Title

Essays on the Uncertainty of Continuous Outcomes

Author

Zimmerman, David

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on the Uncertainty of Continuous Outcomes

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Management

by

David Zimmerman

2022

ABSTRACT OF THE DISSERTATION

Essays on the Uncertainty of Continuous Outcomes

by

David Zimmerman

Doctor of Philosophy in Management

University of California, Los Angeles, 2022

Professor Stephen A. Spiller, Co-Chair

Professor Suzanne Bliven Shu, Co-Chair

People are continually faced with decisions that have uncertain outcomes. Often there are many different uncertainties within each of the options available. This research focuses on uncertainty that is continuous, such as the amount of time it will take for a person to commute home or their monthly grocery expenses. When making choices with uncertain outcomes people must (1) construct judgments about the nature of the uncertainty and then (2) form a preference about these uncertainties. Chapter 1 looks at the impacts of different elicitations metrics on uncertainty judgments. Specifically, how do different, but equivalent metrics, impact the amount and shape of the uncertainty estimated. This research relies on three empirical regularities from the cognitive psychology literature: uncertainty perceptions scale with magnitude, a tendency to neglect units, and a strong prior that uncertainty distributions are symmetric. Capitalizing on these empirical regularities, we find striking inconsistencies in judgments of uncertainty, both the amount and the symmetry, depending on the elicitation metric used. In Chapter 2, we find a robust preference for positively correlated uncertainty over negatively correlated uncertainty in the domain of gains. Based on the stimuli used, only someone with risk-seeking preferences would be expected to choose the option with positively correlated uncertainty. Two common decision heuristics, two different display formats, and three different outcomes all result in the same general preference for positively correlated uncertainty. A third display format, similar to a histogram, was the only manipulation that appeared to align single attribute risk preferences and multi-attribute choice. Both chapters offer new phenomena to be accounted for in future models of uncertainty.

The dissertation of David Zimmerman is approved.

Hal Ersner Hershfield

Craig R. Fox

Suzanne Bliven Shu, Committee Co-Chair

Stephen A. Spiller, Committee Co-Chair

University of California, Los Angeles

2022

CONTENTS

List of Figures

ACKNOWLEDGEMENTS

Chapter 1 is in preparation for submission to a journal. Stephen Spiller, Nicholas Reinholtz, and Sam Maglio are co-authors. Stephen contributed to idea generation, research design, data analysis, and manuscript preparation. Nicholas Reinholtz and Sam Maglio contributed to idea generation, research design, and manuscript preparation. Funding for this research partially came from the Morrison Center for Marketing & Data Analytics.

CURRICULUM VITAE

David Zimmerman

EDUCATION

PhD, Management (Specialization in Behavioral Decision Making) Expected August 2022
Anderson School of Management, University of California, Los Angeles
Advisors: Stephen Spiller and Suzanne Shu
Dissertation Committee Members: Hal Hershfield and Craig Fox

Bachelor of Science, Decision Sciences (additional major: Statistics) 2014
Carnegie Mellon University

PUBLICATIONS

Warmath, D., & **Zimmerman, D**. (2019). Financial Literacy as More than Knowledge: The Development of a Formative Scale through the Lens of Bloom's Domains of Knowledge. *Journal of Consumer Affairs*, *53*(4), 1602-1629.

MANUSCRIPTS UNDER REVIEW

Chin, Alycia, **Zimmerman, David**, Johnson, Heidi, and Shu, Suzanne. Disclosure Design, Consumer Comprehension, and Decisions about Overdraft Services. *Resubmitted for Second Round R&R at Journal of Experimental Psychology: Applied*

WORKING PAPERS

Zimmerman, David, Spiller, Stephen, Reinholtz, Nicholas, and Maglio, Sam (2021), "When Metrics Matter: The Impact of Elicitation Metric on Uncertainty Estimates." *Preparing for submission at JEP: General*

Cooper, Cheryl, Jamison, Julian, Knoll, Melissa, and **Zimmerman, David**. Tools for saving: Using prepaid cards to set aside funds.

Hershfield, Hal H., Kerbel, Susan, and **Zimmerman, David**. Exploring the Distribution and Correlates of Future Self-Continuity in a Large, National Representative Sample

SELECTED WORKS IN PROGRESS

Zimmerman, David (2022) “Risk Loving, Confused, or Misled? The Mystery of Risk Seeking Choices” *In preparation.*

CONFERENCE PRESENTATIONS

* Denotes session chair

When Metrics Matter: The Impact of Elicitation Metric on Uncertainty Estimates. With Stephen A. Spiller, Nicholas Reinholtz, and Sam Maglio. Special Session* at Association for Consumer Research, Oct 2021.

Chin, Alycia, **Zimmerman, David**, Johnson, Heidi, and Shu, Suzanne (2020), “Disclosure Design, Consumer Comprehension, and Decisions about Overdraft Services.” Poster Presentation at Boulder Summer Conference on Consumer Financial Decision Making. Note this conference was cancelled due to Covid-19.

TEACHING EXPERIENCE

Teaching Assistant

Organizational Behavior, Eugene Caruso, 2018
Data Analytics for Marketing and Finance, Stephen Spiller, 2019-2022
Managerial Decision Making, Craig Fox, 2020
Communications Development for Leaders, Craig Fox, 2021
Organizational Behavior, Joyce He, 2021

INTRODUCTION

There are several normative and descriptive theories of how people make choices over risky and uncertain outcomes. From expected value theory (Edwards, 1954) to expected utility theory (von Neumann & Morgenstern, 1944), to subjective probabilities (Finetti, 1937), to prospect theory (Kahneman & Tversky, 1979) and eventually cumulative prospect theory (Tversky & Kahneman, 1992) there have been progressions in the modeling of how people make decisions and how they should make them. In each of these models, the stimuli used to test these theories typically involve simple risky decisions, often with two or three outcomes. As a result, these stimuli allowed for clean tests of the theories. Even after the transition from risk to uncertainty, with the discovery of ambiguity aversion (Ellsberg, 1961; Fox & Tversky, 1995), the structure of the problems retained this simplicity. For instance, people bet on urns with balls of two or three different colors and an unknown distribution. This was one incremental step toward evaluating how people deal with more complicated, uncertain decisions (e.g., investing in retirement savings).

The present research investigates uncertain decisions that have straightforward statistical properties but are more complex than what has often been studied. Instead of having a single outcome to a risky (e.g., a coin flip) or uncertain choice (e.g., choosing a ball from an urn filled with some mixture of red and blue balls), this research examines how people form beliefs about uncertain distributions and assesses preferences about the relationship between attributes of uncertain options. Similar research has examined people's preferences in stock portfolio selection, specifically when asked to allocate to multiple assets with differing correlations (Cornil et al., 2019; Reinholtz et al., 2021; Ungeheuer & Weber, 2021). For all of the work on risk and uncertainty models, many real-world decisions that these models hope to predict involve much more complex uncertainty structures, and this research aims to start bridging this gap.

An unrelated stream of decision-making research has focused on very complex decision-making environments with many interacting uncertainties. Retirement—both the saving and spending phases—is just one example of a decision domain in which people must consider various factors, many of which are uncertain and may interact in very complicated ways. Retirement saving and spending decisions require people to grapple with the uncertainties of longevity, long-term changes in the stock market performance, and the amount of money needed for an enjoyable retired life. Each of these factors is not only highly uncertain, but they may also interact in complex ways; for instance, consider the ways in which uncertainty about how much money is needed to comfortably retire relates to uncertainty about stock market performance and investment. In service of these kinds of considerations, this dissertation aims to incrementally test the current theories of risky and uncertain decisions in situations that are somewhat more representative of the complex decisions people are often required to make.

This dissertation has two chapters, each of which examines how people make judgments and decisions in the face of continuous uncertainty—uncertainty about outcomes that can take a continuous range of values (e.g., the amount of time it takes to drive home) rather than discrete outcomes (e.g., the outcome of a coin flip). While the contemporary risk and uncertainty theories do not explicitly only apply to discrete outcomes, it is important to validate the predictions of the current models for these more complex situations since many real-world decisions involve continuous outcomes.

The first chapter elaborates on three empirically established patterns regarding how individuals evaluate uncertainty regarding a focal quantity. The first is that uncertainty increases as the magnitude of the values used in estimation increases (Hogarth, 1975; Weber et al., 2004). In Chapter 1, we utilized metrics that differ by a known constant to test whether the magnitude of peoples' estimates impacts the amount of uncertainty they give for the same outcome. We find that, when people give estimates with larger magnitudes, they tend to have more uncertainty around those estimates. Second, people tend to be somewhat insensitive to the outcome unit (Burson et al., 2009; Raghubir & Srivastava, 2002). When eliciting estimates that differ by a multiplicative factor, we find that people do not sufficiently adjust their estimates based on the actual outcome units. Further, we find that people using larger units (e.g., dollars) give estimates of uncertainty that are wider than those using smaller units (e.g., cents). Lastly, people tend to give symmetric estimates of the uncertainty (Flannagan et al., 1986). Using metrics that are inversely related (e.g., income in cents/minute or minutes/dollar earned), we found that people give more symmetric estimates in their elicited metric compared to when this metric is transformed.

Chapter 2 examines how people choose between options when they each have multiple uncertain attributes. When the correlations between uncertainties change, the outcome can significantly change, and we used this to test if people have preferences regarding these correlations. We find that people have a surprising preference for positively correlated outcomes, a preference that cannot be explained by any decision heuristic or the most common theories of risk and uncertainty.

This dissertation focuses on how people make judgments and decisions when faced with continuous uncertainty rather than discrete outcome possibilities. The two following chapters examine the possibility of inconsistent judgments depending on how that uncertainty is elicited (Chapter 1) and how people construe the uncertainty (Chapter 2).

CHAPTER 1

WHEN METRICS MATTER: HOW REASONING in

DIFFERENT METRICS IMPACTS JUDGMEENTS OF UNCERTAINTY

Alice receives an offer of $300,000 for her house. Should she accept it, or reject it in hopes for a better one? Bobby is throwing a party and has ten six-packs of beer. Should he buy more to hedge against running out? Claire is shopping for a new, fuel-efficient car. She finds a car online that gets 32 miles per gallon. Is that efficient enough, or should she keep looking for other, potentially more efficient options?

Each of these scenarios involves uncertainty: Is there someone who will offer more than $300,000 for Alice's house? How many beers are Bobby's guests liable to drink? Can Claire find a car with substantively better fuel economy? The subjective distribution of outcomes for these estimates will influence people's actions. For example, if Bobby thinks there is less than a 10% chance his guests will drink more than 60 beers, he might decide against getting more.

At the same time, many quantities can be represented in different yet equivalent metrics[1]. For instance, Alice's $300,000 home with a $240,000 mortgage can be considered in terms of its value ($300,000) or her equity ($60,000). We examine whether the metric used in reasoning about uncertainty (e.g., whether Alice focuses on her home's value or her home equity, whether

[1] Throughout, we use "metric" and "unit of measurement," or simply "unit" interchangeably.

Bobby ponders his supply as “how many beers” or “how many six packs of beer”, and whether Claire considers how many miles she gets from a gallon or how many gallons it takes to go 1,000 miles) influences people’s perceptions of uncertainty. We focus on situations with multiple metrics that have a one-to-one mapping to test whether equivalent metrics can lead to inconsistent judgments of uncertainty.

More specifically, we focus on three types of transformations, both prevalent and potentially suspectable to inconsistencies: (i) addition of a constant (e.g., Alice’s home value vs. home equity), (ii) multiplication by a constant (e.g., Bobby’s count of beers vs. six packs), and (iii) inversion of a ratio (e.g., Claire’s consideration of miles per gallon vs. gallons per thousand miles). In each case, prior research suggests discrepancies may occur in judgments of uncertainty: (i) People’s judgments of risk and uncertainty scale with numeric magnitude (Hogarth, 1975; E. U. Weber et al., 2004), suggesting potential discrepancies between metrics that differ by an additive constant. (ii) People tend to be insensitive to units (Burson et al., 2009; Raghubir & Srivastava, 2002), suggesting potential discrepancies between metrics that differ by a multiplicative constant. (iii) People tend to believe distributions are symmetric (Flannagan et al., 1986), suggesting potential discrepancies between metrics that differ via inversion of a ratio.

Risk Scales with Magnitude

Estimates of subjective risk are positively related to both the standard deviation of the observed data and the mean of the distribution (Hogarth, 1975; E. U. Weber et al., 2004). Thus, if people are told that the mean of a distribution is higher, they expect the distribution to have a higher standard deviation (Hofstätter, 1939; Reinholtz, 2015). For example, if people see distributions with equal variability, they give estimates with more variability for the distribution with the higher mean (Beach & Scopp, 1968; Lathrop, 1967; E. U. Weber et al., 2004).

As a result, considering an unknown quantity in one of two alternative metrics that differ by an added constant might lead to inconsistent perceptions of the variability. Specifically, we predict that people should report greater uncertainty in metrics that result in more numerous estimates than for their less-numerous counterpart metrics. For example, Alice could form expectations for a reasonable value for her home in terms of sale price or her equity. Because these two metrics differ by a known amount (her $240,000 mortgage), they should have the same variability. Nonetheless, we expect reasoning about sale price will have greater uncertainty than reasoning about her equity simply from the higher numerosity of the sale price.

Unit Insensitivity

People are insufficiently sensitive to unit when making evaluations or predictions. This holds in contexts ranging from cell phone plans to calorie information for food (Burson et al., 2009; Pandelaere et al., 2011; Shen & Urminsky, 2012; Wertenbroch et al., 2007). For instance, Americans spend lavishly in Europe but more stingily in Mexico, in part because they evaluate the number on the price tag without fully appreciating that the units are in euros (leading to smaller numbers on price tags) or pesos (larger numbers on price tags). More generally, people working in units that are a multiple of a familiar unit do not sufficiently adjust their evaluations in the less familiar unit (Maglio & Trope, 2011; Pelham et al., 1994; Raghubir & Srivastava, 2002).

This suggests people who reason about uncertainty in different units that are multiples of each other may also be insufficiently sensitive to the unit. For example, Bobby might consider the quantity of beer he expects to be consumed at his party in terms of individual beers or six packs. These metrics should offer a one-to-one mapping: The upper and lower bounds of uncertainty expressed in six-packs ought to be one-sixth that expressed in individual beers. As a result, the interval in six packs ought to be one-sixth the width of the interval in individual beers. Yet, if people insufficiently adjust for unit, they may instead give an interval for six-packs that is greater than one-sixth as wide, representing greater uncertainty versus what is implied in the individual beer metric. More generally, we predict that when one unit is larger than the other, and thus each one-unit increase reflects a bigger change in the larger unit, people will insufficiently adjust the bounds of their prediction intervals. Thus, estimates of uncertainty will be higher for the larger unit.

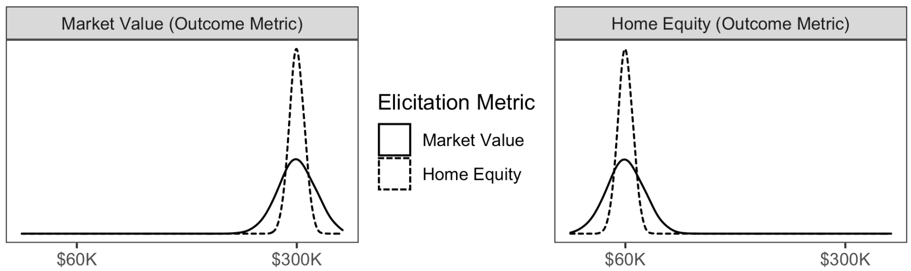
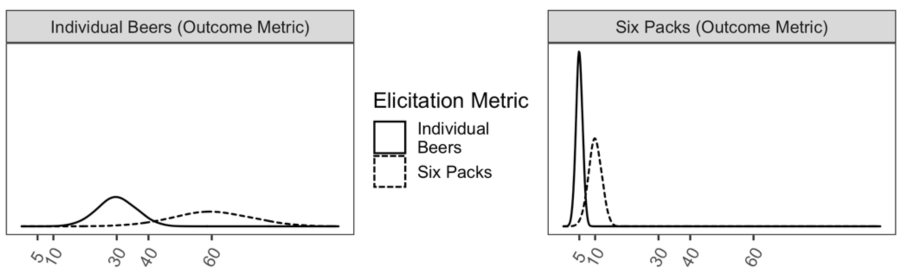
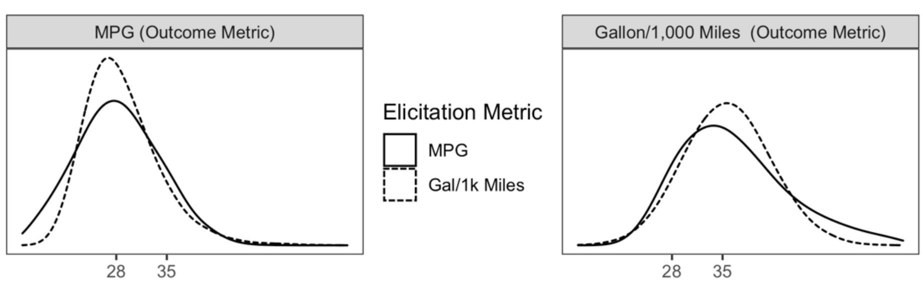
Default Distribution is Symmetric

For estimates, uncertainty can be conceptualized as a distribution of possible value. When people reason about those distributions, they tend to assume the distribution of possible values is symmetric (Flannagan et al., 1986). Moreover, across several elicitation formats (e.g., CDF, PDF, hypothetical future samples, etc.), people express beliefs that are remarkably close to a normal distribution without any requirement to do so (Winkler, 1967).

If people reason that uncertainty should be symmetric, this might lead to an *asymmetric* distribution of uncertainty after using a nonlinear transformation to convert an uncertainty estimate to an alternative metric. Likewise, if symmetric-expecting people initially reason in this alternative metric, converting would lead to asymmetric uncertainty in the original, focal metric. Previous work has shown that inversely-related metrics (e.g., miles-per-gallon and gallons-per-

1,000 miles) lead many people to make incorrect evaluations of the benefits from changes in the two metrics (Langhe & Puntoni, 2016; Larrick & Soll, 2008; Peer, 2010). For example, Claire's car hunt might lead her to reason about the distribution of fuel efficiency for gas-powered cars in the US in miles-per-gallon. Given the findings above, she will likely assume a roughly symmetric distribution. This implies a positively skewed distribution of the scaled reciprocal: gallons-per-1,000 miles. Yet if Claire were to reason about the distribution of fuel efficiency for gas-powered cars in gallons-per-1,000 miles, she would again likely assume a roughly symmetric distribution, necessitating a difference in the uncertainty between those equivalent metrics. Accordingly, we predict that when metrics are inversely related, people should give more symmetric prediction intervals in the focal metric, creating asymmetry upon converting to other metrics. Our key hypotheses are summarized in Figure 1.

Figure 1. Predicted Results of Metrics on Uncertainty Distributions

Metric Relationship	**Outcome Metric A** **Outcome Metric B**	**Prediction**
Risk Scales with Magnitude (Units Differ by Constant)	 Market Value = Home Equity + $240,000 (Mortgage Balance)	H1: Greater uncertainty for more numerous unit Tested in Experiment 1
Unit Insensitivity (Units Differ by Multiple)	 $Individual\ Beers = Six\ Packs * 6$	H2: Greater uncertainty for the larger unit (six packs) Tested in Experiment 2
Symmetry Assumption (Units Inversely Related)	 $\frac{Miles}{Gallon} = \frac{1{,}000}{\frac{Gallon}{1{,}000\ Miles}}$	H3: More symmetric uncertainty for elicited metric Tested in Experiments 3a–c

Note. The line types indicate the elicitation metric for each estimated distribution. The outcome metric columns show estimated uncertainty in each outcome metric. When the outcome metric is different from the elicitation metric, it is the transformation, shown below each pair of plots, of the distribution where the outcome and elicitation metric match in the other plot (e.g., the home equity line in the market value outcome metric is just the elicitation of home equity plus $240,000). The expected differences in the uncertainty between elicitation metrics are stated in the prediction column.

Overview of Experiments

We test three hypotheses in five experiments. Experiment 1 investigates the impact of an additive constant, asking people to estimate either revenue or profit for a business with a fixed cost structure (i.e., revenue = profit + constant). If people are influenced by a numerical magnitude when considering uncertainty, we should expect wider confidence intervals for revenue than for profit. Experiment 2 considers the impact of a multiplicative constant by asking people to estimate the number of eggs sold in a grocery store either in dozens or individual eggs (i.e., individual eggs = dozens × 12). If people insufficiently adjust to the unit when estimating values, then those estimating in dozens of eggs (e.g., the larger unit) will end up with higher real estimates and wider real intervals because people do not sufficiently adjust their nominal intervals for the units in which they report. Experiments 3a, 3b, and 3c consider the impact of ratio transformations (e.g., estimates of Mechanical Turk HIT income in minutes/dollar or cents/minute).

In all studies, people gave an estimate for the upper and lower bounds of an 80% prediction interval, often with a centrality estimate (consistent with the interval elicitation procedure of (Soll & Klayman, 2004)). People were randomly assigned to give estimates in one of two metrics. The focal dependent measure was either the width of the prediction interval (i.e., the 90th percentile - the 10th percentile) or the asymmetry of the prediction interval.

Open Practices Statement

The pre-registrations, data, code, and codebooks for all experiments can be found here: https://researchbox.org/490&PEER_REVIEW_passcode=HVUOIB.

Experiment 1: Fortune Teller

Method

Participants

Two hundred participants (78 women, median age = 36) recruited from Amazon's Mechanical Turk (AMT), a convenience sample, completed this study. The sample size was based on having 80% power to detect a moderate effect size, Cohen's d of about 0.4, a more conservative effect size than what was estimated in a conceptually similar previous study.

Design

Participants were randomly assigned to one of two metric conditions: *profit* (in which they provided a point prediction and 80% prediction interval for a business's weekly profit) or *revenue* (in which they provided a point prediction and 80% prediction interval for a business's weekly revenue).

Procedure

Participants first completed a set of training exercises describing 80% prediction intervals and how to construct them. They then learned about the specific estimation task. First, participants read basic information about a fortune telling booth at a carnival. Specifically, this fortune teller could complete three fortunes per hour for eight hours each day, for a total of 24 fortunes per day, for five days per week. Further, they read that the fortunes earned $30 per completed fortune and that renting the booth cost the fortune teller $1,500 for the week.

Next, participants made predictions about either the revenue or profit, depending on the condition, that this fortune teller would make in a five-day period. Due to the problem parameters described above, revenue was bounded between $0 and $3,600. People estimated the 90th percentile, the average, and the 10th percentile of profit or revenue.

Results

Exclusions

As pre-registered, we excluded data from participants for whom any of the three elicited or implied revenue estimates fell outside of the range [$0, $3,600] (23 people) and anyone who failed a simple attention check at the end of the study (10 additional people). This left a final sample size of 167.

Interval Width (Difference Expected)

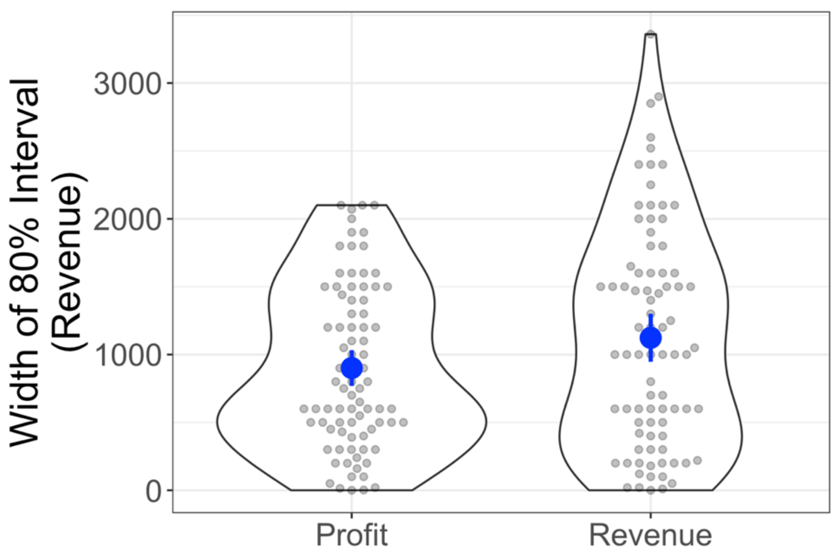
Consistent with our proposal, prediction intervals were significantly wider on average when people gave estimates of revenue (larger magnitudes) than of profit (smaller magnitudes; $M_{Revenue}$ = 1123.3, M_{Profit} = 900.8; $t(165) = 2.00$, $p = .047$, 95% CI = [1.65, 220.91], Cohen's d = 0.31).

Interval Symmetry (No Difference Expected)

Prediction interval symmetry was operationalized as a difference between distances of the central estimate to the edges: [(90th percentile - average) - (average - 10th percentile)]. Consistent with our proposal, prediction interval symmetry did not differ significantly between metrics ($M_{revenue}$ = 2.9, M_{profit} = −15.2, $t(165) = 0.57$, $p = .57$, 95% CI = [-49.54, 90.16], Cohen's d = 0.09).

In experiment 1, we found that people reported wider prediction intervals, and thus greater uncertainty, for metrics with more numerous estimates, revenue, than less numerous estimates, profit.

Figure 2. Prediction Interval Width in Revenue by Elicitation Metric



Note. Error bars represent 95% confidence intervals of the condition means.

Experiment 2: Egg Sales

Our second hypothesis is that unit insensitivity affects uncertainty. If people are insufficiently sensitive to unit, they are likely to report prediction intervals that imply greater uncertainty in real units when each individual unit represents a greater quantity (e.g., dozens of eggs) than when each individual unit represents a smaller quantity (e.g., individual eggs).

Methods

Participants

Three hundred five participants (110 women, median age = 34) recruited from AMT, a convenience sample, completed this study. The sample size was chosen to have 80% power to detect a relatively small effect size, Cohen's d of around .35.

Design

Participants were randomly assigned to one of two metric conditions: *individual* eggs (in which they provided a point prediction and 80% prediction interval for sales in individual eggs) or *dozens* of eggs (in which they provided a point prediction and 80% prediction interval for sales in dozens of eggs).

Procedure

Participants first completed a set of training exercises describing 80% prediction intervals and how to construct them. Then they learned about the specific estimation task. First, participants saw pictures of a particular Trader Joe's grocery store.

Next, participants made predictions about sales of eggs on a single Saturday about seven months away in order to evoke a sense of uncertainty. People estimated the 90th percentile, their best guess[2], and the 10th percentile of sales in individual eggs or dozens of eggs. Finally, they responded to demographic questions and an attention check item.

Results

Exclusions

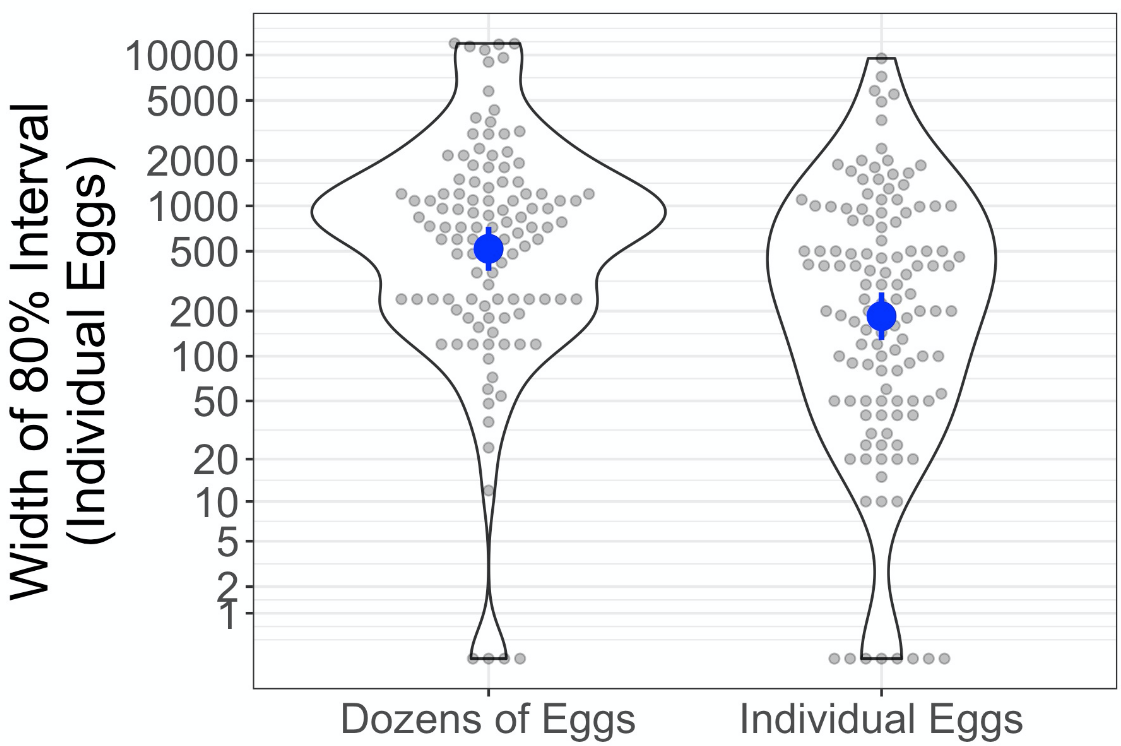
We first converted all estimates in both conditions into individual eggs. As pre-registered, we excluded data from participants for whom any of the three elicited or implied egg estimates were greater than three standard deviations away from the estimated statistic (e.g., a 90th percentile estimate for eggs sold that is greater than three standard deviations away from the mean of all other 90th percentile estimates across both conditions; 10 people) and anyone who failed a simple attention check at the end of the study (56 additional people). This left a final sample size of 239.

Width (Difference Expected)

[2] Due to a programming mistake, participants who were asked to estimate the 90th percentile first were then incorrectly asked to provide the 90th percentile estimate a second time instead of their best guess. We did not find any evidence that this moderated the relationship between the metric manipulation and the focal dependent measures (interval width: $t(235) = –0.16$, $p = .88$; interval symmetry: $t(220) = 0.82$, $p = .41$).

The distribution of interval widths was highly skewed, thus we log-transformed the widths of the prediction intervals for analysis and reported the exponentiated averages below. We find similar results using the pre-registered analysis. Consistent with our proposal, the log prediction interval width was significantly wider when people gave estimates of sales in dozens of eggs than of sales in individual eggs (M_{Dozens} = 649.9, $M_{Individual}$ = 262.8; $t(223) = 4.56$, $p <$ .001, 95% CI = [0.26, 0.65], Cohen's $d = 0.71$). Note that 12 additional people were excluded because they gave the same value for all three questions, thus our outcome statistic cannot be calculated for these people.

Figure 3. Prediction Interval Width in Individual Eggs by Elicitation Metric



Note. The outcome is the natural log of prediction interval width with the labels exponentiated back into individual eggs. Error bars represent 95% confidence intervals for condition means.

Symmetry (No Difference Expected)

Symmetry was operationalized as the ratio between distances of the central estimate to the edges, again log-transformed because of the skew: log((90th percentile - average) / (average - 10th percentile))[3]. As expected, prediction interval symmetry did not differ significantly between metrics (M_{Dozens} = 1.5, $M_{Individual}$ = 1.3, t(193) = 1.47, p = .14, 95% CI = [−0.03, 0.22], Cohen's d = 0.19). Note that 42 additional people were excluded because they gave estimates where the lower bound was equal to the average or the average was equal to the upper bound. Thus, we of the present study cannot calculate the outcome statistic for them.

Discussion

In experiment 2, we found that people give wider (real) prediction intervals when they estimate values with larger units. This is consistent with people insufficiently adjusting for the elicitation unit, where estimates that should be 12 times larger, on average, are not scaled sufficiently. This leads to wider (real) intervals in the dozens condition when all of their estimated values (10th percentile, average, and 90th percentile) are scaled up.

---

[3] Our pre-registered analysis uses a difference score [(90th percentile - average) - (average - 10th percentile) as the operationalization for the dependent variable. Results for this difference score, which also has a log transformation applied, are qualitatively similar.

Experiment 3a: Wage Rates

In experiments 3a–3c, we explore where metrics that differ by a ratio transformation may lead to inconsistent beliefs about uncertainty. If people are more likely to assume a symmetric uncertainty distribution, then ratio transformations will impact the bounds of the prediction intervals. We expect that participants will believe that uncertainty is relatively symmetric in whatever focal metric they consider. We propose this will create skewness upon converting to an alternative metric.

Methods

Participants

One hundred fifty-three participants (61 women, median age = 35) recruited from AMT completed this study. The sample size was chosen to have 80% power to detect a moderate effect size, Cohen's d of around 0.5, based on a previously conducted conceptually similar study.

Design

Participants were randomly assigned to one of two metric conditions: *cents/minute* wage rate (in which they provided a point prediction and 80% prediction interval for the wage rate for their last 100 human intelligence tasks, or HITs) or *minutes/dollar* earnings rate (in which they provided a point prediction and 80% prediction interval for earnings rate for their last 100 HITs). The selected metrics are not exact inversions. Our goal was to select metrics that lead to estimates with similar magnitudes to avoid the impacts of magnitude scaling.

Procedure

Participants first completed a set of training exercises describing 80% prediction intervals and how to construct them. Then they learned about the specific estimation task. As a practice exercise and quality check, participants were asked to construct an 80% prediction interval and point estimate for the heights of adults in the US.

Next, participants made predictions about their earnings rate in the last 100 HITs they completed. People estimated the 90^{th}, 50^{th} and 10^{th} percentiles of earnings rate in minutes spent per dollar earned or cents earned per minute. Finally, they responded to demographic questions and an attention check item.

Results

Exclusions

As pre-registered, we excluded data from participants for whom any of the three elicited or implied estimates were outside of the range [1, 100] in either metric, implying hourly wages of less than $0.60 or greater than $60 (33 people), anyone who gave non-monotonic estimates for the percentiles of heights of US adults in the training exercise (16 additional people), and anyone who failed a simple attention check (5 additional people). This left a final sample size of 99. Note that all estimates are converted to the same metric, and then the range exclusion criterion is applied.

To equate across the two metric conditions, we can convert our outcome variables for all participants into either cents/minute or minutes/dollar. All the results are qualitatively the same, and the effect sizes are very similar, regardless of which outcome metric used. We report estimates and statistical tests in terms of minutes/dollar. Results using cents/minute as the analysis metric are qualitatively similar. Both the width and symmetry outcomes were highly skewed; thus, we applied a log transformation. We exponentiated the condition average for the outcome statistics to ease interpretation.

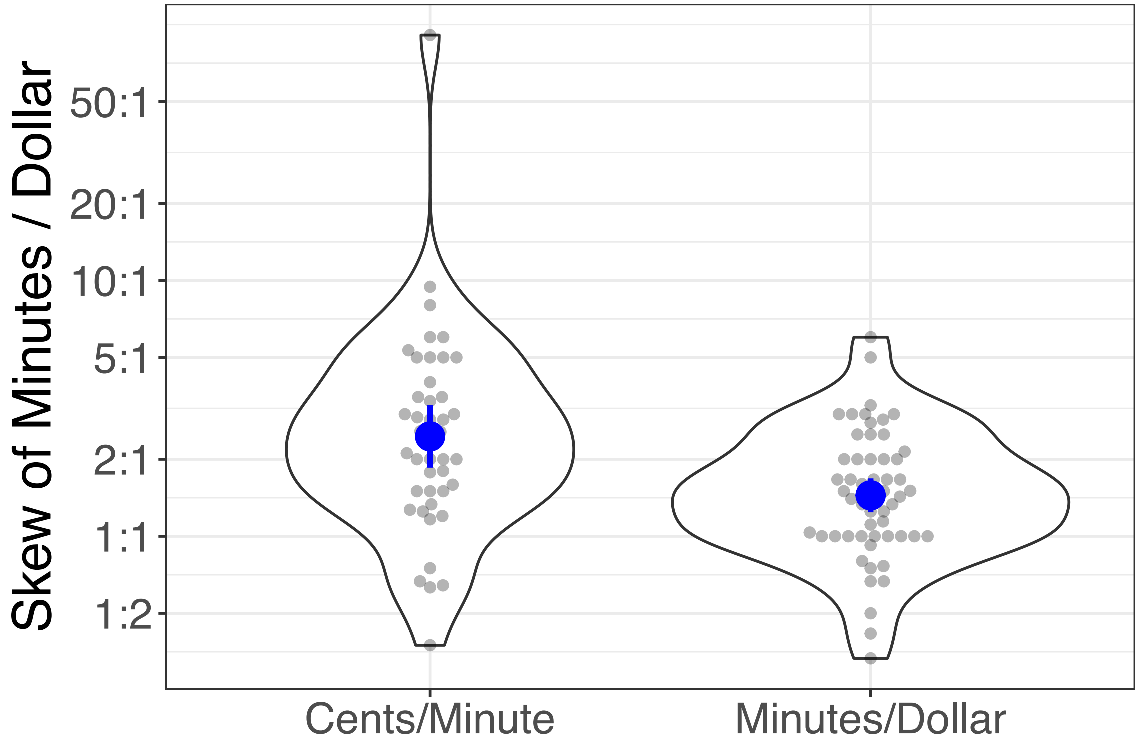
Width (No Difference Expected)

Contrary to our proposal, the log prediction interval widths were significantly different between metrics when estimating payment rates for the last 100 HITs, with wider intervals when elicited as minutes per dollar than cents per minute ($M_{cents/minute}$ = 10.6, $M_{minutes/dollar}$ = 15.1; $t(95) = 2.44$, $p = .017$, 95% CI = [0.03, 0.33], Cohen's $d = 0.5$).

Symmetry (Difference Expected)

Symmetry was operationalized as a ratio between distances of the central estimate to the edges: log((90th percentile - average) / (average - 10th percentile)). Consistent with our predictions, interval symmetry did differ significantly between metrics ($M_{cents/minute}$ = 2.5, $M_{minutes/dollar}$ = 1.4, $t(95) = -3.57$, $p < .001$, 95% CI = [-0.42, -0.12], Cohen's $d = 0.72$). Specifically, as shown in Figure 4, intervals were more positively skewed when transformed from their elicited metric relative to intervals where the elicitation metric and the outcome metric are the same.

Figure 4. Prediction Interval Skew in Minutes / Dollar by Elicitation Metric



Note. The outcome is the natural log of prediction interval skew with labels exponentiated back into the skew outcome and labeled as ratios for interpretability. Error bars represent 95% confidence intervals for condition means.

Experiment 3b: Exchange Rates

Methods

Participants

Two hundred two participants (95 women, median age = 36) recruited from AMT, a convenience sample, completed this study. The sample size was chosen to have 80% power to detect a moderate effect size, Cohen's d of around 0.5, based on a previously conducted conceptually similar study.

Design

Participants were randomly assigned to one of two metric conditions: *cents/Lira* exchange rate (in which they provided a point prediction and 80% prediction interval for the exchange rate from a Turkish Lira to US cents in two weeks) or *Lira/Dollar* exchange rate (in which they provided a point prediction and 80% prediction interval for the exchange rate from a US dollar to Turkish Lira in two weeks). We use these two metrics rather than exact inversions because the historical values of these two metrics have relatively similar magnitudes.

Procedure

Participants first completed a set of training exercises describing 80% prediction intervals and how to construct them. Then they learned about the specific estimation task where they saw historical exchange rate information over the last 11 months in the metric they were about to use for estimation.

Next, participants made predictions about the exchange rate between the Turkish Lira and the US dollar or cents in two weeks in order to evoke a sense of uncertainty. People estimated the 90^{th} percentile, their best guess, and 10^{th} percentile of the exchange rate in US cents per Turkish Lira or Turkish Lira per US dollar. Finally, they responded to demographic questions and an attention check item.

Results

Exclusions

As pre-registered, we excluded data from participants for whom any of the three estimates were outside of the range [13.915, 2.4562] when converted to Lira/Dollar or, equivalently, [40.713, 7.186] in US cents/Lira (25 people) and anyone who failed a simple attention check (8 additional people); these values were chosen as they represented a plausible range of the expectations someone could hold for the exchange rate based on historical rates[4]. This left a final sample size of 169.

To equate across the two metric conditions, we can convert our outcome variables for all participants into either US cents/Lira or Lira/US dollar. All of the results are qualitatively the same, and the effect sizes are very similar, regardless of which outcome metric is used. We report estimates and statistical tests in terms of Turkish Lira/US Dollar. All results are qualitatively similar when the US cents/Turkish Lira is used as the outcome metric. Both the width and symmetry outcomes were highly skewed. Thus, a log transformation was applied. For ease of interpretation, we exponentiated the condition average for the outcome statistics.

Width (No Expected Difference)

Consistent with our proposal, the log prediction interval widths were not significantly different between metrics when estimating the exchange rate between Turkish Lira and US

---

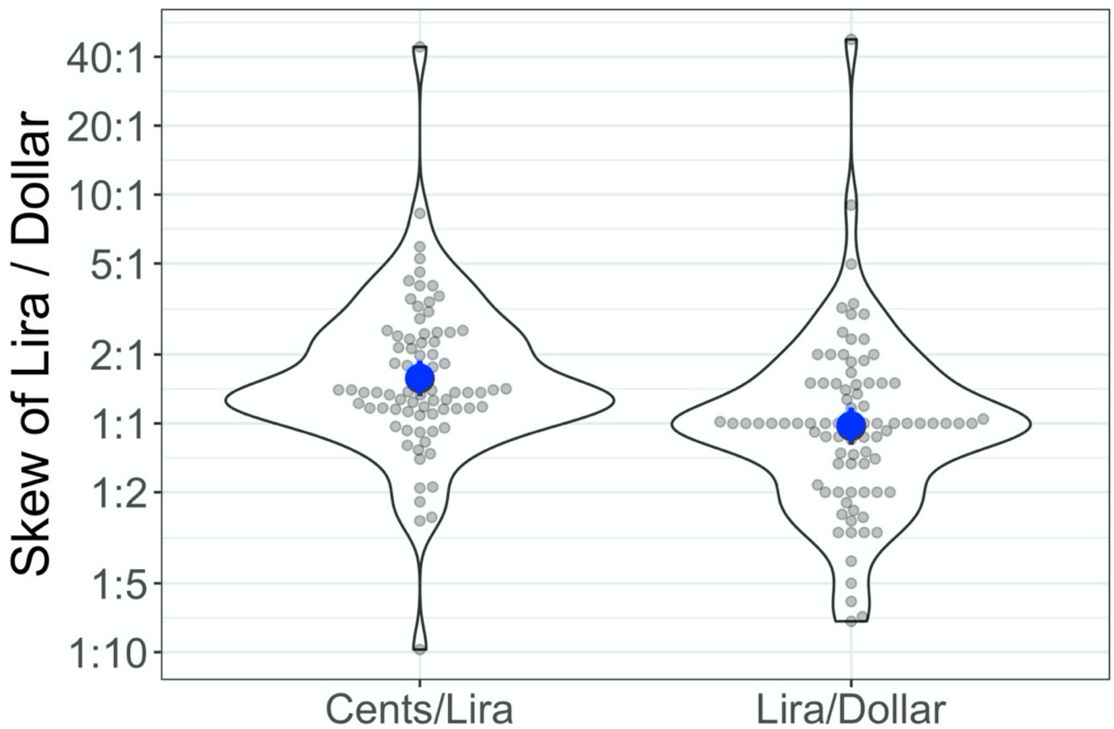
[4] See the pre-registration, Experiment 3b – AsPredicted #64058, for the precise calculations and justification.

Dollars ($M_{cents/lira}$ = 1.46, $M_{lira/dollar}$ = 1.42; $t(166) = -0.26$, p = .79, 95% CI = [-0.12, 0.09], Cohen's d = 0.04). Note that one additional person is excluded in this analysis because they put the same value for all three questions, thus our outcome statistic cannot be calculated for this person.

Symmetry (Expected Difference)

Symmetry was operationalized as a ratio between distances of the central estimate to the edges: log((90th percentile - average) / (average - 10th percentile)). Consistent with our predictions, interval symmetry did differ significantly between metrics ($M_{cents/lira}$ = 1.6, $M_{lira/dollar}$ = 1.0, $t(160) = -3.67$, $p < .001$, 95% CI = [-0.37, -0.11], Cohen's d = 0.58) as shown in Figure 5. Note that seven additional people were excluded because they gave estimates where the lower bound was equal to the average or the average was equal to the upper bound, thus we cannot calculate the outcome statistic for them in both metrics.

Figure 5. Prediction Interval Skew by Elicitation Metric



Note. The outcome is the natural log of prediction interval skew with labels exponentiated back into the skew outcome and labeled as ratios. Error bars represent 95% confidence intervals for condition means.

Experiment 3c (Gas Prices)

Methods

Participants

Three hundred ninety-eight participants (160 women, median age = 37) recruited from AMT, a convenience sample, completed this study. This sample size was chosen to have 80%

power to detect a moderate effect size, Cohen's d of around 0.5, and then increased based on simulations of the downstream outcome choice data resulting from the symmetry effects.

Design

Participants were randomly assigned to one of two metric conditions: *gallons/$40* (in which they provided a point prediction and 80% prediction interval for the distribution of gallons of regular gasoline that could be purchased with $40 in the US) or *dollars/gallon* (in which they provided a point prediction and 80% prediction interval for the distribution of prices for regular gasoline in the US).

Procedure

Participants first completed a filler task in which they calculated miles traveled based on odometer readings. Then they learned about the specific estimation task.

Next, participants made predictions about gas prices across the US in dollars/gallon or gallons/$40. People estimated the 90^{th} percentile and the 10^{th} percentile of the gas prices in the US. They were given the average price in their elicitation metric (dollars/gallon or gallons/$40). People then gave estimates of stopping ratings for hypothetical gas prices for a road trip vignette. People read a scenario about going on a hypothetical road trip where they had to decide whether or not to stop for gas or continue until the next stop. They evaluated five different hypothetical prices in their elicitation metric on a five-item measuring likelihood of stopping (*Definitely stop now*, …, *Definitely continue driving*)[5]. Finally, they responded to demographic questions.

Results

Exclusions

As pre-registered, we excluded data from participants for whom any of the three estimates were outside of the range [6.00, 40] when converted to gallons/$40 or [1.00, 6.66] in dollars/gallon (104 people); these values were selected as they represent a plausible range for gas prices someone could hold given historical rates and geographic variation[6]. This left a final sample size of 294.

We could convert our outcome variables for all participants into gallons/$40 or dollars/gallon to equate across the two metric conditions. All of the results where we expected to

---

[5] We were testing the impacts of metric and the tendency to give more symmetric uncertainty estimates on a downstream outcome, search. This effect would only exist if people give the same average uncertainty, as measured by prediction interval width, but also a tendency to give symmetric intervals regardless of metric. Since we observe a difference in prediction interval width, our test of downstream consequences is confounded. Search was operationalized as the difference in stopping rating from the middle outcome ($3.10/gallon in the price/gallon condition and 12.9 gallons/$40 in the gallons/$40 condition) minus the stopping rating for the best outcome seen ($2.20/gallon in the price/gallon condition and 18.2 gallons/$40 in the gallons/$40 condition). We do not find a significant difference between stopping ratings between metrics ($M_{price/gallon}$ = 1.24, $M_{gallons/\$40}$ = 1, $t(292) = -1.24$, p = .22, 95% CI = [−0.25, 0.06], Cohen's d = 0.14).

[6] See the pre-registration, Experiment 3c – AsPredicted #73452.pdf, for details about the calculation

see differences are qualitatively the same, and the effect sizes are very similar, regardless of which outcome metric was used. We report estimates and statistical tests in terms of price/gallon. All differences are qualitatively similar when using gallons/$40 as the outcome metric. Both the width and symmetry outcomes were highly skewed, thus we applied a log transformation. For ease of interpretation, we exponentiated the condition average for the outcome statistics.

Width (No Expected Difference)

Contrary to our proposal, the log of prediction interval widths was significantly different between metrics when estimating gas prices between price/gallon and gallons/$40 ($M_{price/gallon}$ = 0.9, $M_{gallons/\$40}$ = 1.1; $t(290) = -2.17$, $p = .031$, 95% CI = [-0.22, -0.01], Cohen's $d = 0.25$).

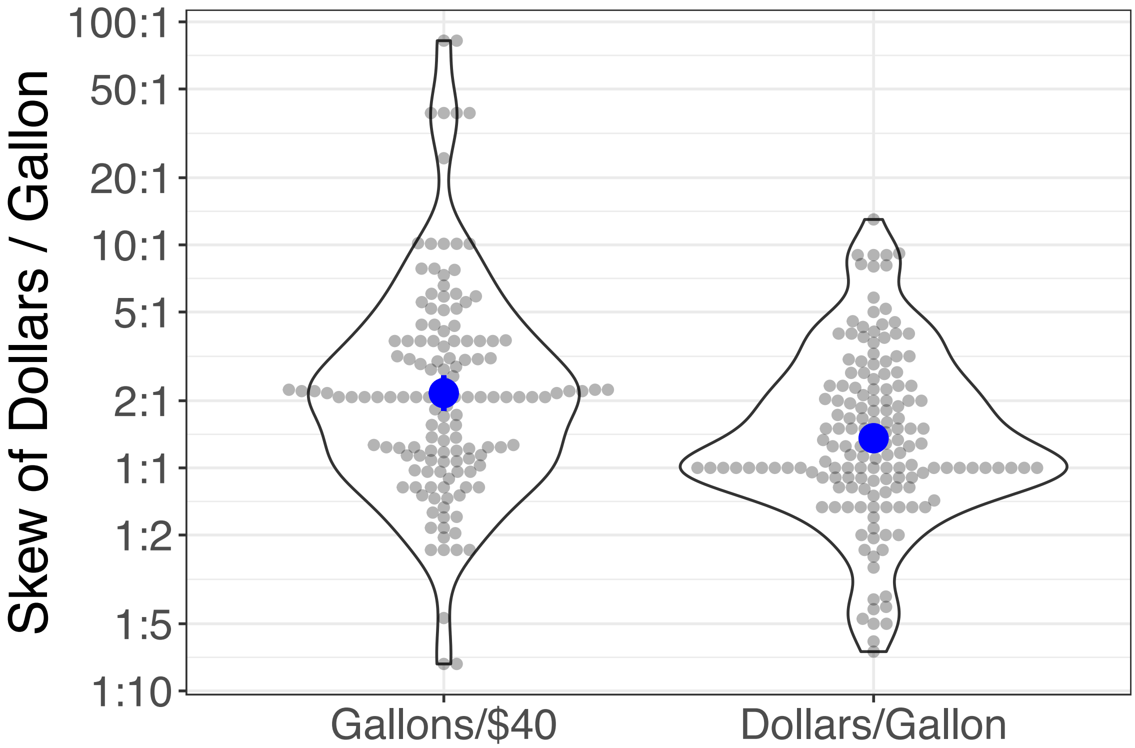
Symmetry (Expected Difference)

Symmetry was operationalized as a ratio between distances of the central estimate to the edges: log((90th percentile - average) / (average - 10th percentile)). Consistent with our predictions, interval symmetry did differ significantly between metrics ($M_{price/gallon}$ = 1.36, $M_{gallons/\$40}$ = 2.2, $t(290) = -4.04$, $p < .001$, 95% CI = [-0.35, -0.12], Cohen's $d = 0.47$). Intervals were more positively skewed when in the transformed metric from their elicited metric relative to intervals where the elicitation metric and the outcome metric are the same.

Discussion

In Experiments 3a–3c, we test whether people tend to give prediction intervals that are more symmetric when the elicitation and outcome metrics are the same than when they are different. We see this result for all three experiments, suggesting that people act as though they are assuming relatively symmetric distributions regardless of which metric they are using for estimation. When their estimates are inverted, the intervals become less symmetric as a result of the symmetry prior combined with this mathematical transformation.

Figure 6. Prediction Interval Skew for Dollars/Gallon by Elicitation Metric



Note. The outcome is the natural log of prediction interval skew with labels exponentiated back into the skew outcome and labeled as ratios. Error bars represent 95% confidence intervals for condition means.

General Discussion

We examined three important means by which simple unit transformations impact uncertainty: addition of a constant, multiplication by a constant, and inversion of a ratio. We observed more complex metric relationships likely produce a combination of the findings and are ripe for examination as the literature matures. Each of the current studies elicited prediction intervals and centrality estimates. The advances in the elicitation of full distributions, such as distribution builders (André; et al., 2021; D. G. Goldstein et al., 2008) or SPIES (Haran et al., 2010), would allow for additional exploration of other biases and more precise investigations. The complex set of numerical processing biases may require more precise measures to detect all factors impacting uncertainty.

There are two unpredicted differences in our interval width results. In the HIT payment context, prediction intervals were narrower in cents/minute compared to minutes/dollar; in the gas purchasing context, prediction intervals were narrower in price/gallon compared to gallons/$40. We speculate that cents/minute (vs. minutes/dollar) and price/gallon (vs. gallons/$40) are more familiar metrics, plausibly leading to less uncertain intervals.

Lastly, we expect that people who regularly convert between, or work with, multiple metrics for the same construct would be less likely to exhibit these inconsistencies. For example, people who can broaden their decision frames can reduce biases (Larrick, 2004). In this case, people may simply need to consider the alternative metric to form a belief about the focal construct, which is consistent with elicitations between the two metrics. For example, explicitly encouraging people to confront both metrics may reduce the impact of the metric. Given the strength of numerical processing biases (Thomas & Morwitz, 2009), a decision aid may be needed to facilitate the realignment between estimates in the two metrics.

CHAPTER 2

RISK LOVING, CONFUSED, or MISLED?
The MYSTERY of RISK SEEKING CHOICES

Introduction

Wendy is driving home from work. She has two familiar routes home: the freeway and surface streets. The freeway could be free of traffic while she makes her way home, or it could become congested. The route over surface streets can also get backed up with traffic or remain relatively smooth. On average, both routes take the same amount of time. In addition to travel time, Wendy also chooses a route based on the number of aggressive drivers she has to deal with on her way home since she particularly dislikes them. Based on her past experiences, if the freeway gets congested, she will likely encounter at least one aggressive driver. However, if it stays clear, she probably will not encounter any aggressive drivers. For the surface streets route, if it stays clear, she will likely encounter at least one aggressive driver. However, if the surface streets have heavy traffic, she is much less likely to encounter an aggressive driver.

This particular situation represents a choice between two options, each with two uncertain attributes. These options vary in the relationship between the uncertainty of the attributes. When Wendy takes the freeway, she will likely get two good outcomes (minimal traffic and fewer aggressive drivers) or two bad outcomes (lots of traffic and many aggressive drivers). Conversely, the surface streets option has the exact opposite relationship between the two attributes. On the surface streets, if there is minimal traffic, she will likely encounter aggressive drivers, but if there is lots of traffic, she is unlikely to encounter any aggressive drivers.

This is one of the many decisions involving multi-attribute outcomes that people encounter daily. Other examples range from a high-stakes medical treatment decision—involving multiple options with varying cost, quality of life, and life expectancy possibilities—to the quotidian such as a choice between two stores when making a weekend pit stop for snacks and drinks. In such decisions, the options may differ in the relationships between the uncertainty for each attribute. For instance, considering the medical decision, one treatment may have outcomes that are positively correlated (e.g., people tend to have a high quality of life and longer life expectancies if the treatment goes well, or lower quality of life and shorter life expectancies if the treatment fails); another treatment option may have outcomes that are negatively correlated (e.g., people tend to have a high quality of life with lower life expectancies or lower quality of life with longer life expectancies).

The current research attempts to understand people's preferences regarding the relationship between multi-attribute outcomes. Several theories make predictions about people's choices in similar situations. Some of these theories focus on how people encode information about the options while others focus on how risk preferences, in a situation of perfect information, predict choices. However, these current theories appear insufficient to explain our most common finding, a preference for a positive correlation between attributes. Moreover, research on risky choices has focused on single-attribute outcomes, with little attention paid to multi-attribute outcomes and, specifically, how the relationship between the risk of the two attributes impacts choice. Thus, by testing current risk and uncertainty theories against more complex decision structures, we hope to better understand the generalizability of these theories.

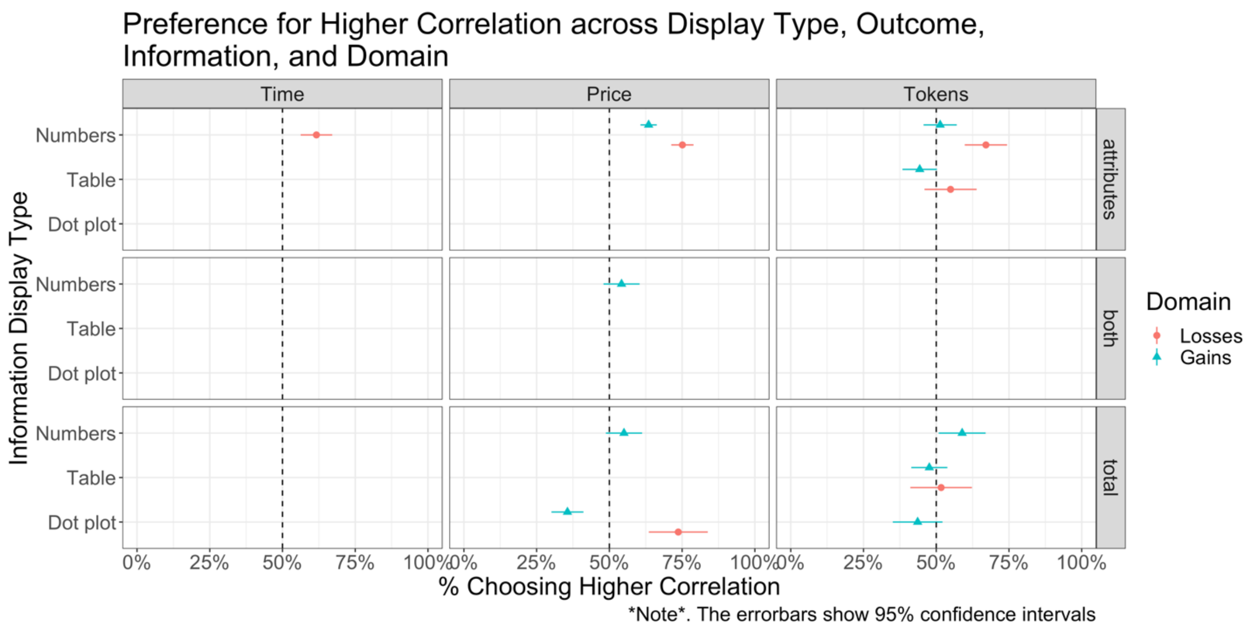
Research Question. This research asks whether people prefer uncertain options that are likely to produce (a) either good outcomes or bad outcomes, or prefer (b) a mixed bag of outcomes. More technically, we ask whether people prefer more correlated or less correlated uncertainties in multi-attribute options. The current theories of risk and uncertainty make some predictions about how people make these kinds of decisions. The most common risky choice model (Neumann & Morgenstern, 1953) predicts that people who are risk-averse to individual attributes will also be risk-averse to the joint outcome. This means that when choosing a commuter route, all else equal, people who are risk-averse to both travel time and likelihood of encountering aggressive drivers should prefer a lower correlation between the two attributes. A lower correlation reduces the variance of the overall outcome. In the special case where each attribute's averages and standard deviations are the same between options, the lower correlation should be strictly better. In the case described, the lower correlation option would be second-order stochastic dominant over the higher correlation option.

Several alternative theories may predict different choice patterns. All of these alternative theories require that people systematically misrepresent the options when making choices. People may use certain heuristics that may also lead to systematic inconsistencies in preferences. For example, the majority rule heuristic (Zhang et al., 2006) is a decision rule by which people keep track of the number of times a particular option does better. Using this heuristic means the preferred option is the option that did better in the most round-wise comparisons, often done at the attribute level. This heuristic implies that people ignore the magnitude of the differences in each comparison when tallying how often an option does better. Returning to the commuting example, Wendy may pair up instances when she took the freeway and surface streets. For each pair, she would just keep track of how many times the freeway resulted in a better overall outcome and not keep track of the magnitude of the differences in travel time and aggressive drivers encountered. We test several heuristics, risk preferences, and uncertainty communication presentations as potential explanations for the surprising choice results. Ultimately, none of these explanations can account for the observed data.

Overview of Results

We investigated people's choice between two uncertain options in 16 studies. We find a surprising preference for positive correlations for price attributes. When people make choices in the domain of gains, they are likely to choose the positive correlation. This is surprising because, if most people are risk-averse in the domain of gains with moderately large probabilities, we would expect people to choose the option with negatively correlated attributes since this option minimizes the chances of an extremely good or bad outcome. However, we find that a significant majority of people prefer the positive correlation.

Figure 7. Preference for Higher Correlation Across Display Type, Outcome, Information, and Domain



Literature Predictions

Risk Aversion

Risky prospects have been studied extensively, primarily focusing on single-attribute outcomes. The extensive literature on risky choices and risk preferences shows that people tend to be risk-averse under a broad range of conditions. When the probability of an outcome is moderate or large, and the outcomes are in the domain of gains, people are generally risk-averse (Abdellaoui et al., 2007; Kahneman & Tversky, 1979). This means that when they are offered a risky prospect, they will accept some money that is less than the expected value of the risky prospect (Pratt, 1964). Risk aversion also implies specific choice patterns when comparing two risky options. For instance, a risk-averse individual would prefer a gamble with the same expected value but a lower standard deviation than the alternative. This case is often called second-order stochastic dominance (Levy, 1992). This research will primarily examine choices among uncertain options where each option has exactly two attributes, and one option is second-order stochastic dominant over the other. Between these choices, risk-averse people, the majority of the population, will choose the option with less variable outcomes but the same expected value.

For the present research, the definitions of risk and uncertainty will be the same as the definitions from Knight (1921). Risk is defined as a situation where the decision maker knows the full outcome space and the exact probability associated with every outcome. Uncertainty is when people do not know the full outcome space, the exact probabilities associated with each outcome, or some combination of those two things. In this research, people do not know the full outcome space and have only approximate estimates of the likelihood of the outcomes they are aware of. Here, we only observe choices under uncertainty, not risky choices. Unless there are exceptional circumstances leading people to savor uncertainty (Golman et al., 2021), we would not expect a preference for options with greater variance. Commonly observed risk-averse preferences would predict that people thus would prefer options with negatively correlated attributes.

Multi-Attribute Risk Preferences

When someone is risk-averse over individual attributes, only innocuous assumptions are required to predict that they will be risk-averse over the joint outcome space of multiple attributes (Epstein & Tanny, 1980). If these assumptions are met, this model predicts that people who are risk-averse for single attribute choices in a certain domain will also be risk-averse over joint outcomes composed of two risky outcomes in the same domain (e.g., someone who is risk-averse for a single monetary gamble should be risk-averse over pairs of gambles). This reaffirms the predicted choice from the single attribute scenario. Risk-averse people will choose the lower variance option when the means are equal in expectation, even if the outcome involves two attributes in the same domain.

Possible Mechanisms

I attempt to explain the surprising trend that people tend to select the positively correlated option by examining various mediators: the majority rule heuristic, the peak-end rule heuristic, the weighting of small probabilities, and risk preferences, as well as various moderators: simultaneous versus sequential learning of information and the use of risk preferences in guiding the decision. Currently, none of the mediators or moderators fully explain the gap between what is observed and what the expected choices would be.

Majority Rule Heuristic

The majority rule heuristic (Zhang et al., 2006) happens when people compare two or more multi-attribute options attribute by attribute and keep a running total of the number of times each option did better, regardless of the magnitude of the differences. Imagine someone were to choose between two vacation spots, such as Malta and Corsica, France, with the following attributes: price, average reviews of accommodations on Yelp, and weather. Using the majority rule heuristic to select the preferred option means a person scores each option by the number of times it has the better attribute value. They would look to see whether Malta or Corsica had a better price, which option had better reviews, and which option had better weather. The chosen option, following this heuristic, would be whichever option did better on at least two of the three attributes, regardless of the magnitude of the differences for each attribute comparison. This is a non-compensatory heuristic that has the potential to lead to choice results not predicted by expected utility theory.

The literature reviewed in this study generally present information about the attribute outcomes through animated displays, such as experience sampling presentation methods (de Palma et al., 2014). Depending on how the information is presented, people could apply this heuristic in multiple ways to the studied paradigms. The most common presentation method used in this paper simultaneously presents historical pricing information for both options. A straightforward application of the majority rule heuristic would be to compare the total prices between each option for every pair of prices shown. After seeing the full sequence, people would simply pick the option with the lowest price most times across all the pairs of prices.

If people were to see information in a histogram format, applying the majority rule would work differently. In this case, one way to apply the rule would be to take the best price from each option and determine which option has the lower price. Next, one would take the second-best prices from each option and compare them, continuing until getting to the highest prices for each option. In the end, a person using this heuristic would select the option with the lowest price most times in this paired comparison process.

Peak-End Rule

The peak-end rule is a well-studied decision strategy for approximating the value people place on events experienced over time. From medical procedures (Kahneman et al., 1993), to advertisements (Baumgartner et al., 1997), to life satisfaction ratings (Schkade & Kahneman, 1998), to coffee purchases (De Maeyer & Estelami, 2012), this heuristic predicts choices surprisingly well using only the evaluations of the most intense point (i.e., the peak) and the end. In the case of advertising, the average of people's ratings for the most intense moment and the last moment of an advertisement is a very good predictor of the retrospective evaluation they give.

Furthermore, in the case where people must choose between two stores, each with a different correlation between their prices, the heuristic has two different ways it may apply. When only the correlations differ between the two options, the end of the experience is just as likely to favor the positively versus negatively correlated outcome option. Thus, evaluations based on the end value would predict that differences in price correlations are irrelevant. Positively correlated prices will have more extreme best and worst values compared to an option with negatively correlated prices. This means the option with the higher correlation between prices will have the most extreme bad outcome. Moreover, if people have a reference price somewhat close to the average, then prices that are equally far from this reference price would get differential weight since the higher prices could be considered losses (Kalyanaram & Little, 1994). Therefore, this suggests that people will consider the highest price the peak. Since the end price should be even in expectation, the highest price dictates the evaluation, predicting that people will prefer the negatively correlated price option.

When switching from attributes where people are losing a resource (e.g., prices where people are paying money) to an attribute where people are gaining a resource (e.g., gaining tokens that translate into money), results in the reverse prediction. People would be expected to not have a reference point within the range of observed values for attributes where people earn tokens. This means that the most intense value for all options would be the highest value. People will also see much higher total values with the positively correlated attributes and thus prefer the positively correlated option. As before, the last observed values are likely to be evenly split in terms of favoring the positive or negative correlation option. If the tokens switch framing, where the token values now represent losses from an endowment of initial tokens, loss-averse people following the peak-end heuristic would evaluate the positively correlated option as worse. Under cumulative prospect theory (Tversky & Kahneman, 1992), people would also become risk-seeking in the domain of losses, pushing them to favor the positively correlated options.

Probability Weighting

Probability weighting for continuous, uncertain outcomes is a relatively novel area for applying cumulative prospect theory. There has been considerable debate on whether people tend to be underweight vs. overweight for low probability events depending on whether the probabilities are learned through experience or description (de Palma et al., 2014). Thus, I will consider the predictions for both underweighting and overweighting. In the case of continuous outcome spaces, all outcomes would likely be considered low probabilities. If people round the outcomes, they could group the total and individual attribute outcomes together when learning about the outcome distribution. The resulting true probabilities could be moderately sized, depending on the exact way in which grouping occurred. When transforming observed probabilities into probability weights, people would use the true probability space from 0% to about 30%. Based on elicitations of probability weighting curves (Palma et al., 2014; Wu & Gonzalez, 1996), relative deviations are expected to occur within the low probability space, even if all the probabilities are considered small. If people are overweighting, a 5% probability is overweighted more than a 10% probability, and if they are underweighting, a 5% probability is underweighted more than a 10% probability. Next, I will turn to the predictions for expected utility theory and prospect theory under the cases of over and underweighting.

Expected Utility and Overweighting. People with preferences that follow expected utility will be risk-averse. Any overweighting of low probabilities, even if it is just a relative overweighting (e.g., 1% is overweighted more than 5%), people would view the positively correlated uncertainty as even riskier, leading to a strong preference for the negative correlation. This is because the most extreme outcomes for the positively correlated option are relatively overweighted, making the most extreme outcomes seem more probable than they are.

Expected Utility and Underweighting. If people do relative underweighting of low probabilities (e.g., 1% chance is underweighted more than a 5% chance), the outcome distributions would be valued as more similar. This means that people would be relatively indifferent between the two options, resulting in something closer to a 50/50 choice share between higher and lower correlated uncertainty options. This is because the most extreme outcomes for the positively correlated option are relatively underweighted relative to all other outcomes across the two options. Because these extreme outcomes are underweighted, the two options appear to be more comparable.

Prospect Theory and Overweighting. Prospect Theory predicts people will be risk-averse in the domain of gains. With relative overweighting of low probabilities, people would view the more correlated uncertainty option as riskier and thus have a preference for the lower correlation between uncertainty. In the domain of losses, people are expected to be risk-seeking on average. This means that overweighting small probabilities would lead people to like the more correlated option since it appears riskier.

Prospect Theory and Underweighting. When people do relative underweighting of small probabilities, the evaluations of the two distributions are expected to be more similar since the more extreme outcomes in the higher correlation option have the lowest probability of occurring. In the domain of gains, prospect theory predicts that people will be risk-averse; however, the underweighting will make the distributions so similar that the expected strength of preference will be relatively weak. In the domain of losses, prospect theory predictions people will be risk-seeking; however, the underweighting of probabilities will lead to similar evaluations for both options. This leads to a prediction of relative indifference between the options with a slight preference for the positively correlated option.

Moderators of choice

Use of Risk Preferences

One strong assumption of this research is that when people consider this a choice under uncertainty, their choices will follow one of the risk or uncertainty decision-making models. If people use a different decision-making framework, then all the predictions based on expected utility theory (Neumann & Morgenstern, 1953) or cumulative prospect theory (Tversky & Kahneman, 1992) may not hold. One example of when risk and uncertainty decision models fail is when people make decisions based on minimal outcomes. In many instances, people get less risk-averse (Prelec & Loewenstein, 1991), often called the peanuts effect. This may be due to feelings of disappointment driving the decision more than the curvature of the utility function (Weber & Chapman, 2005).

A second assumption is that choices under uncertainty will be no more risk-seeking than choices in a risky choice setting. Ambiguity aversion, the tendency for people to prefer options with less uncertainty to ones with more uncertainty (Fox & Tversky, 1995), has an additional prediction about how risk and uncertainty preferences relate to each other. In choice contexts, uncertainty aversion would predict that when the expectation of the risk parameters for the uncertain option is the same as the risky choice, people would prefer the risky option.

For this research, a translation of how this ambiguity aversion applies to choices between two ambiguous options is needed. The risk and uncertainty options in the ambiguity aversion work are aligned by having the expectation of the uncertainty result in the same mean and variance as the risky option. This is often done in contexts where enough information is provided to estimate an expectation over the possible states of the uncertain distribution. For example, many studies use urns with a known number of balls, say 100, but an unknown mixture between two colors: red and blue. The risky option, which is the alternative to the uncertain option, explicitly states that half the 100 balls are red and half are blue. In this case, the actual distribution of colors in the uncertain urn could be very far from a 50/50 split. However, with no information about the distribution, the distribution expectation would be an even split. This means that there is more variance in the outcome of the uncertain choice than in the risky choice. If this additional variability is viewed negatively, people should choose even more conservative outcomes when there is uncertainty than purely focusing on risk. As a result, when comparing elicited risk preferences directly to choices under uncertainty, people would be expected to be even more conservative in the choices under uncertainty compared to a risk preference elicitation measure.

Simultaneous vs Sequential Learning

There is substantial research in learning and memory to understand how people learn things and then hold them in working memory. Researchers have tested whether the presentation format of information impacts peoples' ability to learn information. Most of this research looks at outcomes such as word lists or sequences of images (Barrouillet et al., 2011). When narrowly testing simultaneous versus sequential presentation of stimuli, unfamiliar symbols, in this case, forgetting rates, are similar when the time is constant across presentation formats (Ricker & Cowan, 2014). There is additional research that does not explicitly test the simultaneous versus sequential presentation of information but does offer some insight into the effects of these two methods. Hypothetical outcome plots, when outcomes are drawn from distribution and then sequentially shown using animation, have been compared to static distributions and point estimates with error bars when communicating distributions of random variables (Hullman et al., 2015). In this research, people's estimates of various statistics of the communicated distribution were similar across the three presentation formats. This suggests that sequential learning of the distribution shape may lead to a similar understanding as a static or simultaneous learning presentation format.

Furthermore, while the current evidence suggests that basic recall and distributional estimates are similar, the decision-making process includes additional steps. People must make judgments on their perceptions of risk, which sometimes vary in non-normative ways. For instance, in Cornil et al. (2019), the researchers found that positively correlated investments appeared less risky to people compared to negatively correlated investments, which is incorrect. Negatively correlated investments have the benefit of diversification, which can greatly reduce the variability in the overall balance without any loss in the expected gains. Other research on correlated investments has found different display formats (Laudenbach et al., in press) and promotions of risk-taking (Reinholtz et al., 2021) cause people to diversify more. Different presentation methods may impact other steps in the judgment of uncertain outcomes, specifically how people translate outcomes to risk. Given the current research on presentation methods, which primarily focuses on measuring what information people have stored in working memory rather than their evaluation of that information, it is difficult to make a strong prediction about uncertain choices based on uncertainty formats.

Outline of Studies

Experiment 1 had a surprising finding outside of the pre-registered analyses. Both conditions substantially favored the positive correlation between attribute uncertainty, even though this is most consistent with risk-seeking preferences. To better understand the origin of this preference, the next experiment examined whether the majority rule heuristic would predict the preferences observed for the positive correlation. Here, the majority rule heuristic failed to predict the preferences observed in the study results. A second decision heuristic, the peak-end rule, was the focus of the next study. This heuristic failed to predict the preferences for positive versus negative correlations, and the study provided evidence against the moderator of risk preferences by measuring those directly. Experiment 4 uses attributes of time to expand the generalizability and finds a similar pattern of results: most participants choose the positive correlation of attribute uncertainty over a negative correlation. Finally, drawing from other research on presentation methods of correlated uncertainty, mainly in the investing context, I tested both simplified methods and one previously used alternative.

Detailed Results

Study 1

Research of decisions under uncertainty studied in lab settings have largely considered choices with single-attribute outcomes. I extend this research to uncertain multi-attribute outcomes. Here, CPT would predict that people will prefer positively correlated attributes, the option consistent with risk-seeking, in the domain of losses. Conversely, CPT predicts that people will prefer *negatively* correlated attributes, the option consistent with risk-aversion, in the domain of gains. I wanted to test whether the data in fact supported these predictions. Thus, in Study 1, I tested this by having people make choices in both the loss and gain domain. CPT predicts that people will be more likely to select the option with positively correlated attributes in the domain of losses than gains. Stronger still, CPT predicts that the majority of people will prefer an option with negatively correlated attributes in the domain of gains, and that the majority of people will prefer an option with positively correlated attributes in the domain of losses.

Method

Participants

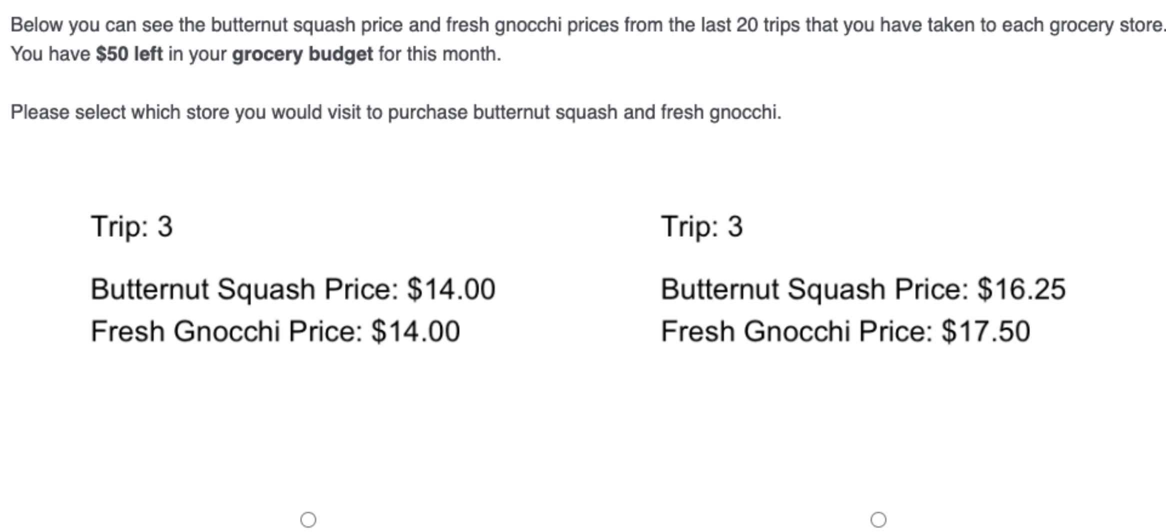
One hundred fifty-one participants (72 women, 1 other, median age = 38) were recruited from Amazon Mechanical Turk (AMT) and completed the study.

Procedure

After an introduction and attention screener, participants were told that they would be making a choice between two grocery stores with a fixed shopping list. Both stores were certain to have the items and the quantities needed, so I asked participants to imagine that the only difference between the stores was the prices they had seen when shopping at each store in the past. As a training example, participants were shown a list of two grocery items and their associated prices. All prices presented were already rescaled for the desired quantities. I then asked participants a set of three comprehension questions about the list, requiring them to answer all questions correctly before proceeding to the main task.

In the main task, participants were given a two-item grocery list and a grocery budget (e.g., \$50). Importantly, this was described as a grocery budget, not their checking account balance, so they would have to dip into another “mental account” if prices exceeded the grocery budget on a trip to the store and not go without these grocery items (Thaler, 1985). As depicted in Figure 8, participants were then shown prices for both items at each of the two stores, intended to represent the prices of these items when visiting each store on a given day (i.e., a “trip to the grocery store”).

Figure 8. Screenshot of Experience Sampling Display of Price Information



After two seconds of observing the prices of the two items on a given trip, new prices of the same items on a different trip appeared for two seconds. This continued for a total of 20 such trips. Participants could select their preferred store for the given shopping list at any point but could only proceed past this page after observing all 20 trips and making a selection. The only difference between the two stores was whether the prices of the two items were positively or negatively correlated between trips (r = .8 or r = -.8; counterbalanced between subjects whether the positively correlated prices were on the left or right).[7]

Grocery budget was a within-subjects manipulation. Participants repeated their store decision—with the same prices on each of the same 20 trips—but with a different budget constraint ($27; order of budgets was counterbalanced between subjects). The average total cost of the grocery list across the 20 trips was the same ($31) across stores and budgets, so participants had a budget surplus for the $50 budget and a deficit for the $27 budget (which would likely require them to use money mentally budget for other purposes). While mean total prices across all 20 trips were the same for both stores, the variance of total prices was constructed such that, for the negatively correlated store, total prices always exceeded $27. However, for the positively correlated store, there were a few outlying trips in which the total price was $27 or less. In this way, for the $27 budget, participants were choosing between a small chance of avoiding a loss (in choosing the positively correlated store) versus a certain loss (in choosing the negatively correlated store).

After making both decisions, participants answered demographic questions and were paid for their time. I was primarily interested in participants' preferences for the positively versus

[7] For ease, I refer to these as the "positively correlated" and "negatively correlated" options. More precisely, I mean the option for which the attributes are positively correlated and the option for which the attributes are negatively correlated.

negatively correlated pricing schemes, both in the domain of gains (i.e., the $50 budget) and losses (i.e., the $27 budget).

Results

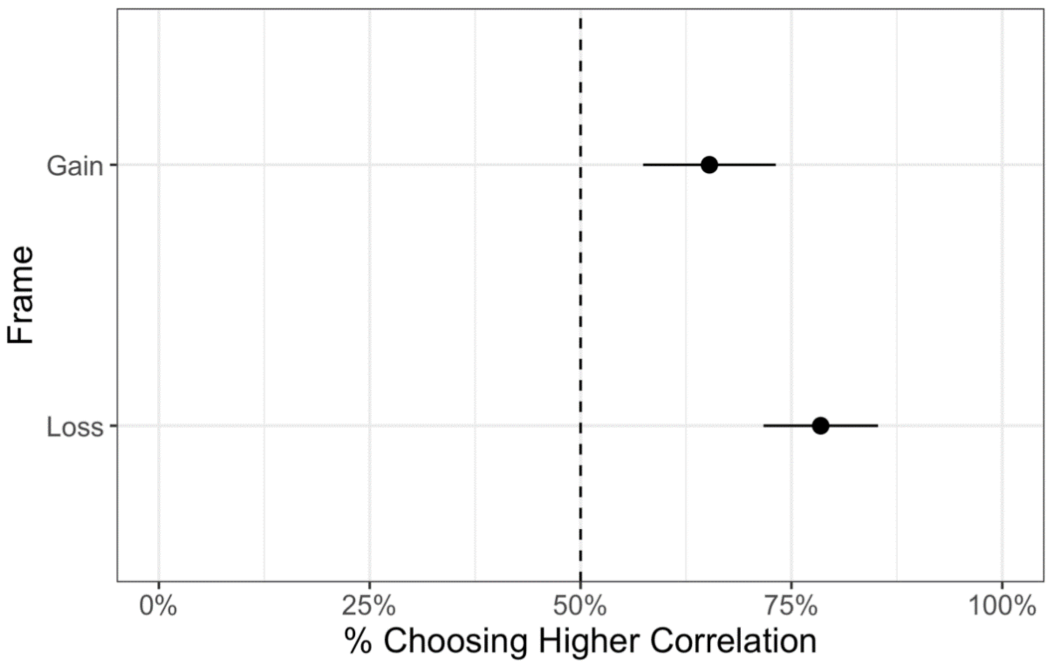
Exclusions

Based on the pre-registered exclusions, I excluded participants who gave the wrong purpose of the study. Seven gave the wrong response after reading a brief study overview.

Choice

The proportion of participants who selected the positive correlation is surprisingly high, particularly in the gain frame. A post-hoc comparison with a 50/50 split in choice share showed that for the gain frame, the proportion of participants selecting the higher attribute correlation option is significantly greater than 50% ($t(143) = 3.84$, $b = 0.15$, $p < .001$). In the loss frame, the same directional result was seen; significantly, more than 50% of participants chose the positively correlated price option ($t(143) = 8.28$, $b = 0.28$, $p < .001$). Prospect theory predicts that people should choose the riskier option more often in the loss frame relative to the gain frame. On average, when participants chose in the loss frame compared to the gain frame, they chose the higher correlation more often ($t(143) = -2.56$, $b = -0.13$, $p = .011$).

Figure 9. Preference for higher correlation by frame



Note. The errorbars show 95% confidence intervals of cell means

Discussion

Prospect theory predicts that people in the domain of gains are risk-averse. Based on the structure of the two options, the option with the lower correlation would be the preferred option for risk-averse people. This is because both options have the same average outcome, but the option with a higher attribute correlation has more variable outcomes, making it riskier. The majority of people selected the higher correlation in both domains. A risk-seeking individual might prefer the positive correlation over the negative one, but this would be an unexpected preference for most participants. In addition, cumulative prospect theory would predict that people choosing in a loss frame would be risk seeking and thus more likely to choose the higher correlation option. It was found that the relative choice share and absolute choice share for this frame are directionally consistent with this prediction. If participants were risk-averse in the gain domain, the relative difference in choice shares between frames would need to be large enough so that most of them selected the negatively correlated distribution in the gain frame, which was not found. The choices participants make in the loss frame provide some confidence that risk preferences are part of the decision for at least part of the sample population.

These decisions are quite difficult given the working memory demands placed on people. Thus, there are several plausible heuristics that people may have used to simplify the decision and could have led to this strong preference for positively correlated attribute outcomes. The next two studies will examine whether the majority rule heuristic and the peak-end rule can explain the choice results, particularly in the domain of gains.

Study 2

Study 1 found that most participants preferred options with positively correlated attributes to options with negatively correlated attributes. This was true even in the domain of gains, contrary to the predictions of prevailing models of decisions under uncertainty. One possible alternative explanation to the findings from Study 1—unrelated to a preference for positively correlated attributes—is that people made a series of round-wise comparisons between the two stores, only tracking which store offered the cheapest grocery bill for each trip. It is possible that as an accidental artifact of the random process used to generate prices, such round-wise comparisons, coupled with a majority-rule decision heuristic, would coincidentally favor the positively correlated grocery store in both domains. If this were the case, the results from Study 1 could be explained by this round-wise majority-rule decision heuristic rather than by a veritable preference for positively correlated attributes (as I suggest). To rule this out, in Study 2, I experimentally manipulated whether such a decision strategy would favor the positively correlated store, the negatively correlated store, or neither store.

Method

Participants

Six hundred six participants (347 women, 6 non-binary, 5 other, median age = 38) were recruited from AMT and completed the study.

Procedure

Study 2 was nearly identical to the design of Study 1 except for the following critical differences. There was a between-subjects manipulation where participants were randomly assigned to one of the following three conditions:

- Positive round-wise-favored: A round-wise majority-rule decision heuristic would favor the positively correlated store.
- Negative round-wise-favored: A round-wise majority-rule decision heuristic would favor the negatively correlated store.
- Round-wise-neutral: A round-wise majority-rule decision heuristic would be indifferent between the positively and negatively correlated stores.

The purpose of Study 2 was to test whether round-wise comparisons affected participants' preference for the positively correlated store in the $50 budget constraint observed in Study 1. Aside from this critical manipulation, there were a few minor changes to the design of Study 2 compared to Study 1. First, I only tested the $50 budget constraint (rather than $50 and $27 budgets). Second, the mean price of the grocery bill was $33 (rather than $31). Third, in order to generalize the findings from choice to evaluation, participants rated their preference for the left vs. right store options (1 = “Strongly prefer left option,” 5 = “Strongly prefer right option”) rather than choosing their preferred store. The main dependent measure of interest was the average preference for the positively correlated store option. Fourth, participants were explicitly told that they should assume the prices they would experience would have a similar distribution of outcomes as seen in the historical information. Finally, the trip numbers above the price list for each store were decoupled (e.g., the prices for Trip 3 may have been listed on the left store while the prices for Trip 18 may have been listed on the right store). I made this change so that making round-wise comparisons would be less justifiable. If participants assumed Trip 1 occurred on the same day for both stores, then they could justify using the majority rule heuristic because it would help them identify which option had the higher probability of resulting in a lower price. Showing two different trips at the same time would make round-wise comparisons irrelevant if they were trying to get the best price.

After making a choice, participants were asked how many times the left option had a better total price, a manipulation check. They then estimated the best, worst, and average of the total price for both options, followed by a risk preference question between five gambles (e.g., Option 1: 5% chance of $465 and 95% chance of $0; Option 2: 25% chance of $81 and 75% chance of $0). Finally, they answered demographic questions and were paid. The design of Study 2 resembled the design of Study 1 in all other ways.

Results

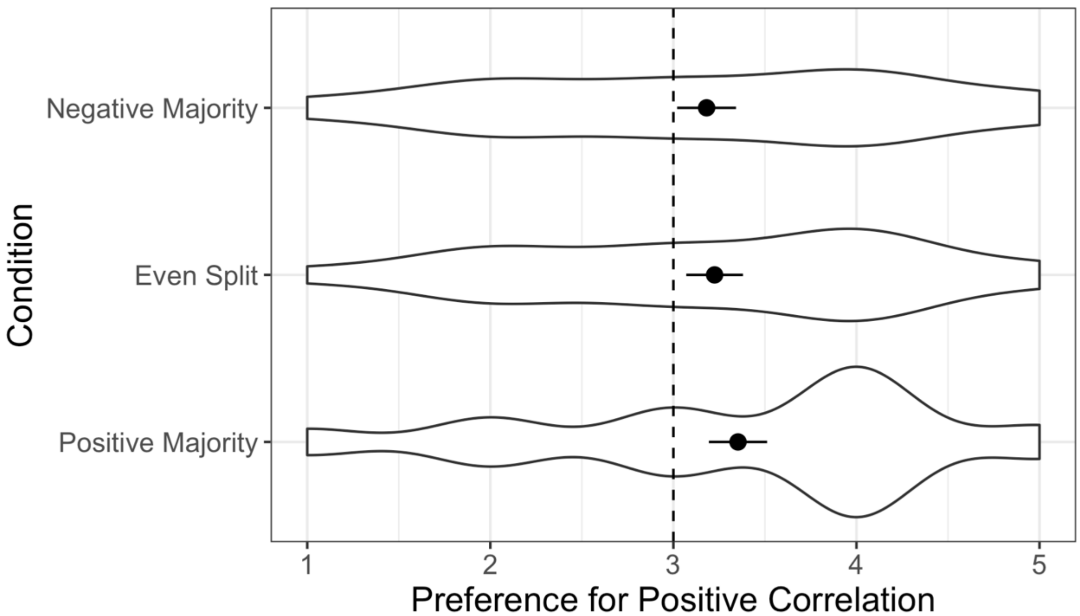
Exclusions

Based on the pre-registered exclusions, I excluded participants who estimated the average total time to be greater than three standard deviations above the mean of the seven average estimates. Twenty-four participants gave at least one estimate, across the lowest, average, and highest total prices for both options, greater than three standard deviations from the mean of all other estimates for the statistic estimated.

Preference

Preference for the left vs right option was recoded to be preferences for the positively correlated attribute option. The pre-registered comparison was the average preference rating in the condition favoring the positive correlation versus the condition favoring the negative correlation. There is not a significant difference in the preference ratings for the positive correlation between the two ($t(580) = 1.50$, $b = 0.09$, $p = .13$). If the preference ratings are turned into a binary rating, with participants who chose the middle option split between the two options equally, a significant difference is not seen in the choice shares ($t(580) = 1.36$, $b = 0.08$, $p = .18$). I calculate the number of times people report the positive correlation had the better price across all three conditions by directly measuring a process measure in the majority rule heuristic. When correlating the true number of times the positively correlated option did better with participant ratings, a statistically significant correlation was not found ($t(580) = 1.36$, $b = 0.08$, $p = .18$). If participants followed this heuristic, then in the even split condition, the average rating would not be statistically different from the mid-point on the scale, and the condition favoring the negative correlation would have an average rating below the scales mid-point. I found that the even split condition had an average rating significantly above the scale's mid-point ($t(579) = 2.78$, $b = 0.23$, $p = .006$). The average preference rating in the negative correlation favoring condition is also significantly above the scale midpoint ($t(579) = 2.31$, $b = 0.18$, $p = .021$), not below it.

Figure 10. Preference Rating for Positive Correlation by Condition



Note. The errorbars show 95% confidence intervals of cell means

Discussion

While participants may be attempting to use the majority rule heuristic, the lack of differences in the manipulation check variable and the unexpected rating results would suggest it was not widely used. There were no detectable differences in the number of times the positively correlated option had better total prices between conditions when it was manipulated. In all conditions, the average rating was above the mid-point of the scale, thus expressing a preference for the positively correlated option, when the majority rule would have predicted ratings below the mid-point for the negative favoring condition, indifference for the evenly split condition, and ratings above the mid-point for the positive favoring condition. As an alternative to the majority rule heuristic, participants could be using the peak-end rule, which approximates an experience by averaging the last outcome experienced and the most intense or peak outcome. In the next two studies, I tested how well the peak-end rule predicted choices between positive and negative attribute correlations.

Study 3A and 3B

Study 2 showed that people's preference for options with positively correlated attributes cannot be explained by using a round-wise majority-rule decision heuristic. However, yet another alternative explanation for the results of Study 1 lingers: the use of a peak-end decision rule (see Kahneman et al., 1993). The peak-end rule is a decision heuristic sometimes used for evaluating events that unfold across time. When using this heuristic, people's global evaluations of an event are disproportionately influenced by both the most intense (the "peak") and the final event (the "end") in the sequence. It is possible that our results from Studies 1 and 2 could be explained by the use of a peak-end decision rule since options with positively correlated attributes are mechanically more likely to yield outlying positive combinations (i.e., higher peaks) compared to options with negatively correlated attributes, whose total payouts are more regressive. In other words, what appears as a preference for options with positively correlated attributes may actually be a preference for options with the best possible outcome (which, mechanically but coincidentally, happens to be options with positively correlated attributes). The purpose of Study 3 was to eliminate this alternative explanation. In Study 3B, I constructed the payouts in a way that would favor one option if participants used a peak-end decision rule and would favor the other option if they were attracted to positively correlated options. This allows for a direct test of the concern that the peak-end heuristic might be an alternative explanation for these results. However, because Study 3B is in a novel decision domain (lotteries), I first sought to conceptually replicate the findings from Studies 1 and 2 in this new domain.

Method

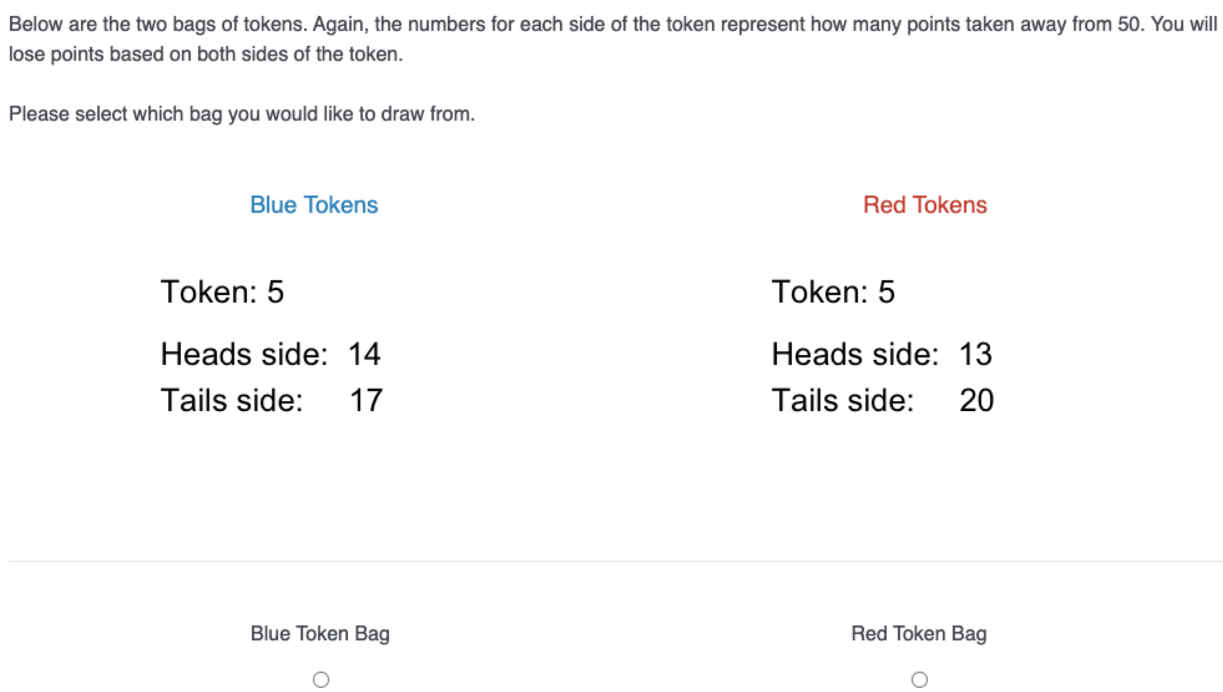
Participants

In 3A, 176 participants (92 women, one non-binary, two prefer not to say, median age = 36, recruited from the AMT) completed the study. For 3B, 176 participants (78 women, two non-binary, two prefer not to say, median age = 37) recruited from AMT completed the study.

Procedure

Study 3A was very similar to the gains-framed choice ($50 budget) from Study 1. However, to test the findings' generalizability, I moved into an unrelated domain: lotteries. Rather than purchasing items from a grocery store, I asked participants to imagine drawing a token from a bag to determine their payoff in a lottery. Participants were asked to imagine there were two bags, each filled with 2000 tokens. Each token had two sides marked with a point value. Participants were told that their total payoff would be determined by adding the point values of the two sides of a randomly chosen token. Similar to Study 1, as depicted in Figure 11, participants then saw the result of 20 draws from both bags using the same "experience sampling" method previously described, leaving 1,980 tokens remaining in each bag. Participants were told the 20 draws would be representative of the actual outcomes they would experience from their selected bag.

Figure 11. Screenshot of Token Information on Decision Page



The critical difference between the two bags was as follows. In one bag, the point values of the two sides of each coin were positively correlated—higher point values on one side were associated with higher point values on the other side—whereas the other bag had negatively correlated point values on the two sides of a given coin. The mean total payoff of each bag was held constant (33). Participants then chose which bag they would rather draw from. To encourage meaningful responses, participants were told that one person would be randomly selected to have their choice honored in order to determine an actual cash bonus, with each point corresponding to one cent. As before, I am interested in people's preference for positively versus negatively correlated outcomes in uncertain decisions. After making their selection, as in Study 2, participants gave estimates of the best, worst, and average of the total payoffs observed from each option.

Study 3B was nearly identical to Study 3A, with the exception that choices in the domain of losses were solicited. Participants were told that they would start with 50 points and would draw a coin to determine how many points would be lost from this initial endowment (the mean loss from each bag was 33 points, resulting in a net of 17 tokens when accounting for the endowment).

Results

Exclusions

As pre-registered, eight participants were excluded for giving the wrong study purpose. An additional two participants were excluded for giving at least one estimate, across the lowest, average, and highest total prices for both options, greater than three standard deviations from the average of all other estimates for the statistic estimated.

Choice

For 3A, the proportion of participants choosing the positive correlation was 53%. This is not significantly greater than the 50/50 choice shares between the two options ($t(162) = 0.83$, $b = 0.03$, $p = .41$). For 3B, the proportion of participants choosing the positive correlation is 67.10%. This is significantly greater than the 50/50 choice shares between the two options ($t(160) = 4.51$, $b = 0.17$, $p < .001$) and the opposite of what the peak-end rule predicts.

Risk Elicitation

Recall that a preference for positively correlated options in the domain of gains runs counter to the predictions of Cumulative Prospect Theory. One potential explanation for these choices is a veritable preference for such options. If our sample of participants (AMT workers) were unusually risk-seeking, this could appear as a preference for positively correlated options. When two options have the same average value, and one has more variability in the outcome than the other, a risk-averse person would prefer the lower variability option, while a risk neutral person would be indifferent. Only a risk-seeking person would prefer to have more variability for the same average outcome and thus choose the positively correlated option. If this were so, the results of the present study would be explained by a preference for positively correlated options directly resulting from participants' (unusually high) underlying risk tolerance. To assess this, I asked a risk preference question unrelated to the main task of the experiment that measured people's risk tolerance, specifically designed to allow for categorizing people as risk-seeking, risk-neutral, or risk averse. This would enable me to test if the observed risk preferences of the sample were in line with typical samples that have demonstrated other aspects of Cumulative Prospect Theory. Indeed, that was my finding. Six participants (3.60% of the sample) were classified as risk-seeking based on the risk-elicitation question in 3A. As for 3B, 10.4% of the sample was classified as risk-seeking based on the risk-elicitation question.

Discussion

If participants closely followed the peak-end rule, it was expected that they would overwhelmingly prefer the positive correlation in 3A, the gain domain study; however, I saw indifference between the options. In the loss domain, this decision heuristic predicts even more poorly. In addition, the peak-end rule predicts that people would prefer the negatively correlated attributes. However, I found that a substantial majority prefer the positively correlated attributes. It is possible that different stimuli, particularly ones with more similar end values, might produce different choice behavior. Stimuli that are more dispersed over the universe of possibilities could more definitely test the predictive power of this heuristic (Goldstein, 2022). There may also be a more complicated set of heuristics that depends on the frame, and thus the peak-end rule would only predict choices with certain decision characteristics. People would need to weigh the end value more heavily in the loss domain to prefer the positive correlation. In the gain domain, weighing the last value more heavily would push people to choose the negative correlation more. While it is not clear that people would end up indifferent in the gain domain, this single pair of studies does not definitely show that the peak-end rule is not predictive of choice. These two studies suggest that people are not using the peak-end rule to approximate the value for these two options when making their choices.

While the risk-elicitation question used here is not a direct measure of tolerance for uncertainty, it seems very unlikely that when 3.60% of the study population is plausibly classified as risk-seeking, it explains why 53% of the sample choose the option most consistent with risk-seeking preferences. In additional studies, participants answered a risk preference measure to check whether the study population is unexpectedly risk-seeking. The next two studies test alternative presentation formats to see how they might impact preferences for the positively correlated attribute option.

Study 4

Study 3 revealed that people do not appear to follow a peak-end decision rule. The studies so far have only considered options with monetary attributes. Further, how attribute information is presented—experience sampling—has been identical across studies. These experimental design decisions may have inadvertently, artifactually given rise to the preference for positively correlated options observed in prior studies. Thus, in Study 4, I changed both the attribute type and presentation format to test the findings' generalizability. In this study, I test the impact of presentation format by comparing experience sampling (as before) to a scatterplot display of the results. Graphical displays may help participants better understand the existence of these correlations and their impact on the distribution of the total outcomes. Further, I test the hypothesis by considering temporal, rather than monetary, attributes of options. People treat time and money differently (Leclerc et al., 1995), and these differences may be especially pronounced in choices under uncertainty.

Method

Participants

Three hundred participants (145 women, median age = 42) recruited from AMT completed the study.

Design

Participants were randomly assigned to see all information in animated, cumulative scatter plots or the experience sampling format. The animated scatter plots cumulatively added points to a scatter plot until the full sample was shown, then the plot cleared, and the sequence started again. As the points were added to the plot, participants also saw the exact values for each attribute.

Procedure

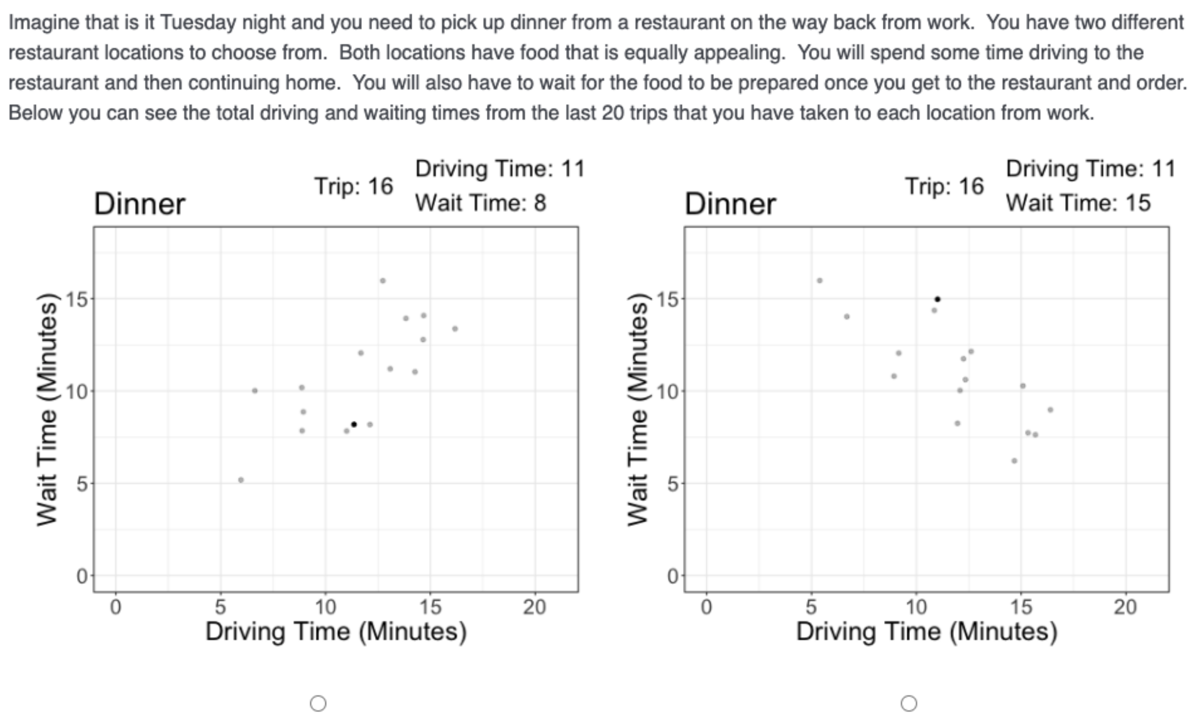
Study 4 is similar to Study 1, with the following critical differences. First, in Study 4, participants considered time rather than financial costs. In this study, participants were asked to imagine choosing between two equally attractive restaurants that each had (a) varying times it took to drive to the restaurant and (b) varying wait times upon arriving.

Second, participants made *two* such choices. As in previous studies, one of the choices was between a positively correlated option and a negatively correlated option. However, the other choice was between two restaurants with positively correlated attributes (order counterbalanced between participants).[8] I expected participants to be indifferent between the two positively correlated options (i.e., approximately 50% of participants choosing each option). I included this in the experimental design to ensure that a preference for positively correlated options (versus negatively correlated options) was not being observed due to unintentionally selecting exceptionally attractive positively correlated options.

[8] One of the two positively correlated restaurants was also an option for the other choice (the choice between a positively and negatively correlated restaurant).

The third difference between Study 1 and Study 4 is that I added a between-subjects manipulation of the presentation format. Participants in the control condition saw attribute information presented using the experience sampling method from Studies 1 - 3. However, participants in the treatment condition (see Figure 12) instead saw attribute information portrayed in a dynamic scatterplot. Each new trip was added as an additional point to a plot of Wait Time and Travel Times. The latest trip to be added was displayed as a black circle, while all previous trips were shown in light gray.

Figure 12. Dynamic Scatterplot Information Display on Decision Page



After each choice, participants saw the options individually and were asked to estimate the average total time, summing across travel and wait times, the errand would take for the presented option. As an additional measure, participants also estimated the likelihood that they would complete the errand in 25 minutes (using a 7-point scale, from “Extremely unlikely” to “Extremely likely”).

Once participants completed the two choice tasks and rated the options, they were given an additional rating task. Here, they saw three other options, sequentially, that varied in their correlations between the timing attributes (r = .8, r = 0, and r = -.8; the order of the correlations was counterbalanced). After seeing each option individually, participants were asked to estimate the average total time and the likelihood of completing the task in 25 minutes using the same 7-point scale as earlier. These additional rating questions were used to test whether participants were sensitive to the impacts these different correlations had on the variability of the total time.

Following that, participants completed a four-item graph literacy scale adapted from (Galesic & Garcia-Retamero, 2011) and a three-item objective numeracy scale adapted from (Fagerlin et al., 2007). Afterward, they were asked to make a choice between risky gambles, just like in Study 2. Finally, they gave their demographic information, including age, gender, income, and education level.

Results

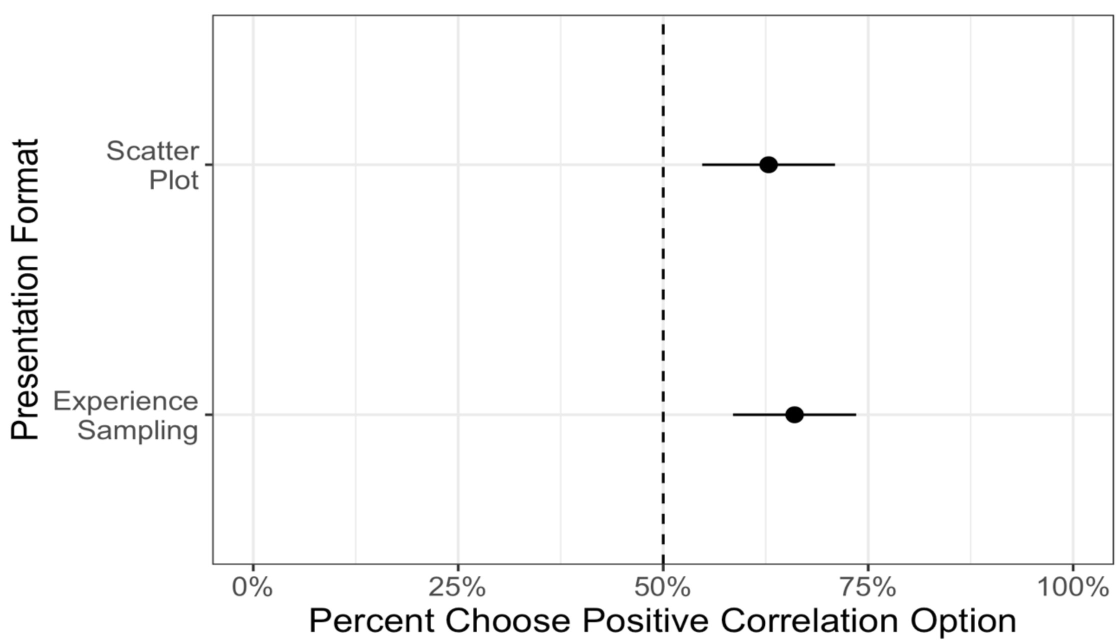
Exclusions

Based on the pre-registered exclusions, I excluded participants who gave an estimate of the average total time greater than three standard deviations from the average across all seven averages. Four participants gave at least one average estimate greater than three standard deviations from the average total time estimate. All seven different options had the same average time and were included in the estimate of the average of the averages.

Choice

I compared the choice shares between the presentation conditions for the positive correlation option to the negative correlation option. There was not a significant difference in choice shares between the two presentation formats ($t(294) = -0.57$, $b = -0.02$, 95% CI = [−0.07, 0.04]). Surprisingly, in each condition, the average proportion of participants who selected the positive correlation was greater than 50% (Scatter plot: $t(139) = 3.14$, $b = 0.13$, $p = .002$, 95% CI = [0.05, 0.21]; experience sampling $t(155) = 4.21$, $b = 0.16$, $p < .001$, 95% CI = [0.09, 0.24], both tests should be considered post-hoc).

Figure 13. Choice Shares for Positive Correlation by Presentation Format



Note. The errorbars show 95% confidence intervals of cell means

Discussion

Studies 1 and 2 show a strong and surprising preference for the positive correlation over the negative correlation option. Here, the presentation format does not appear to impact preferences for the positive correlation. Since it is impossible to know whether participants were considering the outcomes in the domain of gains or losses, it is difficult to know whether they would have such a strong preference for the higher variance option. If I were to assume that participants were directly expressing their risk preferences through this choice, this would imply that around 64.50% of participants in the study are risk-seeking or consider the time attributes to be losses. While this study cannot provide a definitive explanation for the preference for positively correlated uncertainty, two novel aspects do not appear to substantially change the study results. Neither cumulative scatter plots nor time as the outcome attribute seems to have substantially changed the overall choice patterns between higher and lower attribute correlation options. The next experiment tests another presentation format along with simplified experience sampling conditions to see what may impact choice outcomes.

Experiment 5

Study 4 showed that people still prefer positively correlated options even when changing the attribute type (time rather than money) and the presentation format (scatterplots rather than experience sampling). It is worth noting, however, that the presentation formats used up until this point have been somewhat cognitively taxing—requiring participants to combine attribute information (e.g., the total cost of two grocery items per trip) and keep track of this total for two options across many events (e.g., 20 trips to the grocery store). In Study 5, I tested whether simplifying this information can attenuate the preference for positively correlated attributes. In support of this, research on people's investment decisions suggests that alternative display formats that simplify information can weaken people's preference for positively correlated outcomes (Laudenbach et al., in press), but using an adaptation of the quantile dot (Kale et al., 2021). This final study reduces the amount of information participants need to manipulate to test whether this reduced cognitive demand lowers the preference for positively correlated options.

Method

Participants

One thousand six participants (568 women, five non-binary, seven other, median age = 39) were recruited from AMT and completed the study.

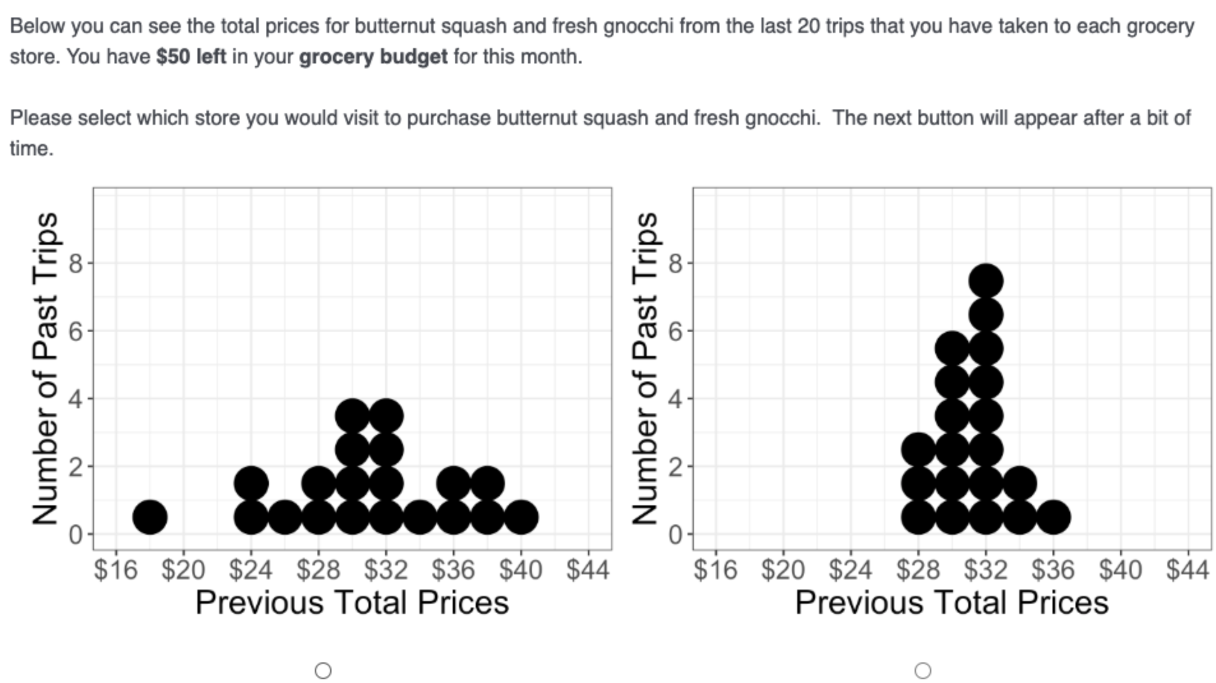
Procedure

Study 5 is similar to Study 1, with the following differences. First, participants only made a single choice (the $50 budget constraint). Second, as in Studies 2 - 3, participants were asked to estimate the best, worst, and average of the total values across the two options after making their choice. Third, as in Study 2, I elicited participants' risk preferences to use as a control in our analysis, using the same question. This method of eliciting risk preferences measures different levels of risk aversion between individuals.

The most significant difference between Studies 1 and 5 is that in Study 5, participants were randomized to one of four different experimental conditions for how price information was displayed.

1. Experience of Attributes: Experience sampling, showing participants the prices of each item on each grocery store trip (the same format used in Studies 1-3)
2. Experience of Attributes and Totals: The same as Condition 1, while also including the total basket price for each trip (i.e., computing the addition for participants)
3. Experience of Totals: Experience sampling, showing people only the total basket price per trip but not including the price of each item
4. Dot plots: A static dot plot (see Figure 14) showing the total basket price for each of the 20 previous trips

Figure 14. Screenshot of Dot Plot Display of Price Information on Decision Page



To keep total price precision consistent across all four conditions, the total prices were always rounded to be even, whole numbers.

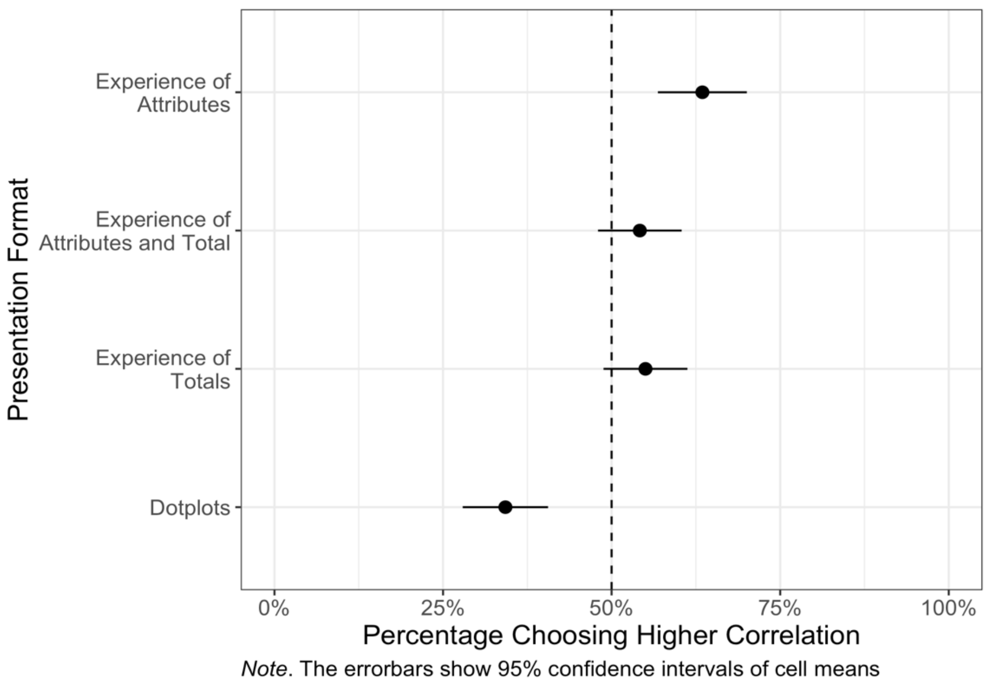
Results

Exclusions

As pre-registered, 31 participants were excluded for giving the wrong study purpose. An additional 45 participants were excluded for giving at least one estimate, across the lowest, average, and highest total prices for both options, greater than three standard deviations from the average of all other estimates for that statistic.

Preference

Figure 15. Choice Shares for Positive Correlation by Presentation Format



Preference for the left vs. right option was recoded as preferences for the positively correlated attribute option. The pre-registered comparisons used a repeated coding scheme. This means that each condition was compared to the neighboring condition in the plot below. Participants' information about the attributes is simplified from top to bottom. The only substantial difference is between the experience sampling presentation of totals compared to the dot plots ($t(923) = -4.58$, $b = -0.21$, $p < .001$). The other differences between neighboring conditions are either not statistically significant (attributes with totals vs totals only: $t(923) = 0.19$, $b = 0.01$, $p = .85$), or only barely significant (attributes only vs attributes with totals: $t(923) = -2.02$, $b = -0.09$, $p = .044$). The dot plot condition is the first condition that is significantly below a 50/50 choice share ($t(923) = -4.76$, $b = -0.16$, $p < .001$).

Discussion

This is the first time that a clear majority of participants in a condition selected the negatively correlated attribute option. This is quite surprising given the persistence of the preference for a positive correlation. In the study context, participants were informed that their remaining budget was $50, which was above any observed total price for either option; thus, they would be choosing in a gain frame. The study was not designed to test why this particular presentation format changes people's preferences substantially and thus cannot provide process evidence. For the two simplified experience sampling conditions, one that includes both attribute and total prices and one that only shows total prices, I do not see a significant decrease relative to the attribute-only condition.

General Discussion

Much of the current research on risk and uncertainty has studied single-attribute options. While progress in this area has been considerable, the uncertainty structures of many real-world decisions are significantly more complex. In light of this, there is a need to refine prevailing models of decision-making that can better predict choices in these more complex environments.

Thus, expanding the set of choices researchers pose to participants will contribute to creating more robust theoretical models and increase the usefulness of these models.

This paper represents such an attempt. Across 5 studies, I observed what looks like risk-seeking preferences—a preference for options with positively correlated attributes—in the domain of gains. Considering the prevailing models of choice under uncertainty, this pattern is both surprising and hard to explain. Additionally, two common decision heuristics, majority rule and peak-end rule, could not predict the choice results observed. Finally, when testing alternative presentation formats, I discovered that only dot plots resulted in choices in the domain of gains consistent with risk aversion (the prediction of CPT). I hope that this research represents a step in the direction of more enriched models of human decision-making under uncertainty.

Beyond the empirical contributions of this paper, I hope this research can be applied to have practical value to people making important, welfare-relevant decisions. For instance, the real-world contexts investigated here—budget-constrained purchase decisions and time-constrained errand completion—could be improved with decision aids designed to support people. When choosing which grocery store to visit, a simple decision aid that uses past shopping behavior and a current grocery list could give distributions of prices for the options people are considering before they leave for the store. As another example, when choosing between different medical treatment options, people may have relatively strong preferences about the relationships between outcomes (e.g., the relationship between expected quality and longevity of life). However, the current informed consent regime does not specifically mandate that people be told about the relationships between such uncertainties (Paterick et al., 2008). Giving people information about how various attributes of treatment options interrelate could have considerable impacts on their choices depending on whether people prefer positively or negatively related uncertainty. These represent just a small number of examples of how further research in this area may help improve the quality of people's decisions.

Future Directions and Open Questions

The insights uncovered by this investigation invite as many questions as it answers. Below is a small selection of open questions and future directions surfaced by this manuscript.

None of the current models of choice under uncertainty—nor any documented decision heuristic that people commonly use—would predict risk-seeking-consistent choice in the domain of gains, as we observed. This preference for positive correlations in the gain domain does not have a clear mechanism to explain it, either in the contexts studied here or in other research on preferences for correlated uncertainty. I observe a surprising preference for positively correlated options across contexts studied here, which range from grocery shopping (price attributes) to selecting tokens in a lottery (point attributes) to running errands (timing attributes), as well as a variety of presentation methods: experience sampling dot plots, and dynamic scatterplots.

Given how often people choose between options with multiple attributes, additional research is needed to understand the cause of this (surprising) preference for positively correlated attributes. The discovery of loss aversion took a long time because it was difficult to create situations that involved real losses. Such a discovery had a radical impact on the models of uncertainty. Studying preferences of the relationship between attributes may be similarly unusual because it is much more difficult to create in a decision and can be difficult to recognize in the real world. Both of these factors may have contributed to why this preference for positively correlated attributes is not an empirical phenomenon incorporated into model building. It may be time to incorporate this as an explainable finding.

Several related investigations have found a preference for positively correlated assets in the investing domain. Multiple paradigms have been used, but they typically involve an allocation decision between investments or selecting an additional investment to add to an initial portfolio. In these studies, there are clear normative prescriptions based on stimuli. People who are even slightly risk-averse, the majority of the population, ought to invest in negatively correlated assets; however, there are strong preferences for the positively correlated assets as well as misunderstandings about the benefits of diversification. Unfortunately, in the investment context, there is a lack of process evidence about why people seem to prefer these positive correlations and, importantly, why certain displays increase the proportion of people selecting the negatively correlated options. However, no one knows why and it contradicts our prevailing models.

Since these tasks are very cognitively demanding, people are still likely to use a heuristic for their judgments. Without additional process evidence, it is hard to ascertain if people are using heuristics that ignore much, if not all, of the uncertainty. In lieu of the CPT value function, there may be an alternative heuristic that results in framing-dependent choices, but for entirely different reasons. Additional evidence that people are treating this decision as one involving uncertainty would help to narrow how people are making this decision. The simplification in the dot plot presentation format likely impacts any heuristics people are using in evaluation.

However, understanding which heuristics people are using for the different presentation formats is a clear next step for additional research.

So far, all the studies in this chapter have used the same attribute types (both attributes have been money or time). While this has simplified the analysis greatly, since we can assume equal attribute weighting, if we change the attribute types—so that they are not reducible to a single metric—we need to understand additional aspects of the decision process. One of the most common models of multi-attribute choice is the weighted additive model. It dictates that people estimate a value for each attribute and then weight each attribute by an importance weight in order to calculate a total value for each option. Because we have been forcing attributes to always be the same type, it allowed us to completely ignore the weighting components needed in multi-attribute decision modeling.

Furthermore, the most important decisions that people make, including who to marry, what city to live in, or what job to take, all have different attributes. Throwing away this richness hampers the ability to make prescriptive predictions about how people could improve the decision-making process, be it structuring how they learn relevant information or a decision aid to help with aggregating so many different attributes. Limiting the types of decisions used in developing these models may also lead to a false sense of generalizability simply because the field has not tested how the models may fail.

The development of risk preference elicitation has found that people often have different risk preferences depending on the domain: monetary, health, or social domains. By studying single attribute outcomes (e.g., the impacts of a medical treatment option just on longevity), the field has neglected the need to understand how risk preferences between attributes may interact with options that have different attributes (e.g., choosing what city to live in impacts both health and income). While the current models assume the aggregation of risk preferences likely has intuitive results, my demonstration of a preference for positive correlations suggests that people may have unexpected preferences over combinations of different attributes. Additional measurement tools will likely be needed to understand the relationships between these domain-specific risk preferences.

Presumably, people are risk-averse to uncertain choices of multi-attribute options in the domain of gains. However, when displaying information using scatterplots or experience sampling, people appear to be risk-seeking (i.e., preferring positively correlated options). Therefore, a better presentation format is needed to represent outcomes of uncertain multi-attribute choices that lead people to make choices in line with their plausibly risk-averse preferences. As shown in Study 5, the only display format that successfully did this, bringing people's choices in line with their single attribute risk preferences, was the dot plot. However, given the nature of dot plots, this presentation format is only possible when all of the attributes of the option are reducible to a single outcome type (i.e., when all of the attributes are only related to time or money, but not, for example, when one attribute is related to time and the other related to money).

For an investing scenario, other studies have found that presentation formats like dot plots, where returns from multiple investments are aggregated, have helped people select negatively correlated investments. Current investing platforms could easily implement such decision aids when people choose between investments. Suppose someone was looking for a new investment and the investments they were keeping was an index fund in the S&P 500. They are considering whether to invest in a real estate fund or developing markets index fund. The decision aid could show them aggregated distributions of historical returns of S&P 500 plus the real estate fund versus the aggregated distribution of the S&P 500 plus the developing markets fund. Based on the prior research, I would expect people to select the more negatively correlated combination as long as the average returns are similar. Future work will seek to take on these and other open questions.

I pause to note that people may believe the static nature of the dot plots explains the preference for negative correlations. Other studies have compared static (showing information in a grid) vs. dynamic. However, in this comparison, I do not see differences in the preference for a positive correlation. Instead of focusing on the static vs. dynamic differences in presentation format, other directions would likely be more fruitful.

Almost all decisions people face are uncertain. Most options that people consider for consequential decisions are multi-attribute. While research in decision science has made significant strides in comprehending the simplest form of these decisions, far more work is needed to understand the richer, more complex version of these decisions. This paper is an effort in that direction.

DISCUSSION

Consider a person who is faced with some uncertain decision. To make a well-considered decision, this person must: 1) form an impression of the degree and nature of the uncertainty associated with each option and 2) form preferences among the options given their sense of these uncertainties. In this dissertation, I demonstrated faults in representing the uncertainty (Chapter 1) and choosing among uncertain options (Chapter 2). Specifically, in this dissertation, I have demonstrated that this process of construing and deciding among uncertain options is rife with normative errors. Chapter 1 showed how people's impressions of uncertainty are inconsistent when the elicitation metrics systematically vary, and Chapter 2 demonstrated an implausibly high preference for options with positively correlated attributes. These errors represent meaningful threats to people's welfare, considering the prevalence and impact of uncertain decisions in people's lives.

Imagine a couple leaving Austin, Texas, and moving to one of two new cities: Chicago or Cincinnati. The couple, both of whom are lawyers, have been searching for jobs for a while, and these are the two cities where they both have reasonable job offers. In Chicago, they have job offers from rival firms, each with a salary offer of $80,000; in Cincinnati, they would be working for the same law firm with a starting salary of $60,000. At both firms, on top of the base salaries, they could receive performance-based bonuses each year, adding uncertainty to their total income from each job.

In order to equate these offers, they must translate the salaries—and the associated uncertainty about their final end-of-year incomes—between the differing costs of living in the two cities. Comparing Cincinnati to Chicago, there is a cost of living adjustment (COLA) of about 25% to consider.

If they were to estimate the distribution of income from each of the Cincinnati offers, they might estimate that there is a:

- 95% chance of making at least $65,000,
- 50% chance of making at least $70,000, and
- 5% chance of making at least $75,000.

Translating these into Chicago-equivalent salaries, adjusting for the COLA should require multiplying each estimate by 1.3. This would result in a prediction of a:

- **A1:** 95% chance of making at least $84,500,
- **A2:** 50% chance of making at least $91,000, and
- **A3:** 5% chance of making at least $97,500.

However, instead of estimating their incomes in Cincinnati and then COLA-adjusting their estimates for Chicago, imagine that they went in the opposite direction: They estimated the Cincinnati offers in Chicago dollars. As was seen from Chapter 1, people tend to insufficiently adjust their uncertainty for changes in the units of their estimates. Given this tendency, because $1 in Chicago buys less than $1 in Cincinnati, this smaller unit might likely lead people to give narrower uncertainty estimates of their incomes in Chicago. Specifically, as an example, they may predict that, for their Chicago salaries, there is a:

- **B1:** 95% chance of earning at least $86,000,
- **B2:** 50% chance of earning at least $91,000, and

- **B3:** 5% chance of making at least $96,000.

Compare the estimated income distribution for the Chicago jobs when estimated in Chicago dollars (B1 - B3) versus when estimated in COLA-adjusted Cincinnati dollars (A1-A3). Note that the estimated uncertainty is smaller in Chicago dollars compared to the larger unit size, Cincinnati dollars; the lower bound (A1) is $6,500 less than the central estimate (A2) in the Cincinnati dollar estimation, whereas the lower bound (B1) is only $5,000 from the central estimate (B2) in the Chicago dollar estimation. A seemingly trivial decision about whether the couple imagines comparing Cincinnati dollars in Chicago or Chicago dollars in Cincinnati can wind up having an outsize impact on their sense of the associated uncertainty. If the couple is targeting some specific absolute income level, the consequences of this arbitrary decision can be vast. In particular, this kind of transformation clearly impacts the distribution's tails.

In this case, the correlation between the two people's incomes at their respective jobs is a feature of the job offers created by their employment at the same firm or competing firms. In Chicago, where they would be employed by rival firms, when one firm performs better, the other firm is likely to perform worse. This would lead to the bonuses at the two firms being negatively correlated (e.g., one person in the couple gets a higher bonus, say $10,000, and the other one gets a worse bonus, say $2,000). However, in Cincinnati, where the couple would be working at the same firm, a good year for the firm would mean a good year for both of them (e.g., both get high bonuses, say $7,500). Also, a bad year for the firm would mean a bad year for both of them (e.g., both get low bonuses, say $1,500), producing a positive correlation in the bonuses. Based on the findings of Chapter 2, it is expected that people would prefer the positively correlated outcomes and thus favor Cincinnati. However, the prevailing models of uncertain choice cannot explain such a widespread preference for positive correlations.

More to the point, a preference for positively correlated outcomes can have significant harmful consequences. Preferring positively correlated options compared to negatively correlated options with the same average outcome leads to choices with greater variability in their outcomes. Here, choosing to live in Cincinnati could mean that in some years, the total compensation is $124,000 but could plausibly go as high as $140,000 (in Cincinnati dollars). On the other hand, if they choose to reside in Chicago, they will likely earn approximately $132,000 (in Cincinnati dollars) annually, with fluctuations of $2,000 or less around the $132,000. When setting up budgets for their future lives, their essential regular monthly expenses (e.g., housing, transportation, student loans, food, and so on) will have some limits they cannot go below without suffering severe consequences. Having positively correlated income increases the chances that their annual income does not cover all of their essential purchases in a given year (i.e., a larger mass of the distribution of possible incomes below this minimum budget for the high variance, positively correlated option). If the couple fails to consider this additional variability associated with the Cincinnati move, they could end up in a dire situation without enough money to pay for their essentials.

Previous research on these types of decisions tended to focus on risky choices with known probabilities (rather than uncertain probabilities) of discrete outcomes (rather than a continuous outcome space). For example, people would be asked to imagine that moving to Cincinnati would result in a 50% chance of earning $130,000 and a 50% chance of earning $132,000, while the

Chicago option would result in a 50% chance of earning $124,000 and a 50% chance of earning $140,000. While insights gleaned from explorations of this sort are essential, the outcome space is continuous. Both judgments and decision-making become far harder to model but nonetheless important when our investigation is broadened to consider continuous outcomes.

Prospect Theory's four-fold pattern of risk has done much to explain a wide range of human behavior dealing with very small probabilities (e.g., the possibility effect; explaining the purchase of lottery tickets and insurance) and very large probabilities (e.g., the certainty effect; explaining the premium people put on guaranteed outcomes). While tail-risk events seem to be exceptionally important, they are also the place where estimates of uncertainty of continuous outcomes are most prone to error. As just one example, the unwarranted presumption of symmetric uncertainty—demonstrated in Chapter 1—may lead people to truncate tail events from skewed distributions. The process by which people convert uncertain outcomes into a decision weight distribution can systematically distort their sense of the likelihood of these events. Hence, extremely good outcomes in a right-skewed outcome distribution may be underweighted or ignored, leading people to reject opportunities they may have actually wanted. The four-fold pattern of risky choice predicted by Prospect Theory demonstrated errors in how people weight tail-risk events; this research reveals additional distortions that may impact the translation from probabilities to decision weights. Additional research may thus help to enrich the current models of uncertainty.

This manuscript is only the first step in a research agenda investigating how people make decisions with complex (though true-to-life) structures and outcomes. In doing so, I hope to better explain and predict—and ultimately assist with—the complex, messy choices that people face in their lives. So many real-world decisions involve highly complex, interacting uncertainty; there is great value in enhancing current choice models. Thus, this dissertation is the first step in that direction.

BIBLIOGRAPHY

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss Aversion Under Prospect Theory: A Parameter-Free Measurement. *Management Science*, *53*(10), 1659--1674. https://doi.org/10.1287/mnsc.1070.0711

André, Q., Reinholtz, N., & Langhe, B. D. (2021). Can Consumers Learn Price Dispersion? Evidence for Dispersion Spillover Across Categories. *Journal of Consumer Research*. https://doi.org/10.1093/jcr/ucab030

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the Law Relating Processing to Storage in Working Memory. *Psychological Review*, *118*(2), 175–192. https://doi.org/10.1037/a0022324

Baumgartner, H., Sujan, M., & Padgett, D. (1997). Patterns of Affective Reactions to Advertisements: The Integration of Moment-to-Moment Responses into Overall Judgments. *Journal of Marketing Research*, *34*(2), 219–232. https://doi.org/10.1177/002224379703400203

Beach, L. R., & Scopp, T. S. (1968). Intuitive statistical inferences about variances. *Organizational Behavior and Human Performance*, *3*(2), 109–123. https://doi.org/10.1016/0030-5073(68)90001-9

Burson, K. A., Larrick, R. P., & Lynch, J. G. (2009). Six of One, Half Dozen of the Other. *Psychological Science*, *20*(9), 1074–1078.

Cornil, Y., Hardisty, D. J., & Bart, Y. (2019). Easy, breezy, risky: Lay investors fail to diversify because correlated assets feel more fluent and less risky. *Organizational Behavior and Human Decision Processes*, *153*, 103–117. https://doi.org/10.1016/j.obhdp.2019.06.001

Ellsberg, D. (1961). *Risk, Ambiguity, and the Savage Axioms*. *75*(4), 643--669.

Epstein, L. G., & Tanny, S. M. (1980). Increasing Generalized Correlation: A Definition and Some Economic Consequences. *The Canadian Journal of Economics*, *13*(1), 16–34.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989x07304449

Finetti, B. de. (1937). Foresight: Its logical laws, its subjective sources (H. E. Kyburg & H. E. Smokler, Trans.). *Annales de l'Institut Henri Poincare*, *7*, 1–68.

Flannagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional Expectations and the Induction of Category Structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 241–256. https://doi.org/10.1037/0278-7393.12.2.241

Fox, C. R., & Tversky, A. (1995). Ambiduity Aversion and Comparative Ignorance. *Quarterly Journal of Economics*, *110*(3), 585--603.

Galesic, M., & Garcia-Retamero, R. (2011). Graph Literacy. *Medical Decision Making*, *31*(3), 444–457. https://doi.org/10.1177/0272989x10373805

Goldstein, D. (2022). Leveling Up Applied Behavioral Economics. In A. Samson, K. Vohs, & A. Shah (Eds.), *The Behavioral Economics Guide 2022*. Behavioral Science Solutions Ltd.

Goldstein, D. G., Johnson, E. J., & Sharpe, W. F. (2008). Choosing Outcomes versus Choosing Products: Consumer-Focused Retirement Investment Advice. *Journal of Consumer Research*, *35*(3), 440--456. https://doi.org/10.1086/589562

Golman, R., Gurney, N., & Loewenstein, G. (2021). Information gaps for risk and ambiguity. *Psychological Review*, *128*(1), 86–103. https://doi.org/10.1037/rev0000252

Haran, U., Moore, D. A., & Morewedge, C. K. (2010). A simple remedy for overprecision in judgment. *Judgment and Decision Making*, *5*(7), 467–476. http://journal.sjdm.org/10/101027/jdm101027.pdf

Hofstätter, P. R. (1939). Über die Schätzung von Gruppeneigenschaften. *Zeitschrift Für Psychologie*, *145*, 1–44.

Hogarth, R. M. (1975). Cognitive Processes and the Assessment of Subjective Probability Distributions. *Journal of the American Statistical Association*, *70*(350), 271–289. https://doi.org/10.1080/01621459.1975.10479858

Hullman, J., Resnick, P., & Adar, E. (2015). *Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering*. https://doi.org/10.1371/journal.pone.0142444

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science*, *4*(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of Decision under Risk. *Econometrica*, *47*(2), 263--292.

Kale, A., Kay, M., & Hullman, J. (2021). Visual Reasoning Strategies for Effect Size Judgments and Decisions. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), 272–282. https://doi.org/10.1109/tvcg.2020.3030335

Kalyanaram, G., & Little, J. D. C. (1994). An Empirical Analysis of Latitude of Price Acceptance in Consumer Package Goods. *Journal of Consumer Research*, *21*(3), 408. https://doi.org/10.1086/209407

Knight, F. H. (1921). *Risk, Uncertainty, and Profit*. Houghton Mifflin Company.

Langhe, B. de, & Puntoni, S. (2016). Productivity Metrics and Consumers' Misunderstanding of Time Savings. *Journal of Marketing Research*, *53*(3), 396–406. https://doi.org/10.1509/jmr.13.0229

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 316–337). Blackwell Publishing Ltd.

Larrick, R. P., & Soll, J. B. (2008). The MPG Illusion. *Science*, *320*(5883), 1593–1594. https://doi.org/10.1126/science.1154983

Lathrop, R. G. (1967). Perceived Variability. *Journal of Experimental Psychology*, *73*(4), 498–502. https://doi.org/10.1037/h0024344

Laudenbach, C., Ungeheuer, M., & Weber, M. (n.d.). How to Alleviate Correlation Neglect in Investment Decisions. *Management Science*. https://doi.org/10.2139/ssrn.3086722

Leclerc, F., Schmitt, B. H., & Dube, L. (1995). Waiting Time and Decision Making: Is Time like Money? *Journal of Consumer Research*, *22*(1), 110. https://doi.org/10.1086/209439

Levy, H. (1992). Stochastic Dominance and Expected Utility: Survey and Analysis. *Management Science*, *38*(4), 555–593. https://doi.org/10.1287/mnsc.38.4.555

Maeyer, P. D., & Estelami, H. (2012). Impact of online consumer reviews on sales and price strategies: a review and directions for future research. *Journal of Product & Brand Management*, *21*(2), 132–139. https://doi.org/10.1108/10610421211215599

Maglio, S. J., & Trope, Y. (2011). Scale and construal: How larger measurement units shrink length estimates and expand mental horizons. *Psychonomic Bulletin & Review*, *18*(1), 165–170. https://doi.org/10.3758/s13423-010-0025-1

Neumann, J. von, & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Palma, A. de, Abdellaoui, M., Attanasi, G., Ben-Akiva, M., Erev, I., Fehr-Duda, H., Fok, D., Fox, C. R., Hertwig, R., Picard, N., Wakker, P. P., Walker, J. L., & Weber, M. (2014). Beware of black swans: Taking stock of the description–experience gap in decision under uncertainty. *Marketing Letters*, *25*(3), 269–280. https://doi.org/10.1007/s11002-014-9316-z

Pandelaere, M., Briers, B., & Lembregts, C. (2011). How to Make a 29% Increase Look Bigger: The Unit Effect in Option Comparisons. *Journal of Consumer Research*, *38*(2), 308–322. https://doi.org/10.1086/659000

Paterick, T. J., Carson, G. V., Allen, M. C., & Paterick, T. E. (2008). Medical Informed Consent: General Considerations for Physicians. *Mayo Clinic Proceedings*, *83*(3), 313–319. https://doi.org/10.4065/83.3.313

Peer, E. (2010). Exploring the time-saving bias: How drivers misestimate time saved when increasing speed. *Judgment and Decision Making*, *5*(7), 477–488.

Pelham, B. W., Sumarta, T. T., & Myaskovsky, L. (1994). The Easy Path From Many To Much: the Numerosity Heuristic. *Cognitive Psychology*, *26*(2), 103–133. https://doi.org/10.1006/cogp.1994.1004

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica*, *32*(1/2), 122. https://doi.org/10.2307/1913738

Prelec, D., & Loewenstein, G. (1991). Decision Making over Time and under Uncertainty: A Common Approach. *Management Science*, *37*(7), 770–786.

Raghubir, P., & Srivastava, J. (2002). Effect of Face Value on Product Valuation in Foreign Currencies. *Journal of Consumer Research*, *29*(3), 335–347. https://doi.org/10.1086/344430

Reinholtz, N. (2015). *Persistence in Consumer Search*.

Reinholtz, N., Fernbach, P. M., & Langhe, B. de. (2021). Do People Understand the Benefit of Diversification? *Management Science*. https://doi.org/10.1287/mnsc.2020.3893

Ricker, T. J., & Cowan, N. (2014). Differences Between Presentation Methods in Working Memory Procedures: A Matter of Working Memory Consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 417–428. https://doi.org/10.1037/a0034301

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. https://doi.org/10.1111/1467-9280.00066

Shen, L., & Urminsky, O. (2012). Making Sense of Nonsense: The Visual Salience of Units Determines Sensitivity to Magnitude. *Psychological Science*, *24*(3), 297–304. https://doi.org/10.1177/0956797612451470

Soll, J. B., & Klayman, J. (2004). Overconfidence in Interval Estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299--314. https://doi.org/10.1037/0278-7393.30.2.299

Thaler, R. (1985). Mental Accounting and Consumer Choice. *Marketing Science*, *4*(3), 199–214. https://doi.org/10.1287/mksc.4.3.199

Thomas, M., & Morwitz, V. (2009). Heuristics in Numerical Cognition: Implications for Pricing. In *Handbook of pricing research in marketing* (pp. 132–149).

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/bf00122574

Ungeheuer, M., & Weber, M. (2021). The Perception of Dependence, Investment Decisions, and Stock Prices. *The Journal of Finance*, *76*(2), 797–844. https://doi.org/10.1111/jofi.12993

Weber, B. J., & Chapman, G. B. (2005). Playing for peanuts: Why is risk seeking more common for low-stakes gambles? *Organizational Behavior and Human Decision Processes*, *97*(1), 31–46. https://doi.org/10.1016/j.obhdp.2005.03.001

Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting Risk Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation. *Psychological Review*, *111*(2), 430–445. https://doi.org/10.1037/0033-295x.111.2.430

Wertenbroch, K., Soman, D., & Chattopadhyay, A. (2007). On the Perceived Value of Money: The Reference Dependence of Currency Numerosity Effects. *Journal of Consumer Research*, *34*(1), 1–10. https://doi.org/10.1086/513041

Winkler, R. L. (1967). The Assessment of Prior Distributions in Bayesian Analysis. *Journal of the American Statistical Association*, *62*(319), 776–800. https://doi.org/10.1080/01621459.1967.10500894

Wu, G., & Gonzalez, R. (1996). Curvature of the Probability Weighting Function. *Management Science*, *42*(12), 1676--1690. https://doi.org/10.1287/mnsc.42.12.1676

Zhang, J., Hsee, C. K., & Xiao, Z. (2006). The majority rule in individual decision making. *Organizational Behavior and Human Decision Processes*, *99*(1), 102–111. https://doi.org/10.1016/j.obhdp.2005.06.004