have been fairly accurate, which is consistent with H5a because there is very little skew in the distribution of fuel expenses.

**Figure 13**

**Mean Budgeted and Actual Spending Across Categories in Study 6**




## Discussion

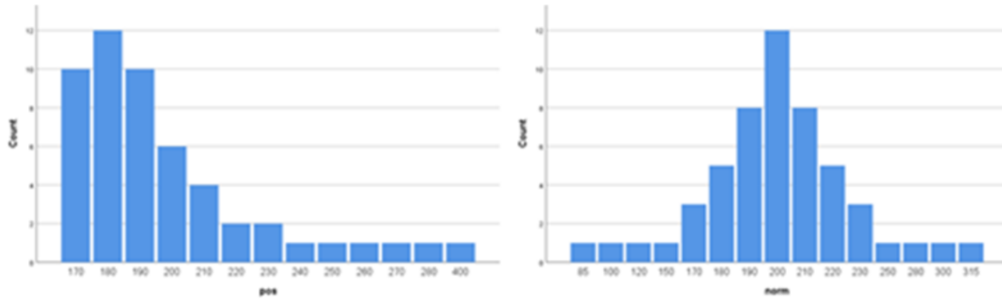
The results of Study 6 support the prototype theory by demonstrating that real world budgets are substantially less (more) accurate when the distribution of expenses is more (less) skewed (H5a). The strengths of this study include a high degree of measurement accuracy and ecological validity; its key limitation is that it is non-experimental and is therefore open to alternative explanations. Thus, in Studies 7 and 8 I use a lab paradigm that manipulates distributional skew to examine the causal effect of skew on predictions.

### 3.7 Study 7: Manipulating Predictions via Distributional Skew

One way to test the veracity of the prototype theory is to experimentally manipulate the mode of the distribution of past expenses and observe subsequent predictions. If the prototype theory is correct, then a positively skewed distribution with mode < mean should lead to a lower prediction than a normal distribution with mode = mean, holding the mean constant. I used Study 7 to test this possibility by presenting participants with 52 weekly expense amounts drawn in random order *without replacement* from either the positively skewed distribution or normal distribution in Figure 14. I then asked participants to predict their expenses for the next week, imagining that the expenses they had seen were an accurate representation of their weekly expenses over the past year. My expectation based on prototype theory was *not* that predictions in the positive-skew condition would exactly equal the mode, because the probability that at least some participants will incorporate distributional information into their prediction is high (as illustrated by the results of the think aloud study), particularly when distributional information is made as salient as it is in this paradigm (Kahneman 2003; Kahneman and Frederick 2002). Rather, my preregistered hypothesis was that predictions in the positive-skew condition would be lower (and farther from the mean) than in the normal condition (H5a).

**Figure 14**

**The Underlying Distributions Used to Manipulate Skew in Study 7**



**Method**

*Participants and Procedure.* Four hundred and one American residents were recruited from Amazon Mechanical Turk to participate in an online study about consumer expenses (49.6% female, $M_{age}$ = 36.9). On the first page of the study participants read the following instructions: "On the next page you will see 52 values presented in quick succession. We want you to imagine that these values represent your weekly spending over the course of one year." On the second page of the study participants were presented with a series of 52 weekly expense amounts, one at a time, every 1.2 seconds. This paradigm was adapted from André, Reinholtz, and de Langhe (2017). It was chosen because it parallels research on perception (e.g., Ariely 2001) and prototype formation is a perceptual process (e.g., Kahneman 2003). The 52 weekly expense amounts were drawn in random order from either the positively skewed distribution on the left hand side of Figure 14 (Min = \$170, Mode = \$180, Mean = \$200, Max = \$400, SD = 38.81, skew = 3.13) or the normal distribution on the right (Min = \$85, Mode = Mean = \$200, Max = \$315, SD = 38.41, skew = 0.00). Mean and range were held constant across conditions, and variance was equated as closely as possible to ensure I was not also manipulating uncertainty. After viewing the expense amounts drawn from one of the two distributions, participants were asked to predict their expenses by responding to the following question: "If the expenses you just saw were an accurate representation of your actual weekly spending over the last year, how much do you estimate you will spend in the next week?"

**Results**

As hypothesized (H5a), predicted expenses were significantly lower in the positive-skew condition ($M$ = \$195.14, SD = 18.95) than in the normal condition ($M$ = \$200.32, SD = 21.44), as confirmed by an independent samples t-test ($t(399) = 2.56$, $p = .011$, $d = .26$). Furthermore, predicted expenses in the positive-skew condition were significantly lower than the \$200 mean of the distribution of past spending ($t(200) = -3.63$, $p < .001$, $d = .26$), but predicted expenses in the normal condition were almost identical to the \$200 mean of past spending ($t(199) = .21$, $p = .83$).

**Discussion**

Study 7 provides support for the prototype theory by establishing a direct causal effect of skew on predictions (H5a). Study 8 replicates and extends this result by examining the mediating role of prototypical predictions.

## 3.8 Study 8: Does Perceived Average Mediate the Effect of Skew on Predictions?

The purpose of Study 8 is to test the hypothesis that the effect of skew on predictions is mediated by what consumers perceive to be their average expenses (H5b). To accomplish this I recruited 351 US residents from Amazon Mechanical Turk (48.3% female, $M_{age}$ = 36.1) to participate in a study that used the same paradigm as in Study 7, but with two adjustments. First, I increased the mean of each distribution to \$300 to ensure that the results observed in Study 7 were generalizable to distributions with higher mean values.[9] Second, in addition to asking participants to predict their expenses for the next week I also asked them "How much do you estimate you spent in an **average week?**" The order of the prediction and perceived average questions was randomized.

**Results**

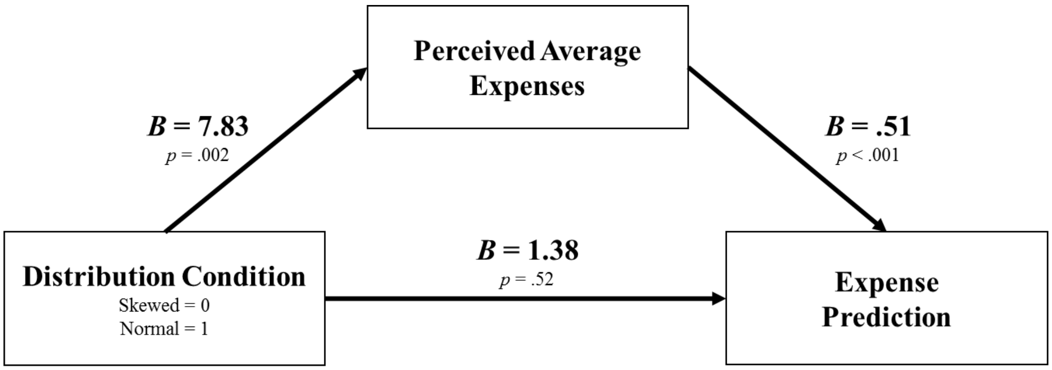
A 2(Distribution Condition: Skewed vs. Normal) x 2(Order: Predict then Estimate Average vs. Estimate Average then Predict) x 2(Judgment Task: Prediction vs. Average Estimation) mixed model ANOVA with distribution condition and order as between-subjects variables and judgment task as a within-subject revealed that the main effect of distribution condition was significant ($F(1, 347) = 9.13$, $p = .003$) and all other effects were not ($F$'s < 1.17 and $p$'s > .28). Consistent with H5a, contrast analysis revealed that predicted expenses in the normal condition ($M$ = \$297.78, SD = 26.58) were significantly higher than predicted expenses in the positively skewed condition ($M$ = \$292.40, SD = 19.62, $p = .030$, $d = .23$). Moreover, predictions in the normal condition did not differ from the \$300 mean of the underlying distribution ($t(150) = -1.03$, $p =.31$) but predictions in the positive skew condition were significantly lower than the \$300 mean ($t(199) = -5.48$, $p < .001$, $d = .39$). Perceived average expenses followed the exact same pattern: they were significantly higher in the normal condition ($M$ = \$298.84, SD = 18.63) than in the skewed condition ($M$ = \$291.01, SD = 26.90, $t(349) = 3.06$, $p = .002$, $d = .29$), and they did not differ from \$300 in the normal condition ($t(150) = -.77$ $p = .45$) but they were significantly lower than \$300 in the skewed condition ($t(199) = -4.73$ $p < .001$, $d = .33$). Finally, as illustrated in Figure 15, a mediation model

[9] Note: I increased the means of each distribution by adding \$100 to every value in the distribution. Thus, although the mean of each distribution increased by \$100 (as compared to Study 7), the shape of the distributions did not change.

with distribution condition as the independent variable, perceived average as the mediator, and prediction as the dependent variable, revealed a significant indirect effect supporting (H5b).[10]

**Figure 15**

**Study 8 Mediation Model**




**Indirect Effect** = 4.00, SE = 1.27, 95% CI = [1.68, 6.68]

### Discussion

Studies 6–8 provide support for the prototype theory by documenting both descriptive and experimental evidence corroborating the Skew → Prototypical Prediction → EPB pathway in Figure 2. In Studies 9a and 9b I extend this investigation of financial misprediction to test the possibility that there exists an *income prediction bias* in which consumers who face variable income *over*-predict their future earnings.

## 3.9 Studies 9a and 9b: Is There an Income Prediction Bias?

The primary purpose of studies 9a and 9b is to determine if consumers who face variable income display an *income prediction bias*. This question is motivated by the rapid growth of the "gig economy," which is defined by temporary, on-call, contract, and freelance work (Katz and

[10] The results of this analysis are substantively unchanged when the order in which participants estimated average expenses vs. predicted future expenses is included as a control variable (indirect effect = 4.01, SE = 1.28, 95% CI = [1.69, 6.69]).

Kruger 2016). Emblematic examples include driving for Uber, selling arts and crafts on Etsy, and even participating in academic research on Amazon Mechanical Turk. The size and speed of the recent shift toward this type of employment has been remarkable. From 2005 to 2015 the number of Americans "gigging" increased by nearly 50%, and 94% of net employment growth in the U.S. economy occurred in gig economy work arrangements (Katz and Kruger 2016).

An important and previously unstudied aspect of employment in the gig economy is that it involves variable income, and it is currently an open question as to whether or not individuals in this circumstance systematically mispredict their weekly or monthly earnings. On the one hand, there are reasons to believe that people will accurately predict their future income. For example, people may engage in income targeting and simply work for as long as it takes to hit their target (Camerer et al. 1997). Notably, this type of behavior is said to be encouraged by some gig economy intermediaries who gamify their apps to keep people working on their platform for longer periods of time (Scheiber 2017). It may also be the case that the motivation to make optimistic predictions in the context of income is far less than, for example, the context of expenses (Peetz and Buehler 2009), because, according to loss aversion, losses (expenses) hurt more than gains (income) feel good (Kahneman and Tversky 1979b). Finally, there is at least one study in the labor economics literature showing that undergraduate students' post-graduation salary expectations are fairly accurate when compared to their self-reported salaries four years after graduation (Webbink and Hartog 2004). Taken together, these findings would seem to suggest that income prediction accuracy is the norm.

However, a complete evaluation of this evidence requires acknowledgement of three contravening points. First, there is some controversy around the existence of income targeting (Farber 2005; Ottenginer 1999), and it has been shown that income targeting among Uber drivers

– who are perhaps the most quintessential example of gig economy workers – dissipates quickly as drivers gain experience (Sheldon 2017). Second, research suggests that optimistic financial predictions are prevalent even in the absence of a motivational goal (Peetz and Buehler 2009; Study 3 of the present research). Finally, the results in the labor economics literature are mixed. Where Webbink and Hartog (2004) found students' salary expectations to be accurate, others have found them to be wildly optimistic (Jerrim 2015; Betts 1996; Smith and Powell 1990).

In addition to the ambiguity of the evidence suggesting that income prediction accuracy might be the norm, there are also reasons to believe that income predictions will be optimistically biased. Most notably, the present research has shown that financial predictions are persistently optimistic in the related context of expenses (see also Peetz and Buehler 2009; Ulkumen et al. 2008). Moreover, the tendency to produce optimistic predictions has been documented for a diverse array of behaviors ranging from consumers' tax filing times (Buehler, Griffin, and MacDonald 1997) and charitable giving (Epley and Dunning 2000), to students' school work (Buehler et al. 1994), to infrastructure mega-projects (Flyvbjerg 2008). Thus, given that optimism in the context of income is predicting more than one ends up earning, I hypothesize that gig economy workers display an *income prediction bias* in which they *over*-predict their future gig income.

Studies 9a and 9b were designed to explore whether or not an income prediction bias exists, not to conclusively determine the underlying psychological mechanism(s). However, if gig economy workers do over-predict their future earnings it would represent an exciting opportunity to extend the prototype theory because the available evidence suggests that many gig economy workers have *negatively* skewed income with mode > mean (in contrast to consumer expenses, which are positively skewed with mode < mean). For example, on a week to week basis Uber drivers who change the amount of time they spend driving are more likely to drive 11-25% *less* in the next week than 11-25% more (Hall and Krueger 2018). This is consistent with the fact that 83% of Uber drivers drive part time, and approximately two-thirds work another job which is more often than not full-time (Hall and Krueger 2018). This implies that most drivers face a fairly strict upper bound on the amount of time they can drive because of their other job. However, there are myriad unforeseen factors that can interfere with a person's ability to drive – a fact that is exemplified in numerous studies of the planning fallacy (see Buehler, Griffin, and Peetz 2010 for a review) – which suggests that income will tend to be negatively skewed because the frequency of driving less than usual will be higher than the frequency of driving more than usual. Thus, if Studies 9a and 9b do reveal an income prediction bias in which gig economy workers over-predict their future gig income, it will open the door to future research that tests the possibility this bias occurs because of prototypical prediction based on negatively skewed income with mode > mean.

**Method**

*Participants and Procedure.* Participants in Study 9a were Uber drivers recruited from the Uber driver subreddit (https://www.reddit.com/r/uberdrivers/). Forty two drivers started the study (4.8% female, $M_{\text{age}}$ = 33.8) and twenty-seven drivers completed it (7.4% female, $M_{\text{age}}$ = 35.5). Participants in Study 9b were Amazon Mechanical Turk (AMT) workers. Two hundred workers started the study (41.5% female, $M_{\text{age}}$ = 34.7) and one hundred and twenty-nine completed it (40.3% female, $M_{\text{age}}$ = 34.9).

Study 9a and 9b followed the same basic procedure: participants were recruited on a Sunday afternoon and asked to predict how much money they would earn from their gig in the next week. They were also asked to predict how many hours they thought they would work at their gig, and their estimated hourly wage. One week later, participants were re-contacted and asked to report their actual earnings and hours. Participants in Study 9b were also asked to report their averagely weekly AMT income. If income over-prediction is caused in part by prototypical prediction wherein "average" refers to the mode of a negatively skewed distribution, then it should be the case that predicted income is similar to average income, but actual income earned is significantly lower than both.

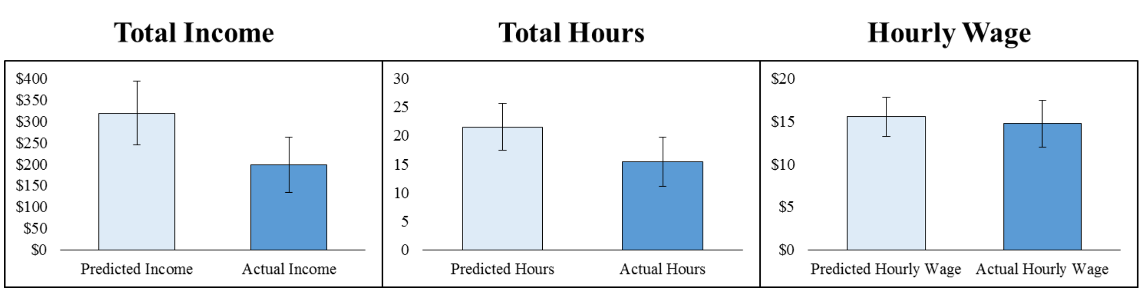
**Results**

As illustrated in Figure 16, the Uber drivers in Study 9a over-predicted their gig income by 60.2% (Mean difference = \$119.82, 95% CI = [62.81, 176.63], $t(26) = 4.32$, $p < .001$). They also predicted they would work 40.0% more hours than they ended up working (Mean difference = 6.5 hours, $t(26) = 4.89$, 95% CI = [3.74, 9.16], $p < .001$). However, they were fairly accurate at predicting their hourly wage (Mean difference = \$0.79, 95% CI = [-2.11, 3.70], $t(24) = .56$, $p =$ .58).

**Figure 16**

**Uber Driver Prediction Accuracy for Uber Income, Hours, and Hourly Wage in Study 9a**

Error Bars Represent 95% Confidence Intervals

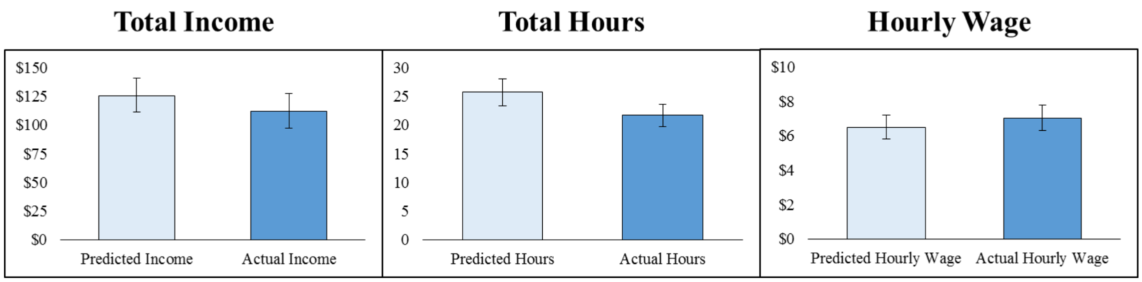



As illustrated in Figure 17, the AMT workers in study 9b displayed the same tendencies as the Uber drivers in Study 9a. AMT workers over-predicted their gig income by 11.6% (Mean difference = \$13.05, 95% CI = [1.01, 1.23], $t(128) = 2.14$, $p = .034$). They also predicted they would work 18.6% more hours than they ended up working (Mean difference = 4.0 hours, 95% CI = [2.33, 5.72], $t(128) = 4.70$, $p < .001$). However, they were fairly accurate (and even somewhat conservative) at predicting their hourly wage (Mean difference = –\$0.55, 95% CI = [-1.18, .08], $t(128) = -1.72$, $p = .088$).

**Figure 17**

**AMT Worker Prediction Accuracy for AMT Income, Hours, and Hourly Wage in Study 9b**

Error Bars Represent 95% Confidence Intervals




Finally, predicted income (M =\$124.72 , 95% CI = [110.57, 140.70]) and average income (M = \$123.87, 95% CI = [110.31, 139.10]) in Study 9b were very similar ($t$(128 = .26, $p = .79$). However, actual income earned during the target week (M = \$111.94, 95% CI = [97.79, 128.15]) was significantly lower than average income ($t(128) = 2.00$, $p = .048$). Although purely descriptive, these results are consistent with a prototype account of income prediction bias.

**Discussion**

Studies 9a and 9b provide what is to the best of my knowledge the first evidence that gig economy workers display an *income prediction bias* in which they *over*-predict their future earnings. Understanding the psychological causes of this bias is an important question for theorists. Controlling for between-subject variation in average weekly income, actual income earned during the target week was negatively skewed in Study 9b (skew = –1.24). This raises the possibility that the income prediction bias can be explained by the prototype theory, because a prototypical prediction based on the mode of *negatively* skewed distribution with mode > mean would result in *over*-prediction. Alternatively, it could be the case that income predictions are based on predicted hours, and the bias is therefore a special case of the planning fallacy in which workers base their predictions on narrow plans for the future rather than their past experience. Understanding the downstream consequences of income prediction bias is also an important question for practitioners, policy-makers, and consumers themselves. For example, if a person believes they will earn more money in the future than is realistic, will they fail to save enough money to fully cover their expenses? If so, how can income prediction accuracy be improved? Investigating these possibilities (among others) represents exciting directions for future research.

## 4. General Discussion

The present research develops and tests a prototype theory of consumer expense misprediction. The logic underlying the theory is that consumers' expense predictions are based on prototype attributes that represent typical or mode expenses. Prototypical prediction then leads to an expense prediction bias in which consumers under-predict their future expenses because the distribution of expenses is positively skewed with mode < mean. Accordingly, an "atypical intervention" was developed that improves prediction accuracy by making atypical expenses more cognitively accessible when consumers make their predictions. I next discuss the implications of this work for theory and practice, as well as directions for future research.

*Understanding the Expense Prediction Bias*

The prototype theory deepens our understanding of the expense prediction bias in several ways. First, it provides a parsimonious explanation for why the bias occurs, revealing for the first time the interactive effect that prototype attributes and distributional skew have on prediction accuracy. Moreover, as highlighted in the theoretical development, the prototype theory has the ability to explain several past findings regarding expense misprediction including distribution neglect (Peetz and Buehler 2012; Peetz et al. 2015), narrow bracketing of exceptional expenses (Sussman and Alter 2012), and the inverse relationship between prediction confidence and accuracy (Ulkumen et al. 2008). However, the persistent null result regarding the association between confidence and accuracy in Study 3 suggests that beliefs about this relationship may need to be updated. The evidence supporting the prototype theory also suggests that expense prediction bias is primarily a function of a basic cognitive process (i.e., prototypical prediction), rather than an individual difference like motivation to save (cf. Peetz and Buehler 2009). Similarly, the results of the present research indicate that the bias is not an instance of pure base-rate neglect (cf. Peetz and Buehler 2012). Instead, it can be said that consumers have easy access to the wrong base-rate (i.e., the mode, not the mean) in their minds. Finally, the prototype theory offers the first intervention that both informs theory and is practically useful in the field.

*The "Average" Represented by Prototype Attributes*

Past research on the use of prototype attributes in judgment and decision making has implied that the "average" represented by prototype attributes is often a simple mean. For example, Kahneman and Frederick (2002) conceptualize the peak-end rule—the phenomenon that global evaluations of a temporally extended experience can be predicted by averaging the peak and the end of the experience (Redelmeier and Kahneman 1996)—as an instance in which people are substituting a prototype attribute for an extensional one. It is therefore a notable contribution of the present research that it identifies a circumstance in which prototype attributes represent outcomes that follow the *mode* of a distribution.

An important direction for future research is to investigate when and why prototype attributes represent the mode versus mean (or median) of a distribution. Past research has concluded that individuals are remarkably adept at mean-identification with respect to sets of shapes (Ariely 2001), numbers (Andre et al. 2017), and faces (Haberman and Whitney 2009). However, the distribution of stimuli in these studies was normal with mode = mean. This raises the intriguing possibility that the perceived average in these studies could have actually been the mode, and that the prototype theory may generalize well beyond expenses.

*Distributional Skew and Prediction Accuracy*

Documenting the relationship between distributional skew and prediction accuracy is an important contribution of the present research because it has received little attention in the psychological literature on prediction. However, the observation that skew should be accounted for when constructing predictions is not entirely without precedent. Reference class forecasting (RCF) is a strategic planning process that urges forecasters to compare a project's potential outcomes (e.g., total cost or completion time) to those of similar past projects in order to produce more accurate predictions (Lovallo and Kahneman 2003; Kahneman and Tversky 1977). The application of RCF to large infrastructure projects like hydroelectric dams has been the focus of research on capital cost forecasting in part because the distribution of cost overruns for these projects is positively skewed, and RCF accounts for skew by utilizing the full distribution of the reference class (Flyvbjerg 2008; Ansar et al. 2014).

One key difference between the prototype theory and work on reference class forecasting is the underlying prediction process. The prototype theory argues that predictions are based on the mode of past spending behavior. In contrast, RCF argues that predictions are based on a narrow set of future plans for the project at hand (Lovallo and Kahneman 2003). This difference is not at all surprising – consumers have ample personal experience to draw from when predicting their own expenses, whereas the planners of large, often one-of-a-kind projects almost certainly do not. However, this contrast does help illuminate one limitation of the prototype theory – its explanatory power is confined to situations in which predictors have a distribution to draw from.

A second difference between the present research and work on reference class forecasting is that the prototype theory argues expense predictions will be optimistic when the relevant distribution is positively skewed, but relatively accurate when it is normal (Studies 6–8). Research on RCF makes no such claim, and some evidence implies that forecasters under-budget for infrastructure projects even when the distribution of actual costs is normally distributed (Ansar et al. 2014). A third, related difference is that the atypical intervention derived from the prototype theory solves under-prediction in the case of positive skew by drawing attention to outcomes that lie predominately in the right tail of the distribution, whereas RCF draws attention to the whole distribution. The former has the advantage of being easily applied; the latter has the advantage of being applicable to distributions of any shape.

Reference class forecasting is a relatively effortful procedure that involves selecting a reference class, establishing a probability distribution for it, and comparing the target outcome to the distribution (Flyvbjerg 2008; Lovallo and Kahneman 2003). This makes it unlikely to gain much traction among consumers given that expense predictions are often made spontaneously (Peetz et al. 2016), and the growing popularity of personal finance apps (Barba 2018) suggests that consumers are actively looking for less effortful ways to predict and manage their spending. Nonetheless, I think that applying RCF in the context of consumer expense predictions is a worthwhile academic exercise because it can help eludicate the extent to which consumers are consciously aware of the skew in their expense distribution.

*Temporal Asymmetry*

Finally, the present research contributes to a nascent literature on temporal asymmetry which hypothesizes that people think about the future in more prototypical terms than the past (Kane, Van Boven, and McGraw 2012; Van Boven, Kane, and McGraw 2008). By comparing perceived typicality of past versus future expenses, I extend this work to the domain of money. This provides a notably conservative test of the temporal asymmetry hypothesis because money is a relatively concrete and predictable resource (MacDonnell and White 2015; Zauberman and Lynch 2005), whereas the hypothetical people, places, and events that participants have been asked to mentally represent in other studies of temporal asymmetry are arguably much more ambiguous. Therefore, because prototypes are generalizations, it is reasonable to believe that people will rely on prototypes less when they are thinking about a resource like money, which has very specific uses. Nonetheless, I find consistent evidence that representations of future expenses (predictions) are more prototypical than representations of past expenses (recall). Another direction for future research is to directly compare and contrast the strength of asymmetries in resources like time versus money.

*Implications for Consumers, Financial Literacy Organizations and Firms*

An important contribution of the present research is that it provides a more comprehensive understanding of expense prediction bias as a phenomenon. For example, I present the first studies to identify the magnitude and persistence of the bias in non-student samples. This research is also the first to study the bias longitudinally and in the field, and to measure monthly expense predictions against actual expenses for the target month. The implications of my findings in this regard are clear: the magnitude of the bias (approximately \$100 per week or \$400 per month in Study 3) is large enough to be economically meaningful for many consumers. Thus, the prosocial benefit of this research is also clear—any consumer can make use of the atypical intervention to improve his or her expense prediction accuracy and make better informed decisions regarding their spending, borrowing, or saving behavior.

One promising channel through which the atypical intervention can be disseminated is financial literacy organizations. Currently, the modal approach of such organizations is to educate their stakeholders about debits and credits, interest rates, and so on. However, this approach is both time consuming and appears to have very limited impact (Fernandes, Lynch and Netemeyer 2014). In contrast, the atypical intervention does not need to be learned, per se, but it can be easily provided (e.g., through text messages) and used to effectively increase prediction accuracy.

The findings herein also have practical implications for for-profit firms. For example, companies in the FinTech sector that are developing and managing budgeting apps can leverage the results to design their products in a way that helps users set more realistic budgets. Given that 63% of North Americans with a smartphone have at least one financial app on their phone (Barba 2018)—the key function of which is often budgeting—this could confer a substantial product advantage. Furthermore, because many behaviors that follow a skewed distribution may be subject to prototyping, I believe that the atypical intervention can also be used to inform the design of products that aim to improve consumers' predictions with respect to calories, exercise, time management, and a host of other variables that can positively impact consumers' well-being. Indeed, yet another important direction for future research is to test the explanatory power of the prototype theory and the effectiveness of the atypical intervention beyond the context of expense prediction.

*Additional Directions for Future Research*

The argument that under-predicting future expenses can lead to problematic downstream consequences like undersaving is fairly intuitive: if, for example, a consumer saves for future expenses based on a prediction of $2,200/month but ends up spending $2,600/month, their savings will be insufficient. This intuition is consistent with the results of Study 5 which show that consumers recognize that higher (lower) future expenses require higher (lower) savings. However, it should be noted that research has not yet established a causal link between expense prediction bias and undersaving (or other related consequences like overspending), nor has it quantified the potential benefits of improving prediction accuracy in terms of real world consumer behavior. This is an important direction for future research that I believe will be aided by research partnerships with personal finance apps like Money Dashboard, because these apps allow researchers to observe consumers' budget forecasts and a number of consequential financial behaviours including savings, debt, payday loan use, credit card use, and so on.

Relatedly, future research should also examine factors that may interact with the atypical intervention to produce unintended negative outcomes. For example, it is possible that the atypical intervention causes some consumers (e.g., those with low income) to experience stress or negative affect because a higher expense prediction might increase feelings of scarcity or financial constraint. It is also possible that the atypical intervention may lead to higher spending (versus control) if consumers use their prediction as a reference point and actively track their spending against it (Heath and Soll 1996). Understanding if, when, and why these negative outcomes might occur is an important next step for mapping the psychology of financial decision making and its behavioral consequences.

Finally, another exciting direction for future research is to systematically investigate the extent to which consumers are aware of their prediction errors, and the extent to which they learn from their errors over time. The objective definition of expense prediction bias is the under-prediction of future expenses; but does under-prediction *feel* like an error to consumers? The results of Study 3 offer some preliminary insight into this question. On average, predictions did not change significantly from week one of the study to week four ($t(175) = 1.31$, $p = .19$), nor did prediction accuracy significantly improve ($t(175) = .92$, $p = .36$). However, somewhat ironically, prediction confidence did increase ($M_{\text{week1}} = 4.87$, $\text{SD}_{\text{week1}} = 1.31$, $M_{\text{week4}} = 5.05$, $\text{SD}_{\text{week4}} = 1.09$, $t(175) = 1.96$, $p = .051$, $d = .15$). This implies that consumers either did not know they were underpredicting, or that underprediction simply did not *feel* erroneous. To more fully understand this phenomenon future research should provide consumers with explicit feedback regarding their prediction accuracy and measure the extent to which that feedback is incorporated into subsequent predictions.

## References


Alter, Adam L. and Daniel M. Oppenheimer (2009), "Uniting the tribes of fluency to form a metacognitive nation," *Personality and Social Psychology Review*, 13 (3), 219-235.

André, Quentin, Nicholas Reinholtz, and Bart de Langhe (2017),"Variance Spillover in Intuitive Statistical Judgments," in *NA - Advances in Consumer Research*, 45, ed. Ayelet Gneezy, Vladas Griskevicius, and Patti Williams, Duluth, MN: Association for Consumer Research, 336-340.

Ansar, A., Flyvbjerg, B., Budzier, A., & Lunn, D. (2014). Should we build more large dams? The actual costs of hydropower megaproject development. *Energy policy*, *69*, 43-56.

Ariely, D. (2001), "Seeing sets: Representation by statistical properties," *Psychological Science*, *12*, 157–162.

Barba, R. (2018), https://www.bankrate.com/personal-finance/smart-money/americans-and-financial-apps-survey-0218/

Beach, L. R., & Swenson, R. G. (1966), Intuitive estimation of means. *Psychonomic Science*, *5*(4), 161-162.

Berman, Jonathan Z., An TK Tran, John G. Lynch Jr, and Gal Zauberman (2016), "Expense neglect in forecasting personal finances," *Journal of Marketing Research*, 53 (4), 535-550.

Betts, J. R. (1996), What do students know about wages? Evidence from a survey of undergraduates. *Journal of human resources*, 27-56.

Buehler, Roger, Dale Griffin, and Johanna Peetz (2010), "The planning fallacy: Cognitive, motivational, and social origins," *Advances in Experimental Social Psychology*, 43, Academic Press, 1-62.

Buehler, R., Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and social psychology bulletin*, *23*(3), 238-247.

Buehler, Roger, Dale Griffin, and Michael Ross (1994), "Exploring the 'planning fallacy': Why people underestimate their task completion times," *Journal of Personality and Social Psychology*, 67 (3), 366.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997), Labor supply of New York City cabdrivers: One day at a time. *The Quarterly Journal of Economics*, *112*(2), 407-441.

Canadian Task Force on Financial Literacy (2019). https://www.practicalmoneyskills.ca/personalfinance/savingspending/budgeting/

CB Insights (2018), Global FinTech Report Q1 2018. https://www.cbinsights.com/research/report/FinTech-trends-q1-2018/

Consumer Federation of America (2018). https://paydayloaninfo.org/facts

Dougherty, M. R., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180.

Dunn, E. W., Gilbert, D. T., & Wilson, T. D. (2011). If money doesn't make you happy, then you probably aren't spending it right. *Journal of Consumer Psychology*, *21*(2), 115-125.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of personality and social psychology*, *58*(4), 568.

Epley, N., & Dunning, D. (2000). Feeling" holier than thou": are self-serving assessments produced by errors in self-or social prediction? *Journal of personality and social psychology*, *79*(6), 861.

Epley, Nicholas and Thomas Gilovich (2006), "The anchoring-and-adjustment heuristic: Why the adjustments are insufficient," *Psychological Science,* 17 (4), 311-318.

Farber, H. S. (2005). Is tomorrow another day? The labor supply of New York City cabdrivers. *Journal of political Economy*, *113*(1), 46-82.

Federal Reserve Bank of New York (2018), Household Debt Continues its Increase in the First Quarter of 2018. https://www.newyorkfed.org/microeconomics/hhdc.html.

Fellowes, Matthew and Willemin, Katy (2013), The Retirement Breach in Defined Contribution Plans. *HelloWallet Research Reports*.

Fernandes, Daniel, John G. Lynch Jr, and Richard G. Netemeyer. "Financial literacy, financial education, and downstream financial behaviors." *Management Science* 60.8 (2014): 1861-1883.

Flyvbjerg, B. (2008). Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice. *European planning studies*, *16*(1), 3-21.

Gilbert, D. T., Morewedge, C. K., Risen, J. L., & Wilson, T. D. (2004). Looking forward to looking backward: The misprediction of regret. *Psychological Science*, *15*(5), 346-350.

Gilbert, D. T., & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science*, *317*(5843), 1351-1354.

Gilbert, D. T., & Wilson, T. D. (2009). Why the brain talks to itself: Sources of error in emotional prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1335-1341.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.

Gillespie, P. (2017), Intuit: Gig economy is 34% of US workforce. *CNN Money*.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718.

Jerrim, J. (2015). Do college students make better predictions of their future income than young adults in the labor force?. *Education Economics*, *23*(2), 162-179.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The big five inventory: Versions 4a and 54 [Technical Report]. *Berkeley: University of California, Institute of Personality and Social Research*.

Johnson, Marcia K. and Carol L. Raye (1981), "Reality monitoring," *Psychological Review,* 88 (1), 67.

Johnson, E. J., G. Häubl, and A. Keinan (2007), “Aspects of endowment: A query theory of value construction,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33* (3), 461.

Kahneman, Daniel (2003), “A perspective on judgment and choice: Mapping bounded rationality,” *American Psychologist*, *58* (9), 697.

Kahneman, Daniel and Shane Frederick (2002), "Representativeness revisited: Attribute substitution in intuitive judgment," *Heuristics and Biases: The Psychology of Intuitive Judgment,* 49, 81.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, *3*(3), 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237.

Kahneman, D., & Tversky, A. (1977). *Intuitive prediction: Biases and corrective procedures*. Decisions and Designs Inc Mclean Va.

Kahneman, Daniel and Amos Tversky (1979), “Intuitive prediction: Biases and corrective procedures,” *TIMS Studies in Management Science, 12,* 313-327.

Kane, Joanne, Leaf Van Boven, and A. Peter McGraw (2012), "Prototypical prospection: Future events are more prototypically represented and simulated than past events," *European Journal of Social Psychology,* 42 (3), 354-362.

Katz, L. F., & Krueger, A. B. (2016). *The rise and nature of alternative work arrangements in the United States, 1995-2015* (No. w22667). National Bureau of Economic Research.

Kirby, Kris N. and Nino N. Maraković (1996), "Delay-discounting probabilistic rewards: Rates decrease as amounts increase," *Psychonomic Bulletin & Review,* 3 (1), 100-104.

Kruger, Justin and Matt Evans (2004), "If you don't want to be late, enumerate: Unpacking reduces the planning fallacy," *Journal of Experimental Social Psychology,* 40 (5), 586-598.

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological review*, *125*(1), 1.

Lynch Jr., John G., Richard G. Netemeyer, Stephen A. Spiller, and Alessandra Zammit (2010), "A generalizable scale of propensity to plan: The long and the short of planning for time and for money," *Journal of Consumer Research,* 37 (1) 108-128.

MacDonnell, Rhiannon and Katherine White (2015), "How construals of money versus time impact consumer charitable giving," *Journal of Consumer Research,* 42 (4), 551-563.

Oettinger, Gerald S. (1999). "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors," *Journal of Political* Economy 107, 360-392.

Oxford English Dictionary (2019). https://tinyurl.com/vemu8y4

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), 1579-1591.

Peetz, Johanna and Roger Buehler (2012), "When distance pays off: The role of construal level in spending predictions," *Journal of Experimental Social Psychology,* 48 (1), 395-398.

Peetz, Johanna and Roger Buehler (2013), "Different goals, different predictions: Accuracy and bias in financial planning for events and time periods," *Journal of Applied Social Psychology*, *43* (5), 1079-1088.

Peetz, Johanna, Roger Buehler, Dereck J. Koehler, and Esther Moher (2015), "Bigger not better: Unpacking future expenses inflates spending predictions," *Basic and Applied Social Psychology*, *37* (1), 19-30.

Peetz, Johanna, Melanie Simmons, Jingwen Chen, and Roger Buehler (2016), "Predictions on the go: Prevalence of spontaneous spending predictions," *Judgment and Decision Making,* 11 (1), 48.

Pew Charitable Trusts (2012), "Payday Lending in America: Who Borrows, Where They Borrow, and Why," *Safe Small-Dollar Loans Research Project*.

Rick, Scott I., Cynthia E. Cryder, and George Loewenstein (2007), “Tightwads and spendthrifts,” *Journal of Consumer Research, 34* (6), 767-782.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*(1), 3-8.

Rosch, E. and C. B. Mervis (1975), “Family resemblances: Studies in the internal structure of categories,” *Cognitive Psychology, 7,* 573–605.

Scheiber, N. (2017). How Uber uses psychological tricks to push its drivers’ buttons. *The New York Times*, *2*.

Scheier, Michael F., Charles S. Carver, and Michael W. Bridges (1994), "Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A re-evaluation of the Life Orientation Test," *Journal of Personality and Social Psychology,* 67 (6), 1063.

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological science*, *9*(5), 340-346.

Schwartz, Lisa M., Steven Woloshin, William C. Black, and H. Gilbert Welch (1997), "The role of numeracy in understanding the benefit of screening mammography," *Annals of Internal Medicine,* 127 (11), 966-972.

Sheldon, M. (2016). Income targeting and the ridesharing market. *Unpublished manuscript. Available at: https://static1. squarespace. com/static/56500157e4b0cb706005352d*, *56*, 1457131797556.

Smith, H. L., & Powell, B. (1990). Great expectations: Variations in income expectations among college seniors. *Sociology of Education*, 194-207.

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.

Tam, Leona and Utpal Dholakia (2014), "Saving in cycles: How to get people to save more money," *Psychological Science,* 25 (2), 531-537.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207-232.

Tversky, Amos and Daniel Kahneman (1974), "Judgment under uncertainty: Heuristics and biases," *Science,* 185 (4157), 1124-1131.

Tversky, A., & Kahneman, D. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291.

Tversky, Amos and Derek J. Koehler (1994), "Support theory: A nonextensional representation of subjective probability," *Psychological Review,* 101 (4), 547.

Ülkümen, Gulden, 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), 245-256.

UK Office of National Statistics (2019). https://www.ons.gov.uk/economy/inflationandpriceindices

US Federal Reserve (2016), *Report on the Economic Well-Being of US Households in 2015*. https://www.federalreserve.gov/2015-report-economic-well-being-us-households-201605.pdf.

Vallone, R. P., Griffin, D. W., Lin, S., & Ross, L. (1990). Overconfident prediction of future actions and outcomes by self and others. *Journal of Personality and Social Psychology*, *58*(4), 582.

Van Boven, Leaf, Joanne Kane, and Peter A. McGraw (2008), "Temporally asymmetric constraints on mental simulation: Retrospection is more constrained than prospection," ed. K. Markman, W. Klein, and J. Suhr, *The handbook of imagination and mental simulation* (pp. 131–149). New York: Psychology Press.

Webbink, D., & Hartog, J. (2004). Can students predict starting salaries? Yes!. *Economics of Education Review*, *23*(2), 103-113.

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of personality and social psychology*, *78*(5), 821.

Winkielman, Piotr, Jamin Halberstadt, Tedra Fazendeiro, and Steve Catty (2006), "Prototypes are attractive because they are easy on the mind," *Psychological Science,* 17 (9), 799-806.

Yang, Sha, Livia Markoczy, and Min Qi (2007), "Unrealistic optimism in consumer credit card adoption," *Journal of Economic Psychology,* 28 (2), 170-185.

Zauberman, Gal and John G. Lynch, Jr. (2005), "Resource slack and propensity to discount delayed investments of time versus money," *Journal of Experimental Psychology: General,* 134 (1), 23.

## Appendix A: Exploratory Variables Measured in Study 2

The following variables were measured in Study 2. My intuition was that the atypical intervention would *decrease* perceived financial slack, discretionary spending, and borrowing, and *increase* saving/investment. The data did not support my intuition – the effect of prediction condition on each of these variables was not significant ($p$'s > .20). However, these results taught me more about study design than they did about the effect of the atypical intervention on downstream decision making, hence why I present them here and not in the main text. Specifically, these results taught me that the effect of the intervention on downstream decision making is best understood one decision at a time, with no intermediate questions. So, for example, I now know that the best way to understand the effect of the intervention on intention to save is to manipulate predictions then immediately measure intention to save, not to manipulate predictions *and* measure recall (in counterbalanced order), have participants complete an expense listing task (for prediction and recall, also in counterbalanced order), indicate perceived typicality of their past and future expenses, then complete several decision making tasks, one of which measures intention to save. This learning is reflected in the design of Study 4, and many other studies that I've conducted for research related to my dissertation.

**Financial Slack** (Zauberman and Lynch, 2005)

Using the scale below, please indicate how much spare money you expect to have in the next week, compared to an average week in your life. (1 = Very little spare money; 7 = A lot of spare money).

**Discretionary Spending**

Imagine that a friend invites you to go out for a fancy dinner next week. You will each pay for your own food and drinks. How much money would you be willing to spend on dinner, including all your food, drinks, taxes, and tip?

$0 - $10
$11 - $20

$21 - $30
$31 - $40
$41 - $50
$51 - $60
$61 - $70
$71 - $80
$81 - $90
$91 - $100
More than $100

**Borrowing**

Imagine that next week you find yourself with an unexpected bill. For example, suppose that you use your vehicle to get to work, and it requires an expensive repair that is not covered by insurance. Until you get your car repaired, you will have to walk to work, which will add an extra 60 minutes onto your commute each way.

Now imagine that to help cover the cost of fixing your vehicle you are able to take out a $350 loan which will need to be repaid in 2 weeks along with the lender's fee. Using the scale below, please indicate the highest lender's fee you would be willing to pay to be able to borrow the $350.

$0
$5
$15
$20
$25
$30
$35
$40
$45
$50
More than $50

**Repayment Confidence**

Assuming that you took the loan offered in the previous question, how confident are you that you would be able to pay back the loan (including the lender's fee) within 2 weeks? (1 = Extremely confident; 7 = Extremely unconfident).

**Financial Allocation**

Imagine that you have just inherited $1,000 that you weren't expecting. How much of the $1,000 would you use for each of the following? (Please note that your total must equal $1,000).

Spending: $ [Number Box; range 0-1000]

Saving: $ [Number Box; range 0-1000]

Debt repayment: $ [Number Box; range 0-1000]

Total: $ [Number Box; range 0-1000]

**Available Resources**

Note: This measure was included so that I could gain some insight as to whether or not EPB is associated with socio-economic status. There was no correlation between this measure and EPB ($r(1023) = .04$, $p = .16$).

Imagine that you have to pay an unexpected bill immediately. For example, suppose that you require an expensive medical procedure that is not covered by insurance. Considering all possible resources available to you (including savings, borrowing, etc.), what is the maximum dollar amount that you could come up with on short notice? Please enter the amount below.

$ [NUMBER BOX, range 0-99999]

## Appendix B: Individual Difference Measures Used in Study 3

**Trait Optimism** (Scheier, Carver, and Bridges, 1994)

Please indicate your level of agreement with each of the following statements. (1 = Strongly Disagree; 5 = Strongly Agree)

In uncertain times, I usually expect the best.

It is easy for me to relax. (Filler item)

If something can go wrong for me, it will. (Reverse coded)

I'm always optimistic about my future.

I enjoy my friends a lot. (Filler item)

It is important for me to keep busy. (Filler item)

I hardly ever expect things to go my way. (Reverse coded)

I don't get upset too easily. (Filler item)

I rarely count on good things happening to me. (Reverse coded)

Overall, I expect more good things to happen to me than bad.

**Short-term Financial Propensity to Plan** (Lynch et al., 2010)

Please indicate your level of agreement with each of the following statements. (1 = Strongly Disagree; 6 = Strongly Agree)

I set financial goals for the next few days for what I want to achieve with my money.

I decide beforehand how my money will be used in the next few days.

I actively consider the steps I need to take to stick to my budget in the next few days.

I consult my budget to see how much money I have left for the next few days.

I like to look to my budget for the next few days in order to get a better view of my spending in the future.

It makes me feel better to have my finances planned out for the next few days.

**Spendthrift-Tightwad** (Rick, Cryder, and Lowenstein, 2008)

Below is a scenario describing the behaviour of two shoppers. After reading about each shopper, please answer the question that follows.

Mr. A is accompanying a good friend who is on a shopping spree at a local mall. When they enter a large department store, Mr. A sees that the store has a "one-day-only-sale" where everything is priced 10-60% off. Mr. A realizes that he doesn't really need anything, but can't resist and ends up spending almost $100 on stuff.

Mr. B is accompanying a good friend who is on a shopping spree at a local mall. When they enter a large department store, Mr. B sees that the store has a "one-day-only-sale" where everything is priced 10-60% off. Mr. B figures he can get great deals on many items that he needs, yet the thought of spending money keeps him/her from buying the stuff.

In terms of your own behaviour, who are you more similar to, Mr. A or Mr. B?

(1 = Mr. A; 5 = Mr. B)

**Numeracy** (Schwartz et al., 1997)

Please answer the question below without a calculator, but feel free to use scratch paper.

Let’s say you have 200 dollars in a savings account. The account earns 10% interest per year. How much would you have in the account at the end of two years? (Answer provided in a free response text box).

**Linear vs. Cyclical Time Orientation** (Tam and Dholakia, 2014)

A **cyclical time orientation** is the belief that life consists of many small and large cycles, that is, events that repeat themselves. According to this orientation, the future will be similar to the past.

A **linear time orientation** is the belief that life consists of separate and progressive time compartments such as the past, present, and future. According to this orientation, the future will be different from the past.

To what extent do you believe that life is cyclical versus linear? (1 = Life is completely cyclical; 7 = Life is completely linear)

**Openness to Experience** (John, Donahue, and Kentle, 1991)

Below there are a number of characteristics that may or may not apply to you. For each statement please indicate the extent to which you agree or disagree with that statement.

I see Myself as Someone Who... (1 = Strongly Disagree; 7 = Strongly Agree)

Is original, comes up with new ideas.

Is curious about many different things.

Is ingenious, a deep thinker.

Has an active imagination.

Is inventive.

Values artistic, aesthetic experiences.

Prefers work that is routine. (Reverse coded).

Likes to reflect, play with ideas.

Has few artistic interests. (Reverse coded).

Is sophisticated in art, music, or literature.

**Temporal Discounting for Losses and Gains** (Kirby & Maraković 1996)

To measure temporal discounting for losses each of the following items was presented as a binary choice (0 = smaller soon, 1 = larger later). The same items were then used to measure temporal discounting for gains, but "pay" was replaced with "receive."

Would you prefer to **pay** $54 today, or $55 in 117 days?

Would you prefer to **pay** $55 today, or $75 in 61 days?

Would you prefer to **pay** $54 today, or $60 in 111 days?

Would you prefer to **pay** $54 today, or $80 in 30 days?

Would you prefer to **pay** $47 today, or $50 in 160 days?

Would you prefer to **pay** $67 today, or $75 in 119 days?

Would you prefer to **pay** $69 today, or $85 in 91 days?

Would you prefer to **pay** $49 today, or $60 in 89 days?

Would you prefer to **pay** $41 today, or $75 in 20 days?