PROCEDURE INVARIANCE VIOLATIONS

IN CONSUMER BEHAVIOR

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

---

by
Sangsuk Yoon
May 2018

Examining Committee Members:

Vinod Venkatraman, Advisory Chair, Department of Marketing & Supply Chain Management

Nathan Fong, Department of Marketing & Supply Chain Management

Crystal Reeck, Department of Marketing & Supply Chain Management

David Smith, External member, Department of Psychology

ABSTRACT

Although prior studies have widely examined how descriptions of task environment influence consumer preference, the effect of procedure elicitation methods on consumer preference have not yet been explored thoroughly. To address this issue, this three-essay dissertation investigates the effect of preference elicitation methods on consumer preference in three different domains: anchoring, risky choice and decision framing. This dissertation also uses a multi-method approach that includes behavioral experiments, meta-analysis, *p*-curve analysis, eye-tracking, and computational modeling to deeply understand the impact, robustness, and underlying processes of procedural manipulations in the three domains. The overall results show that changes in decision processes not only affect consumer preference immediately in all three domains, but also have long-term effects. Critically, these findings imply that the impact of procedural manipulations on consumer preference is robust and is not a short-term distortion of preference. Thus, marketers and policy makers can utilize different procedural elicitation methods to shape long-term consumer preferences, and need to consider decision procedure in setting up marketing strategies. Limitations and future research directions are discussed in the last chapter of the dissertation.

DEDICATION

This dissertation could be successfully completed with the support and guidance of a group of supportive people. I wish to thank my advisor, mentor, and dissertation chair Vinod Venkatraman, my best supporter and co-worker Nathan Fong, my dissertation committee Crystal Reeck and David Smith, and the Center for Neural Decision Making at the Fox School of Business. I also thank my friends Khoi Vo, David Marston, Megha Joshi, Xudong Chen, and Gary Zhou for helping me inside and outside of the school. Finally, but most certainly, I would like to thank my wife Mingyeong Choi, my parents, and my brother for their unconditional love and support through it all.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1.

INTRODUCTION

Understanding the characteristics of human preference is a crucial aspect in predicting how people make judgments and decisions. Normative theory of human preference assumes that people have inherent preference which is absolute, stable, consistent, and precise (March, 1978). Innate and inherent preference is revealed through given judgment and decision tasks. Such preference is not affected by the way it is elicited (procedure invariance principle), or the way given task environment are described (description invariance principle; Tversky, Sattath, & Slovic, 1988). However, in contrast to the normative theory, numerous studies in marketing, psychology, and behavioral economics have shown that the procedure invariance and description invariance assumptions can be easily falsified. For example, people are risk averse in gain framing, while they are risk seeking in loss framing, falsifying the description invariance assumption (Tversky & Kahneman, 1981). In another example, people prefer the P-bet (i.e., a high probability of winning gamble) in a choice task, while they prefer $-bet (i.e., a high payoff gamble) in a valuation task, falsifying the procedure invariance assumption (Lichtenstein & Slovic, 1971). Scholars in this preference construction stream have argued that human preference is not simply revealed rather preference is constructed as a function of given decision contexts and decision agents' cognitive and emotional status at the moment of making decisions (Payne, Bettman, & Johnson, 1988; H. A. Simon, 1955).

Numerous prior studies in consumer contexts also have shown that consumers show inconsistent preferences depending on the description of attributes of items (Levin & Gaeth, 1988), service benefits (Ganzach & Karsahi, 1995), price discount (Gendall,

Hoek, Pope, & Young, 2006; Khan & Dhar, 2010), and payment methods (Gourville, 1998). Although abundant studies have examined consumer behavior in contexts where descriptions of the task environment vary, relatively less research have been conducted to understand how different procedures eliciting preference affect consumer preference. Moreover, recent replication attempts of procedure manipulation studies have shown a null effect or small effects than those of the original studies' (e.g., Klein et al., 2015; Rohrer, Pashler, & Harris, 2015; Shanks et al., 2015). For example, recently behavioral priming studies in which participants are asked to think about specific occasions or status before answering main questions could not be replicated (Harris, Coburn, Rohrer, & Pashler, 2013; Pashler, Coburn, & Harris, 2012; Pashler, Rohrer, & Harris, 2013). To address the lack of research in this area and weak evidence, this three-essay dissertation investigates anchoring effects on product valuations, description-based and experience-based risky decision making, and choose vs. reject decision framing effect on preferential choices.

CHAPTER 2 examines anchoring effects on product valuations. Anchoring effects refer to a phenomenon where consideration of a prior standard of comparison influences subsequent numeric judgments, even when the standard of comparison is explicitly non-informative to the given task (Tversky & Kahneman, 1974). Anchoring effects have been shown in diverse contexts (e.g., factual questions, physical length estimation, mathematics problem), including product valuations. Anchoring effects on product valuations have been regarded as evidence supporting preference construction, but recent replication attempts have shown inconsistent results (Bergman, Ellingsen, Johannesson, & Svensson, 2010; Fudenberg, Levine, & Maniadis, 2012; Jung, Perfecto,

& Nelson, 2016; Maniadis, Tufano, & List, 2014). In CHAPTER 2, I attempt to reconcile previous inconsistent replication attempts results by conducting three experiments testing the effect of different experimental procedures used in previous replication studies of anchoring, a cross-study analysis, a *p*-curve analysis, and a meta-analysis to show that the anchoring effect on willingness-to-pay (WTP) is robust evidence of preference construction.

CHAPTER 3 examines whether the anchoring effect on WTP is transient. If the anchoring effect on WTP is a simple response bias, then the effect on WTP will decay and disappear. However, if the anchoring effect on WTP influences underlying preferences, then the effect will persist over time. To address this research question, two multi-session experiments were conducted to investigate whether the anchoring effect on WTP persists over time by manipulating the time gap between anchoring procedures and WTP responses. Study 1 examines whether immediate WTP responses after anchoring procedures moderate the long-term anchoring effect on WTP, and Study 2 examines whether repeating WTP responses moderates the long-term anchoring effect on WTP. The findings of these two chapters imply that the anchoring effect on product valuations is a robust phenomenon and is not just a simple response bias.

CHAPTER 4 examines how different procedures of learning payoffs and risk levels influence people's preferences for risky alternatives. More specifically, I investigate risk preference in two presentation formats: description-based and experience-based decision formats. Consumers make various risky decisions, such as retirement saving, mortgage, medical treatment, insurance decisions. Numerous risk preference studies have used gambling tasks in which possible outcomes and associated risky levels

are explicitly described. However, in real-life, it is rare for us to know the possible outcomes and associated probabilities exactly. Instead, we have to learn the payoffs and risk levels from repeated experience or observations. Previous literature has shown that people overweight a rare event in a binary gambling task when the payoffs and associated risk levels are explicitly described, while they underweight the rare event when they are asked to learn the information from repeated sampling (Hertwig, Barron, Weber, & Erev, 2004b).

Because it is hard to differentiate decision strategies in a binary choice gamble, and risky decisions in real-world are more complex than a binary choice task, CHAPTER 4 examines how people make risky decisions in the description-based and experience-based decision formats using a more complex gambling task (ternary outcome-ternary alternative). Utilizing a multi-alternative and multi-attribute gambling task in experience-based decision format is expected to provide a higher ecological validity than a simple binary choice task or description-based decision format. Especially, I focus on the role of the overall probability of winning heuristic (i.e., making decisions based on the overall possibility that one can achieve positive outcomes regardless of the amount of the positive outcomes) in both decision formats. The findings of CHAPTER 4 indicate that even though people use different decision strategies between the two decision formats, people rely on the overall probability of winning heuristic comparably in both formats.

CHAPTER 5 focuses on the effect of procedural manipulations on preferential choices. Especially, this chapter examines how different decision modes that people employ in previous choice decisions influence subsequent preferential choices. In many consumer purchase occasions, consumers make the same purchase decisions repeatedly,

thus confront same or similar alternatives repeatedly (e.g., purchasing a car or laundry detergent, deciding a restaurant for dinner). In each time, consumers do not rely on a single decision strategy; rather they employ different strategies. For example, people might purchase an item by actively choosing what they like the most among available items in a choice set or by passively rejecting unattractive alternatives in the choice set until they find the most attractive item (Chernev, 2009; Dhar & Wertenbroch, 2000; Foo, Haji, & Sakai, 2014; Ganzach, 1995; Huber, Neale, & Northcraft, 1987; Meloy & Russo, 2004; Shafir, 1993). Thus, this essay investigates the underlying cognitive mechanisms of the choose and reject decisions and how the two decision modes people employ in previous choice decisions affect their perception on the considered items in later preferential choices.

It has been debated how people make the choose and reject decisions. The compatibility hypothesis posits that people process task compatible information (i.e., choosing-positive attributes vs. rejecting-negative attributes) more than task incompatible information, while the commitment hypothesis posits that people process information more in the choose decision than in the reject decision. A compromised model, the accentuation model, posits that the choose decision is a default decision mode for people while the reject decision is not. Thus, the reject decision is harder, and people are more deliberate in the reject decision than in the choose decision. These three models provide distinctive explanations about the underlying mechanism of the choose and reject decisions and predictions about the effect of the choose and reject decisions on subsequent consumer preferences. Investigating the underlying mechanisms using eye-tracking and computational modeling methods and examining the effect of choose vs.

reject decision frame on subsequent preferences are important in reconciling the three theoretical claims. The results of CHAPTER 5 show that people are more deliberate in the reject decision than in the choose decision and that the effect of the two decision frames on the alternatives in a choice set is greater in the choose task than in the reject task.

This dissertation seeks to understand the effect of procedural manipulations on consumer preferences in three different decision domains; preferential judgments (CHAPTER 2 and CHAPTER 3), risky decision making (CHAPTER 4), and preferential choice (CHAPTER 5). Recent replication crisis emphasizes the importance of rigorous testing of research findings. The current dissertation also uses a multi-method approach including behavioral experiment, meta-analysis, p-curve analysis, eye-tracking and computational modeling, and this multi-method approach is expected to cross-validate the key behavioral findings in this dissertation. Moreover, a majority of consumer preference studies have focused on the effect of experimental manipulations in situations where temporal proximity between manipulation and response is close. Investigating whether the experimental manipulations influence consumer preferences in the short term or can have persistent effects is critical in understanding whether procedure manipulations simply distort responses or can influence underlying preferences. Based on the findings of the three essays, theoretical and practical implications, limitations, and future research directions are discussed in CHAPTER 6.

CHAPTER 2.

ESSAY 1A: THE ROBUSTNESS OF ANCHORING EFFECTS ON PRODUCT VALUATIONS

The anchoring effect refers to the tendency for people's numeric judgments to be influenced by an initially considered value (Tversky & Kahneman, 1974). For example, when tasked with estimating the length of the Mississippi River (Jacowitz & Kahneman, 1995), depending on whether they are first asked whether the Mississippi River is longer or shorter than a smaller or larger "anchor" (70 vs. 2,000 miles), respondents will provide a smaller or larger estimate (median estimates 300 and 1,500). The anchoring effect has been studied in many domains, such as factual questions (Chapman & Johnson, 1994; Epley & Gilovich, 2001, 2005, 2006; Mussweiler & Englich, 2005; Mussweiler & Strack, 1999; Tversky & Kahneman, 1974; Wegener, Petty, Blankenship, & Detweiler-Bedell, 2010), physical and temporal length estimation (LeBoeuf & Shafir, 2006, 2009), answers to math questions (Smith & Windschitl, 2011), legal judgments (Mussweiler, 2001b), performance judgments (Thorsteinson, 2011; Thorsteinson, Breier, Atwell, Hamilton, & Privette, 2008), and purchase quantities (Wansink, Kent, & Hoch, 1998).

Anchoring effects have also been documented in a variety of valuation tasks (Ariely, Loewenstein, & Prelec, 2003; Bergman et al., 2010; Chapman & Johnson, 1999; Northcraft & Neale, 1987). Notably, Ariely et al. (2003) showed that, in consequential decision making situations where real money and products were at stake, people's valuations of consumer products were strongly and systematically influenced by numbers that should have no bearing on their own preferential judgments. Ariely and colleagues

showed several items to participants in a class and asked them whether they were willing to buy the given items for a dollar amount derived from the last two digits of their Social Security numbers (SSN, hereafter; see Figure 1). The same participants were also asked their willingness-to-pay (WTP, hereafter) using an incentive compatible Becker-DeGroot-Marschak (BDM, hereafter) mechanism (Becker, DeGroot, & Marschak, 1964). Importantly, the experimental items used were common consumer products, for which participants would be expected to have well-defined preferences, and the task was incentive compatible, so their responses were consequential. The results showed that people's WTP for the given items was strongly and systematically influenced by the anchoring number, although their relative valuations across the given items gave the appearance of coherence. This finding contradicts the claim that choices reflect stable underlying preferences, and supports theories of constructed preferences, according to which preferences are context and task dependent (Slovic, 1995).

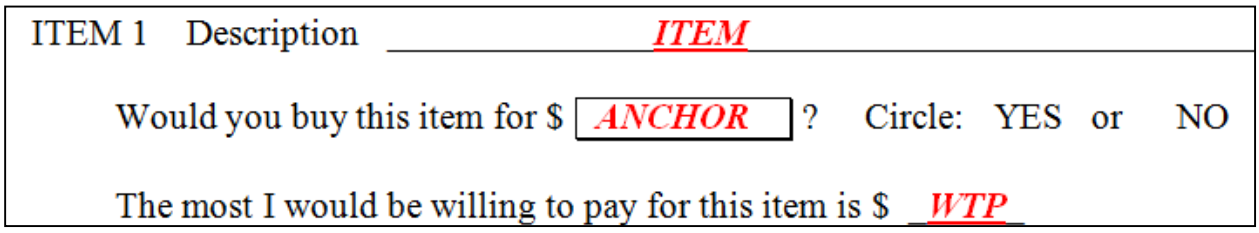

Figure 1. An example of questions for one item used in the task. Item names were provided by the experimenter. Participants had to write an anchoring number in the "ANCHOR" box across the six items before seeing the actual items. Each item was then presented sequentially, and participants were asked to answer both the anchor question (Yes or No) and the WTP question (dollar amount with no restriction).

Despite the theoretical importance of the result, empirical evidence for the anchoring effect on consumer product valuations has only been documented in a handful

of published studies. The recent massive replication of classic anchoring effects (for factual questions) across 36 different samples and settings showed stronger effects than the original study (Klein et al., 2015), whereas several other (non-anchoring) psychological effects were not replicable in the large-scale replication effort. These findings highlight the robustness of anchoring in factual judgments, and the importance of replication in experimental social science. In contrast, replications of anchoring for product valuations have been inconsistent, as failed replication attempts were recently published (e.g., Fudenberg et al., 2012).

In a set of three experiments and a cross-study analysis, we assess the replicability of anchoring effects for preferential judgments, and find that anchoring effects on WTP are robust to differences in experimental settings and procedures. Our review of prior studies reveals highly divergent results, but leaves open the possibility that the discrepancies were caused by procedural differences. We attempt to explain the divergence by systematically replicating the Ariely et al. (2003) study with several variations of experimental procedures. While previous studies report a wide range of results, our studies show strong anchoring effects near the middle of the range. However, we do not find that procedural differences could account for the differences in previous studies. Moreover, we introduce two new variations of the experiment that further establish the procedural robustness of the effect. In Study 3, we vary the anchoring number between items, within subjects, so that participants' attention to any particular anchoring number is weaker than in the classic anchoring paradigm. We also vary whether we instruct participants to consider the amount that they spent on similar

previous purchase experiences, to investigate the overriding effect of internal reference prices, yet still found significant anchoring effects.

Finally, combining our data with that from previous valuation studies, we report a cross-study analysis which shows that the anchoring effect is consistently found, yet the magnitude varies across studies. Additionally, a *p*-curve analysis (Simonsohn, Nelson, & Simmons, 2014) shows no indication that reported anchoring effects for consumer product valuations result from publication bias or other censoring, and a meta-analysis using effect sizes from previous studies confirms a significant anchoring effect on WTP. Comparing across studies, we do not find that procedural differences account for variation in the magnitude of anchoring effects. We do find that the percentile rank gap, a measure of the overlap in the ranges of anchor values and consumer valuations (proposed by Jung et al., 2016), affects the strength of the anchoring manipulation, but it does not fully account for variation across studies. This finding supports theories of anchoring where the effect requires that the distribution of participants' valuations coincides with the range of anchor values used.

Background

While anchoring effects have been shown in a variety of valuation tasks, relatively few focus on preferential judgments where participants report valuations for familiar consumer products. Bergman et al. (2010) found a significant anchoring effect for five out of six items, and Simonson and Drolet (2004) showed similar results with different items (e.g., toaster, cordless phone, backpack, and headphones). However, Fudenberg et al. (2012) did not find significant anchoring effects for any of six items;

similarly, they did not find an effect in several willingness-to-accept (WTA) studies. More recently, Maniadis et al. (2014) failed to replicate the anchoring effect on willingness-to-accept (WTA) for unpleasant hedonic experiences, casting further doubt on the robustness of anchoring effects on valuations. The authors suggested that anchoring effects are simply not robust in valuation tasks. Additionally, the effect of anchoring on preferential judgments in field experiments, where real payments were made in natural settings, have also shown weak or null effects (Alevy, Landry, & List, 2015; Jung et al., 2016).

Fudenberg et al. (2012) did not provide a specific explanation for why they obtain such different results, leaving open the possibility that it was due to differences in experimental protocol (in one study, they ruled out that the difference was due to whether participants were provided a thorough description of the BDM procedure, which did not change the non-significant anchoring effects). In particular, they used a different anchoring mechanism and recruiting method. While Ariely et al. (2003) used the last two digits of participants' SSN as the anchoring numbers, Fudenberg et al. (2012) had participants actively generate random numbers from a computer (RN, hereafter). One possibility is that the perceived randomness of an anchoring number influences people's inferences about the anchoring number (Frederick, Mochon, & Danilowitz, 2013). For example, in Experiment 3 of Chapman and Johnson (1999), 34% percent of participants reported that they considered SSN to be informative. In addition, a recent study showed that SSN can be predicted based on personal information, such as place and date of birth (Acquisti & Gross, 2009); even the perception that information is embedded in a person's SSN could cause inferences to be more prevalent than with randomly generated numbers.

Thus, the actively-generated numbers may have been more transparently random to the Fudenberg et al. (2012) participants, resulting in no anchoring effects. We test this hypothesis in our Study 1.

Separately, the premise for participation in the experiment could be driving differences across studies. Ariely et al. (2003) conducted the experiment in a classroom as a part of a class activity, but Fudenberg et al. (2012) recruited people from a participant email list and conducted the experiment in sessions dedicated to running the experiment in a laboratory. The classroom demonstration context may have provided enough of a "cover story" that participants were less likely to perceive the anchoring number as transparently irrelevant to the valuation task. In the lab study, participants may have been more attuned to the arbitrariness of the anchoring manipulation, and thus less likely to draw inferences about the relevance of the anchor. We tested this hypothesis in two ways: in Study 1, the experiment was conducted in both a classroom context and in a lab study using email recruiting. In Study 2, the experiment was conducted in a class, but the cover story was varied to frame the activity as a demonstration of either marketing research or academic research.

In addition to the differences in experimental procedures used in the previous studies, we investigated the moderating roles of knowledge from previous purchasing experience and repetition of an anchoring number (Study 3). Previous product valuation anchoring studies used a single anchoring number across multiple items. It is possible that repeating the same anchoring number multiple times increases people's attention to the anchoring number, leading to significant anchoring effects on product valuation. In Study 3, participants were exposed to different anchors for different items. Separately,

people might have experience purchasing similar items that could dilute the effect of an experimenter-given anchor in a valuation anchoring task. In Study 3, we vary whether we have participants think about their previous purchasing experiences for the same or similar items.

We also test the robustness of the anchoring effect on preferential judgments through a cross-study analysis (Study 4) that combines our data with data from prior studies. Of the prior studies, three used very similar procedures for WTP outcomes that should be considered direct replications (Ariely et al., 2003; Bergman et al., 2010; Fudenberg et al., 2012). As the anchoring number generation method and premise for participation are confounded among the prior studies, the addition of data from our first two studies is needed to disentangle these factors. A summary of these studies is presented in Table 1.[1] Among the studies, only Bergman et al. (2010)'s study was conducted in a country other than the United States (Sweden), and it also used a different currency (Swedish krona; $1 was around SEK 6 at the time the experiment was conducted). The other two studies were conducted in the United States and used dollars as the monetary unit. In the original study, the experiment was conducted in an MBA class and the BDM procedure was reconciled using participant's own money. The experiment was introduced as a classroom demonstration. For the other two studies, however, undergraduate students were recruited, and they were endowed with money by the experimenter when resolving the BDM procedure. These studies were explicitly introduced as experiments.

[1] We thank the authors of these studies for making their data available. Simonson and Drolet (2004) also conducted an anchoring effect study using WTP tasks with slightly different items, but with a similar procedure. However, as they did not use an incentive compatible method and decisions were hypothetical (only one study investigated the effect of anchoring on WTA using an incentive compatible method), we excluded their study from this analysis.

Table 1. Comparison of anchoring studies on valuation

	Ariely et al (2003)	Bergman et al (2010)	Fudenberg et al (2012)	Current Article
Participants	MBA	Undergraduate	Undergraduate	Undergraduate
N of participants	55	116	78	392
Items	Cordless Mouse	Radio Transmitter for mp3 Players	Cordless Mouse	Cordless Mouse
	Cordless Keyboard	Chocolate Truffles	Cordless Keyboard	Bluetooth Keyboard
	Average Wine	Average Wine	Academic Planner	Normal Dinner
	Rare Wine	Rare Wine	Financial Calculator	Fancy Dinner
	Design Book	Design Book	Design Book	Art Book
	Belgian Chocolates	Belgian Chocolates	Milk Chocolates	Belgian Chocolates
Cash Endowment for Winners	No	-	Yes	Yes
Anchoring Number	SSN	SSN	RN	SSN, RN
Cover Story	Class Activity	Experiment	Experiment	Class Activity, Experiment
Location (currency)	United States ($)	Sweden (SEK)	United States ($)	United States ($)
Purpose	Susceptibility to irrelevant anchor in valuation task and relative preference order	The effect of cognitive ability on the anchoring effect	Robustness of anchoring effect in willingness-to-accept and willingness-to-pay	Reconciliation of previous anchoring effect studies

Study 1. The Effect of Recruitment and Anchoring Number Types

Design and Participants

Our first study examined whether the different experimental procedures used in previous studies contributed to the inconsistency of their results. We used a 2 × 2 between subject design: recruiting method (classroom or email), and anchoring number types (SSN or RN anchoring). We were not necessarily predicting an interaction effect; rather, we were exploring whether one or both factors together could account for the differences in previous findings, through some combination of the mechanisms described

above. A total of 116 university undergraduate students participated in our study, receiving class credit or $5 in exchange for their participation. Participants in the classroom condition (N = 66; SSN = 26 and RN = 40) were undergraduate students taking Marketing Research classes, while participants in the laboratory condition (N = 50; SSN = 22 and RN = 28) were recruited using a SONA system, an electronic platform developed to recruit participants, rather than through the class. Participants were not asked to provide their demographic information.

Procedure

We conducted the experiment using an incentive compatible BDM mechanism. Before the experiment, participants were informed that they would have a chance to buy one of the six items based on their answers. We used a similar item composition to the original (Ariely et al., 2003) study's (cordless trackball, cordless keyboard, normal wine, fancy wine, art book, and chocolates), except that we substituted a cordless mouse for the trackball, and two different types of dinner coupon (normal and fancy dinner coupons) for the two wines (since we could not sell alcohol to the participants). Then, we showed the six items and, for each item, asked participants for three responses in the following order: whether they want to buy the given items for the random dollar amount (Yes/No response), the highest price they are willing to pay, and their estimate of the item's market price. In the SSN condition, the anchoring number was chosen by transforming the last two digits of the participant's SSN into a dollar amount, and in the RN condition, random numbers were generated by drawing from a deck of 100 numbered cards with a range of 0 to 99. After responding to all the questions, the winners (who were endowed

with enough additional funds ($80 in this study) to cover the purchase of the items), item, and resolution method (e.g., based on the Yes/No or WTP response) were randomly decided by rolling dice.

Result

We tested for the presence of anchoring effects using Pearson correlation coefficients, as used in previous studies (Ariely et al., 2003; Bergman et al., 2010; Fudenberg et al., 2012). As seen in Table 2, for the four similar items, excluding two dinner coupons, the Pearson correlation coefficients between anchoring number and WTP ranged from .12 to .29. The range of the Pearson correlation coefficients was from 0.32 to 0.52 in Ariely et al. (2003), and -0.04 to -0.01 in Fudenberg et al. (2012). Our results replicated the anchoring effect at a strength near the middle of the range for previous studies. We also examined the ratio of average WTP for participants in the highest quintile of anchoring numbers to those in the lowest quintile and compared the magnitudes with prior studies (Table 2). The range of this WTP ratio across items in our study was 1.26 to 1.81, which is also near the middle of prior studies (the ratio in Ariely et al. (2003)'s study was 2.16 to 3.46 and that in Fudenberg et al. (2012)'s study was 0.77 to 0.92).

Table 2. Average stated WTP (standard deviation) sorted by quintile, Pearson's r, and rank-sum test result in Study 1

Quintile	Art Book	Bluetooth Keyboard	Belgian Chocolates	Cordless Mouse	Fancy Dinner	Normal Dinner
1 (N=25[a]) 0-19	$13.71 (9.00)	$36.66 (31.12)	$9.58 (4.31)	$20.00 (13.71)	$27.62 (11.95)	$20.06 (8.15)
2 (N=22) 20-39	$15.82 (10.35)	$30.95 (12.79)	$14.39 (10.36)	$17.82 (10.48)	$42.27 (22.08)	$28.05 (12.18)
3 (N=23) 41-65	$26.52 (21.92)	$47.92 (22.17)	$13.74 (8.05)	$24.39 (13.86)	$43.13 (16.13)	$30.87 (14.82)
4 (N=23) 66-78	$27.04 (30.56)	$50.26 (28.34)	$13.57 (8.61)	$26.39 (20.79)	$51.87 (23.35)	$36.74 (15.35)
5 (N=23) 79-99	$24.78 (14.13)	$48.48 (31.89)	$12.04 (5.89)	$29.91 (20.84)	$51.35 (27.55)	$35.52 (23.01)
Average	$21.56	$42.85	$12.62	$23.69	$42.99	$30.09
Pearson's *r*	.29	.26	.12	.26	.41	.37
p-value	.002	.005	.191	.006	< .001	< .001
Ratio (fifth/first)	1.81	1.32	1.26	1.50	1.86	1.77
Rank Sum Test *p*-value	.003	.043	.137	.040	< .001	.003
ALP Ratio	2.34	3.46	2.16	3.03		
BEJS Ratio	2.68		2.11			
FLM Ratio	0.81	0.92	0.87	0.77		

[a] *The number of responses in Art Book and Belgian Chocolates is 24 because of nonresponse.*

Note: Fudenberg et al. (2012) slightly changed the items used because of legal constraints (wine cannot be sold to people below 21 years old in the USA) and participant preferences (trackballs are not popular among current undergraduates). We substituted fancy and normal dinners for the fancy and normal wines, and used cordless mice instead of trackballs.

To investigate the effects of the different anchoring mechanisms and recruiting methods, we regressed the WTP responses on the anchoring values interacted with the experimental conditions (Table 3). The main effect of the anchoring number was statistically significant in all specifications. As one unit of anchoring number increases, people's WTP increases $0.19 (Table 3, model 2). The main effects of the random

number and recruiting method were not significant, nor were there significant interactions with the anchoring number (Table 3). Not only do these results fail to explain the difference between previous studies, but the difference between recruiting conditions appears to run in the opposite direction: the anchoring effect appears mildly stronger in the email recruiting conditions than in the classroom conditions.

A single experiment offers limited evidence for a null result. To provide some context for our findings, we examined the statistical power to detect a significant interaction between the anchor and procedural variations. As our objective was to reconcile divergent previous results, we computed statistical power to detect a difference in slopes equivalent in magnitude to the difference in anchoring effects between Ariely et al. (2003) and Fudenberg et al. (2012), which was 0.25. For α (the allowable false-positive rate) of 5%, the power to detect a significant interaction was 71% for a single item, and was 98% when we account for 6 items per participant.[2] Thus, *ex ante* our study had ample opportunity to detect an interaction of large magnitude. We also examined the confidence intervals for the interactions (anchor×anchor type and anchor×recruit type), but they did not exceed 0.1 (95% CI for anchor × anchor type: -0.03 to 0.07; CI for anchor × recruit type: -0.01 to 0.09). While our results are consistent with a range of values for procedural effects that includes small effect sizes, they are inconsistent with large effects that could account for divergent results between previous studies.

[2] Power was computed using the G*Power 3 power analysis program (Faul, Erdfelder, Lang, & Buchner, 2007). The effect size to detect was computed by running a regression of WTP on anchoring number, using the Ariely et al. (2003) data combined with the Fudenberg et al. (2012) data. The model included item fixed effects and a dummy variable interacted with the anchor; the coefficient for the interaction term was used for our power calculations, and the standard deviation of the residuals was used as an estimate for the standard deviation of the error term. Since a participant's responses to multiple items may have unknown correlation structure, we treated the 6-item power computation as the ability to detect an interaction in a model for the participant-level average WTP (across 6 items); this reduces the estimated standard deviation of the error term without assuming responses within a participant are i.i.d. (statistical power computed between subjects, with *N* remaining at 116).

Table 3. Regression analysis for Study 1.

	(1) No Item FE	(2) Item FE	(3) Anchor Type	(4) Recruit Type	(5) Both
Anchor	0.19***	0.19***	0.19***	0.19***	0.19***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Anchor Type			0.21		0.17
			(1.48)		(1.49)
Anchor×Anchor Type			0.02		0.02
			(0.03)		(0.03)
Recruit Type				-0.20	-0.28
				(1.43)	(1.49)
Anchor×Recruit Type				0.04+	0.04+
				(0.02)	(0.03)
Anchor Type×Recruit Type					0.16
					(1.49)
Anchor×Anchor Type×Recruit Type					0.01
					(0.03)
Intercept	19.34***	Item FE	Item FE	Item FE	Item FE
	(1.66)				
N	694	694	694	694	694

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: Standard errors in parenthesis. Models (2) through (5) include fixed effects for each item. Anchor Type is coded as -1 for SSN and 1 for RN. Recruit Type is coded as -1 for classroom and 1 for email recruiting. Standard errors clustered on participant do not affect the pattern of results.

Study 2. The Effect of Cover Story

Our second experiment investigated the effect of two different cover stories on anchoring effects. In the original study (Ariely et al., 2003), the authors informed participants that they would learn how to construct demand curves using two different methods. Participants were informed that one is based on yes/no questions with different random prices and the other is based on directly eliciting WTP. In Fudenberg et al. (2012)'s study, however, it was more salient to participants that they were participating in

an experiment, and the different expectations in this setting may have affected people's tendency to make inferences about the random number (Frederick et al., 2013). In this study, we investigated whether differences in cover story contribute to inconsistent anchoring effects in product valuations.

Design and Participants

We had two quasi-random groups in this study, to whom the procedure was described as either a class activity or an experiment. We conducted the study in two consecutive sessions of a Marketing Research class, with the first session assigned to the class activity condition and the second session assigned to the experiment condition. This order was randomly determined by tossing a coin. Participants in the classroom activity condition were informed that they would learn how to elicit demand information, while participants in the experiment group were informed that they would participate in an experiment. A total of 82 university undergraduate students participated in this study, and were randomly assigned to either the classroom activity condition ($N = 44$) or the experiment condition ($N = 38$). Both groups were taking the Marketing Research class with the same instructor and they were not asked to provide any demographic information.

Procedure

Overall the procedure was identical to the RN-Classroom condition in Study 1, except we substituted a pair of headphones for the chocolates. Both groups were debriefed and shown how to generate demand curves using the data from Study 1.

Results

We tested for the presence of anchoring effects using Pearson correlation coefficients as in Study 1. Four out of six items in Study 2 showed a significant correlation between anchoring number and WTP, with one additional item marginally significant (Table 4). To examine whether WTP in the fifth quintile is greater than that in the first quintile, we ran a rank-sum test, and found that four out of six items showed that WTP in the fifth quintile was significantly greater than that in the first quintile.

Table 4. Average stated WTP (standard deviation) sorted by quintile, Pearson's r, and rank-sum test result in Study 2.

Quintiles	Art Book	Bluetooth Keyboard	Cordless Mouse	Fancy Dinner	Headphone	Normal Dinner
1 (N = 17) 0-20	$9.82 (7.85)	$27.35 (13.40)	$16.88 (7.35)	$30.29 (12.05)	$52.29 (38.54)	$25.24 (10.50)
2 (N = 16) 21-38	$21.56 (21.79)	$45.31 (16.68)	$25.69 (7.72)	$50.31 (18.48)	$48.62 (46.12)	$30.31 (8.46)
3 (N = 18) 39-58	$15.94 (10.09)	$37.78 (17.25)	$24.89 (10.39)	$54.56 (15.83)	$39.44 (21.07)	$38.78 (16.03)
4 (N = 15) 59-71	$25.67 (13.61)	$53.00 (17.40)	$28.67 (11.41)	$67.00 (15.33)	$80.67 (86.25)	$46.33 (15.86)
5 (N = 16) 72-95	$21.56 (29.40)	$55.19 (28.49)	$24.69 (10.46)	$49.69 (26.74)	$66.06 (39.49)	$41.25 (15.44)
Average	$18.65	$43.27	$24.04	$50.02	$56.63	$36.18
Pearson's r	.21	.42	.26	.39	.13	.46
p-value	.056	< .001	.018	< .001	.240	< .001
Ratio (fifth/first)	2.20	2.02	1.46	1.64	1.26	1.64
Rank sum	-1.26	-2.99	-2.26	-2.41	-1.1	-3.07
p-value	.207	.003	.024	.016	.271	.002

To test for the moderating effect of the cover stories, we regressed WTP on anchoring number, interacted with an indicator for cover story. The regression shows that

the main effect of the anchoring number was statistically significant, but the interaction effect between anchoring number and cover story was not statistically significant (Table 5). This result provides further evidence for anchoring effects in product valuations, but varying the cover story provided to participants did not moderate the anchoring effect.

We again computed statistical power to detect the difference in anchoring effects between Ariely et al. (2003) and Fudenberg et al. (2012). The power to detect a significant interaction was 51% for a single item, while the power to detect a significant interaction was 87% when we account for 6 items per participant. The 95% confidence interval for the anchor × cover story interaction also did not exceed 0.1 (95% CI: -0.10 to 0.08), suggesting that any effects of cover story are small.

Table 5. Regression analysis for Study 2.

	(1) No Item FE	(2) Item FE	(3) Cover Story
Anchor	0.24***	0.24***	0.25***
	(0.05)	(0.05)	(0.05)
Cover Story			0.16
			(2.41)
Anchor×Cover Story			-0.01
			(0.05)
Intercept	27.00***	Item FE	Item FE
	(2.66)		
N	492	492	492

*** $p < .001$

Note: Standard errors are in the parentheses. Model (2) and Model (3) include fixed effects for each item. Cover story is coded as -1 for Experiment and 1 for Class activity.

Study 3. Within-Subject Anchoring and Self-Generated Anchors

Our first two studies investigated whether procedural differences accounted for differences in previously reported studies that used the same experimental paradigm. To

continue probing the strength of the original study's theoretical claims, we also considered whether other characteristics of the procedures used could account for the observed effects. Previous preference anchoring replication studies and our first two experiments used one random number per subject, which was either the last two digits of their SSN or a single random number, and participants answered whether they want to buy each of the six items for that one random price. Participants were instructed to copy the number six times, before answering any questions about the items (Figure 1). Thus, the repeated processing of the same number might have increased attention towards the anchoring numbers and contributed to the anchoring effects (rather than the evaluation of the anchor as a possible price). For example, Wilson, Houston, Etling, and Brekke (1996) showed, using factual questions, that sufficient attention towards random numbers is a necessary condition for anchoring effects. In their experiment, participants who copied anchoring numbers five times showed stronger anchoring effects than those who copied the anchoring numbers once, and there was no difference in response between the no-anchor group and the group which copied the anchoring numbers once (but also see Brewer & Chapman, 2002, who argue such "basic" anchoring effects are fragile).

If the repetition of the anchoring number is necessary to obtain the large anchoring effects previously observed, it could be argued that the effect is a procedural artifact, and does not provide strong evidence for the arbitrariness of preferences. We also note that, while this feature of the procedure did not vary across prior studies, the different participant populations and their respective experimental settings could have generated differences in attention to the anchoring number. In Study 3, we investigated

whether the anchoring effect would be lessened by eliminating the repetition in anchoring numbers; participants in this study received a different anchoring number for each item.

We also test for the possible moderating effect of self-generated anchors. In addition to incentive compatibility, another characteristic that distinguishes valuation tasks from other numeric judgment tasks might be the role of previous experience. Epley and Gilovich (2001) showed that participants could use a self-generated anchor when they are asked to estimate the answer to questions for which they have easily accessible standards of comparison. For example, when estimating the year George Washington was elected as president, they could use the information that the answer is later than the year the U.S. declared independence (self-generated anchor, in 1776), and adjust from the self-generated anchor. In preference anchoring tasks, participants might use the information from their experience of purchasing the same item or items in the same item category. Thus, this self-generated anchor might dilute the effect of an experimenter-given anchor (Fudenberg et al., 2012). In Study 3, we tested whether considering prices at which they have purchased the same or similar items previously could impact the effect of random experimenter-given anchoring on WTP.

Procedure

A total of 216 university undergraduate students were recruited for the experiment, receiving class credit for participation (M_{age} = 20.10 years old, SD = 0.15, female = 126). The overall procedure was identical with the email recruiting-RN condition of Study 1, except that we replaced the dinner coupons and the chocolates with different consumer products (humidifier, electric toothbrush, and vacuum cleaner), and that participants

selected different random numbers for each of the six items. The anchoring numbers were tested at discrete levels: each item had its own deck of 30 cards in which there were 10 low anchor, 10 middle anchor, and 10 high anchor cards. The low anchors were below the 10^{th}-15^{th} percentile, the middle anchors were at the 50th percentile, and the high anchors were at the 80^{th}-95^{th} percentile of WTP from one of our previous experiments (Table 6). Participants selected one card for each item from the corresponding deck. Before starting the task, participants were randomly assigned to either a self-generated anchor group (N = 114) or a control group (N = 102). Participants in the self-generated group (assigned at the experimental session level, so that procedures remained the same for all participants in a given session) were asked to think about any experience they may have had of purchasing the six items (or similar items), and were asked to state the amount they paid at that time, while those in the control group were not. The same incentive compatible procedures were used. At the end of the task, participants were asked to provide demographic information and complete two individual difference measures: need for cognition (Cacioppo & Petty, 1982) and the cognitive reflection test (Frederick, 2005).

Results

Among them, 21 participants selected anchoring numbers from the wrong decks, and one participant did not provide answers for the self-generated anchor questions, so we excluded these participants' responses from the analysis. We also excluded four responses where WTP was missing. Before examining the effect of self-generated anchors, we checked whether the anchors affected the price information people used in

the self-generated anchor condition. A one-way analysis of variance (ANOVA) for each item did not find any differences in self-generated anchors among the three anchor conditions (Table 6). Thus, it does not appear that participants in different anchor conditions used different self-generated anchors in the task.

Table 6. Anchors and summary statistics for the self-generated anchors in Study 3

	Anchors			Self-Generated Anchors					
	Low	Middle	High	Low	Middle	High	$F(2, 101)$	p-value	η^2
Art Book	\$5	\$10	\$28	\$20.53 (33.86)	\$18.03 (14.20)	\$25.60 (25.05)	0.81	.448	.02
Humidifier	\$5	\$20	\$36	\$46.14 (25.18)	\$71.12 (67.03)	\$57.14 (50.62)	2.10	.127	.04
Bluetooth Keyboard	\$8	\$20	\$40	\$45.30 (21.61)	\$45.94 (21.16)	\$51.93 (30.10)	0.79	.458	.02
Cordless Mouse	\$5	\$15	\$30	\$23.21 (13.86)	\$29.86 (47.59)	\$24.94 (9.35)	0.50	.611	.01
Electric Toothbrush	\$10	\$30	\$70	\$22.79 (18.50)	\$32.90 (26.59)	\$31.93 (30.67)	1.61	.204	.03
Vacuum Cleaner	\$10	\$25	\$63	\$57.08 (42.57)	\$57.15 (41.53)	\$54.72 (30.52)	0.05	.955	.0009

Note: There was one missing responses for Art Book and Vacuum Cleaner.

Since we used discrete anchors, we ran a 2 (self-generated anchor: no vs. yes) × 3 (anchor: low vs. middle vs. high) factorial analysis of variance (ANOVA) for each item for the analysis, excluding four responses where WTP was missing. Five out of six items showed a significant main effect of anchoring, but did not show a significant interaction between anchoring and the use of self-generated anchor (Table 7). A post-hoc analysis using a Tukey Honest Significant Difference (HSD) test showed that WTP in the high anchor condition was greater than WTP in the low anchor condition across all the five items that showed a significant main effect for the anchoring number at the 95% confidence level.

The sixth item (the humidifier) also showed a significant main effect of anchoring, but the patterns were slightly different between the self-generated anchor condition and the no self-generated anchor condition. A post-hoc analysis using a Tukey HSD test showed that WTP in the high anchor condition was greater than WTP in the low anchor condition in the no self-generated anchor condition, whereas there was no difference in WTP between the low and high anchor conditions in the self-generated anchor condition. Additional comparisons for the self-generated anchor condition revealed that WTP in the middle anchor condition was greater than WTP in the low anchor, but there was no difference in WTP between the middle and the high anchor conditions. Thus, the significant interaction between anchoring and the use of self-generated anchor for the humidifier resulted from a higher WTP in the middle anchor condition in the self-generated anchor condition. We note that the middle anchor group reported the highest reference price in the self-generated anchor condition, and although the difference in self-generated anchors was not significant across anchoring conditions (the self-generated anchoring task was completed before the experimental anchoring task), this may have contributed to the different pattern observed for this item.

We also regressed WTP on the anchoring number interacted with the use of self-generated anchor (coded -1 for the no self-generated condition and 1 for the self-generated condition) with item fixed effect, treating the anchor as a continuous variable. The average anchoring effect is significant ($b = 0.27$, $SE = 0.03$, $t = 10.43$, $p < .001$), with a magnitude slightly larger than in Study 1 ($b = 0.19$) and Study 2 ($b = 0.24$). However, neither the main effect of the use of self-generated anchor ($b = 1.05$, $SE = 0.71$, $t = 1.48$, $p = .140$) nor the interaction between anchor and the use of self-generated anchor ($b =$ -

0.02, $SE = 0.02$, $t = -1.00$, $p = > .250$) were significant. Thus, we conclude that neither varying the anchoring number across items for a participant nor prompting participants to use previous experience or knowledge as self-generated anchors reduces or eliminates the anchoring effect.

Table 7. Summary statistics and ANOVA results for Study 3

			WTP		Anchor	Condition	Interaction
Item	Self-Generated	Low	Middle	High	$F(2, 188)$	$F(1, 188)$	$F(2, 188)$
Art Book	No	\$9.62	\$11.36	\$17.96	15.19	0.79	0.42
		(6.68)	(6.47)	(11.04)	$p < .001$	$p = .376$	$p = .659$
	Yes	\$10.09	\$14.13	\$17.85	$\eta^2 = .139$	$\eta^2 = .004$	$\eta^2 = .004$
		(6.58)	(9.25)	(10.14)			
Humidifier	No	\$29.83	\$28.85	\$40.40	3.31	0.17	4.96
		(17.42)	(11.56)	(16.88)	$p = .039$	$p = .680$	$p = .008$
	Yes	\$27.80	\$36.97	\$32.05	$\eta^2 = .033$	$\eta^2 = .0008$	$\eta^2 = .049$
		(12.73)	(20.31)	(11.30)			
Bluetooth Keyboard	No	\$28.07	\$32.93	\$34.32	4.65	1.09	2.13
		(11.57)	(14.85)	(11.38)	$p = .011$	$p = .297$	$p = .122$
	Yes	\$33.82	\$29.14	\$39.45	$\eta^2 = .046$	$\eta^2 = .005$	$\eta^2 = .021$
		(13.64)	(8.87)	(17.35)			
Cordless Mouse	No	\$15.67	\$18.13	\$21.84	13.85	3.67	0.74
		(7.93)	(8.44)	(9.82)	$p < .001$	$p = .057$	$p = .480$
	Yes	\$16.42	\$20.11	\$26.22	$\eta^2 = .126$	$\eta^2 = .017$	$\eta^2 = .007$
		(6.81)	(6.03)	(11.91)			
Electric Toothbrush	No	\$22.63	\$30.86	\$40.26	8.97	0.09	0.12
		(18.31)	(17.08)	(26.26)	$p < .001$	$p = .762$	$p = .891$
	Yes	\$23.24	\$30.27	\$37.35	$\eta^2 = .087$	$\eta^2 = .0004$	$\eta^2 = .001$
		(13.04)	(19.61)	(27.91)			
Vacuum Cleaner	No	\$33.97	\$37.44	\$51.43	15.47	0.09	0.28
		(21.61)	(14.51)	(20.73)	$p < .001$	$p = .761$	$p = .758$
	Yes	\$34.56	\$40.80	\$50.42	$\eta^2 = .142$	$\eta^2 = .0004$	$\eta^2 = .003$
		(13.45)	(13.83)	(18.72)			

Note: There was one missing responses in Art Book and Vacuum Cleaner and two missing responses in Humidifier.

We also examined individual differences in preference anchoring. Summary statistics are presented in Appendix A. Consistent with Bergman et al. (2010), there was no strong evidence that participants with a higher CRT score are less susceptible to anchoring in preferential judgments. In only one of the six items, participants with a

higher CRT showed smaller anchoring effects on WTP. Similarly, we did not find a significant moderating effect of NFC on the relationship between anchor and WTP.

Study 4. A Cross-Study Analysis

Our previous three studies showed that the anchoring effect is robust to variation in anchoring number type, recruiting method, cover story, repetition of the anchoring number, and the salience of previous purchasing experiences. In this section, we compare the effect sizes from multiple anchoring studies to test for differences across studies and test for factors that may help explain any such differences.

First, we pooled the data from the studies listed in Table 1 with the data from our Study 1 and Study 2, and ran a linear regression analysis. We ran a linear regression with study/item fixed effects and the interaction between anchor and studies. We used dummy variables as indicators for each study, setting the data from our Study 1 and Study 2 as the baseline. Discrete anchors were used in Study 3, and the additional factor that we tested (the prompt to use a self-generated anchor) was unique to this study, so this factor could not explain differences across previous studies. Thus, we did not include Study 3 in this regression, but we do include Study 3 in a subsequent meta-analysis based on *z* scores, to help characterize the distribution of effect sizes. Since one study was conducted using a different currency and the studies were conducted in different years and locations, we also standardized anchoring number and WTP within each study.

The anchoring effect from our two studies is statistically significant ($b = 0.23$, $SE = 0.03$, $t = 9.13$, $p < .001$; Model 1 in Table 8), and there was no difference between our studies and Bergman et al. (2010)'s study ($b = 0.03$, $SE = 0.04$, $t = 0.82$, $p = .412$). The

anchoring effect for Ariely et al. (2003)'s study was greater than our result (b = 0.12, SE = 0.05, t = 2.17, p = .030), although additional testing found that their effect was not significantly different from that in Bergman et al. (2010) (b = -0.08, SE = 0.06, t = -1.44, p = .150). Fudenberg et al. (2012)'s study showed a significantly smaller anchoring effect than our two experiments (b = -0.24, SE = 0.05, t = -5.09, p < .001), and the effect was not different from zero (b = -0.01, SE = 0.04, t = -0.30, p = .767). Further testing found that their effect was smaller than every other reported study (vs. Ariely et al: b = 0.36, SE = .06, t = 5.75, p < .001; vs. Bergman: b = 0.28, SE = 0.05, t = 5.31, p < .001). We also note that when analyzing each individual item for each experiment, Ariely et al. (2003) is the only study in which all items (six out of six) had a significant anchoring effect individually, and Fudenberg et al. (2012) is the only study in which none of the items had a significant effect. In all other experiments, one or two individual items had marginally significant or non-significant effects, which is not surprising for reasonably powered experiments.

We also examined whether the combined data could help identify factors that moderate the anchoring effect. We first retested whether the experimental procedures tested in our studies (recruit type, anchor type, and cover story) moderated the anchoring effect on WTP when incorporating additional studies that varied on these procedures. We did not find a significant interaction between the anchor and any of these procedural variables (Models 2, 3, and 4 in Table 8).

Table 8. A cross-study analysis result

	(1) Study Effect	(2) Recruit Type	(3) Anchor Type	(4) Cover Story	(5) Percentile Gap
Anchor	0.233***	0.254***	0.230***	0.229***	0.214***
	(0.026)	(0.029)	(0.031)	(0.032)	(0.026)
Anchor × Ariely et al (2003)	0.119*	0.148**	0.126	0.147*	0.190**
	(0.055)	(0.057)	(0.072)	(0.074)	(0.061)
Anchor × Bergman et al (2010)	0.0344	-0.037	0.042	0.015	0.020
	(0.042)	(0.060)	(0.063)	(0.044)	(0.042)
Anchor × Fudenberg et al (2012)	-0.244***	-0.315***	-0.245***	-0.264***	-0.155***
	(0.048)	(0.064)	(0.050)	(0.050)	(0.059)
Anchor × Recruit Type		0.050			
		(0.029)			
Recruit Type		0.133***			
		(0.030)			
Anchor × Anchor Type			0.004		
			(0.031)		
Anchor Type			-0.028		
			(0.030)		
Anchor × Cover Story				-0.024	
				(0.032)	
Cover Story				-0.100**	
				(0.032)	
Anchor × Percentile Gap					0.003**
					(0.001)
Percentile Gap					-0.008
					(0.005)
Intercept	Item FE	Item FE	Item FE	Item FE	Item FE
N	2,680	2,680	2,680	2,680	2,680

** p < .05, ** p < .01, *** p < .001*
Recruit type (0: classroom vs. 1: email recruiting), anchor type (0: SSN vs. 1: RN), and cover story (0: experiment vs. 1: class activity) are dummy coded.

In an additional analysis, we examined whether the distributional gap between anchors moderated the inconsistent anchoring effects on WTP. In a recent paper, Jung et al. (2016) posited that the range of anchors used in the study might moderate anchoring effects on WTP. In a series of field experiments, the authors showed that the percentile

rank gap between anchor values in the corresponding WTP distribution predicts significant anchoring effects on WTP. Intuitively, a large percentile rank gap means the anchor values span a wide range of the distribution of valuations, suggesting they also span a range of plausible prices, increasing the anchoring effect. Conversely, anchors that do not have a meaningful difference in WTP percentiles may have a weaker effect; for example, a ceiling effect where the provided anchor values are all well above participants' WTP may not differ in their effect on responses, so no anchoring effect would be detected (as in Chapman & Johnson, 1994). Unlike Jung et al. (2016)'s studies, the anchoring studies introduced in this paper used continuous anchors, so we operationalize percentile rank gap using the interquartile range, the difference in percentiles corresponding to $25 and $75 within the WTP distribution for each item and study.

We regressed WTP on anchor interacted with the percentile rank gap and with study and item fixed effects. The result showed a significant interaction between anchor and the percentile rank gap ($b = 0.003$, $SE = 0.001$, $t = 2.59$, $p = .010$), indicating that the anchoring effect on WTP was stronger when the percentile rank gap between utilized anchors was greater. Consistent with previous results, we found a significant main effect of anchoring on WTP ($b = 0.21$, $SE = 0.03$, $t = 8.09$, $p < .001$), and the main effect of the percentile rank gap of anchors was not significant ($b = -0.008$, $SE = 0.005$, $t = -1.46$, $p = .143$). Finally, we note that comparing the coefficients for each study across models, it appears that the percentile rank gap could contribute a small amount to the difference between our results and Fudenberg et al. (2012). This finding indicates that the distribution of valuations relative to the range of anchors moderates the anchoring effect, but on its own does not account for the observed differences across studies.

The cross-study analysis shows that anchoring studies consistently find an effect, with Fudenberg et al. (2012) providing the only exception. Our analysis does not rule out the potential effects of publication bias or other selective reporting. However, an inspection of the *p*-values across studies reveals a strong right skew that is inconsistent with selective reporting. A *p*-curve, as proposed by Simonsohn et al. (2014), is depicted in Figure 2, and the visual pattern is quite clear. Related diagnostic tests indicated that the observed anchoring effects were not driven by publication bias or selective reporting: the observed *p*-curve was significantly right skewed (binomial test: $p < .001$), was not flatter than a null *p*-curve generated by tests with 33% power (binomial test: $p > .99$), and was not left skewed (which is one proxy of checking selective report; binomial test: $p > .99$). The average estimated statistical power was 90%. Additionally, a funnel plot indicated that the distribution of effect sizes was not indicative of selection bias (Figure 2).

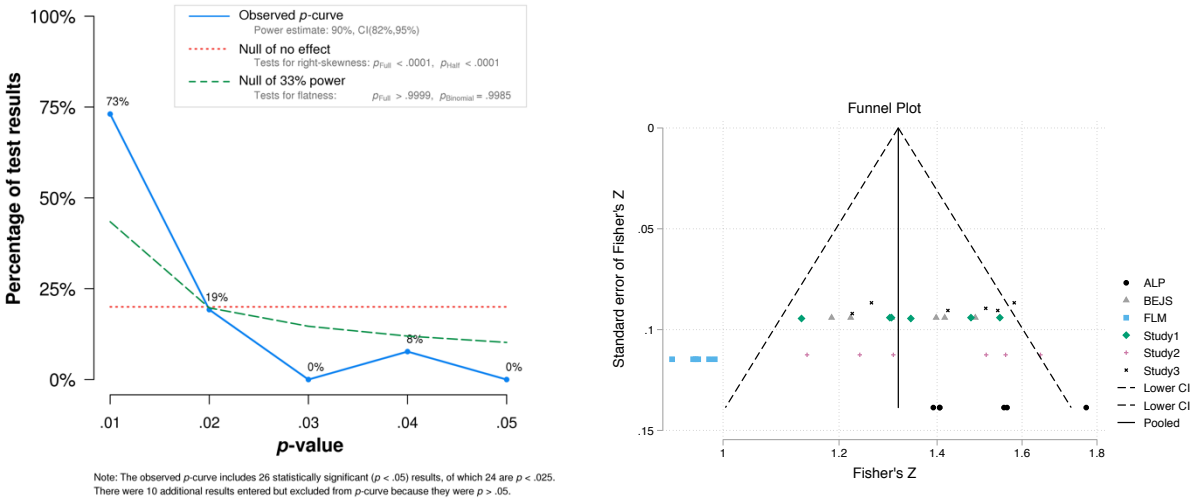

Figure 2. The results of a *p*-curve analysis (left) and funnel plot (right). P-values were from the Pearson correlation for each item in each study. Since discrete anchors were used in Study 3, *p*-values were from the main effect of anchor in the ANOVA results for each item. In the funnel plot, the effect sizes (Cohen's *d*) for Study 3 was calculated based on the *t*-tests between low and high anchor conditions for each item and transformed to *z* scores.

We also conducted a meta-analysis using the Pearson correlation between WTP and anchoring numbers for each item in each study, and the Pearson correlations were converted to the Fisher's Z scale for the analysis (Figure 3). Since Study 3 used discrete anchors, we transformed Cohen's *d* calculated from a *t*-test for each item between low and high anchor conditions into Fisher's Z scale. For the estimation, we used '*metafor*' package in R software and estimated the average effect by using a multilevel random effect model to reflect possible dependence across six items within each study (Konstantopoulos, 2011; Viechtbauer, 2010). Consistent with the cross-study regression presented above, the meta-analysis showed that, for the 30 items from five studies, the effect of anchoring on WTP was significant (average Fisher's Z = 0.27, *SE* = 0.06, *z* = 4.23, $p < .001$, transformed $r = .27$). This analysis further highlights the robustness of the anchoring effect, while also showing how the effect varies across items and studies.

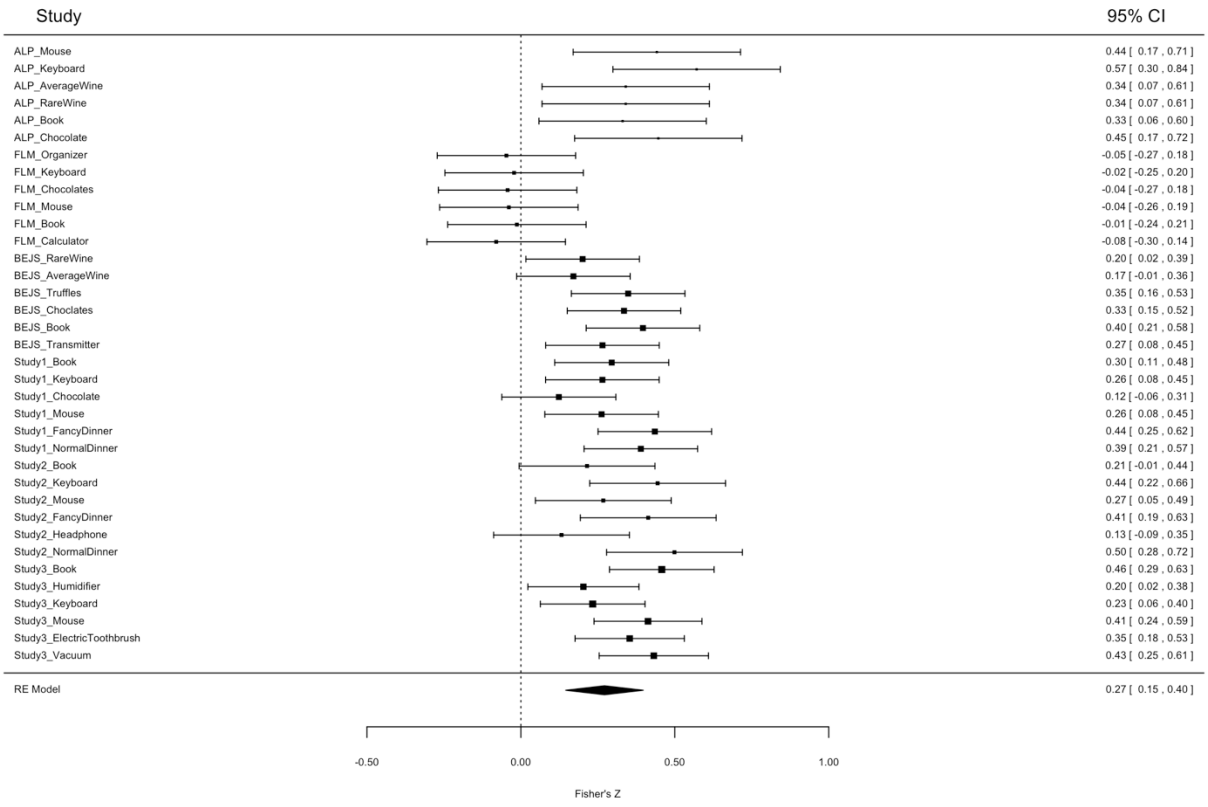

Figure 3. A meta-analysis result (mean and 95% confidence interval of Fisher's Z are in the right end of the figure).

To summarize the findings of our cross-study analysis, the differences between studies appear to be too large to attribute to chance, though from our Study 1 and Study 2, the observed procedural differences do not account for the gaps. The average effects are smaller than in the original Ariely et al. (2003), but still of an economically meaningful magnitude: an unstandardized regression estimated an average marginal effect where a change of one unit in the anchoring number changed WTP by 0.23 in the corresponding currency. The distribution of the test statistics across studies suggests that, by and large, the studies were sufficiently powered, with no indication of selective reporting. We conclude that the smaller effects typically observed (relative to the original study) are not

so small that they invalidate the main theoretical point in Ariely et al. (2003) that preferences can exhibit substantial arbitrariness. Additionally, the meta-analysis confirmed a significant anchoring effect on WTP, and the *p*-curve analysis and the funnel plot indicated that publication bias is unlikely to account for the observed effect sizes.

Conclusion and General Discussion

In this paper, we examined the robustness of anchoring effects in preferential judgments. While the anchoring effect has been shown to be robust in other types of numeric judgment tasks (e.g., general knowledge), and a recent Many Labs project showed stronger effects than those of the original study (Klein et al., 2015), the anchoring effect in valuation tasks has been shown to be inconsistent (Bergman et al., 2010; Fudenberg et al., 2012; Maniadis et al., 2014). Valuation tasks are distinctive in that (1) tasks are incentive compatible (Camerer et al., 2016), and (2) participants' existing preferences can play an important role (Fudenberg et al., 2012). Thus, the finding that task irrelevant random anchors influence people's WTP has been thought to be strong evidence showing preference construction (Lichtenstein & Slovic, 2006; Slovic, 1995).

Despite its theoretical importance and distinctive characteristics, the replicability of anchoring on preferential judgments has not been thoroughly and systematically evaluated. We conducted a thorough examination of the robustness of anchoring effects on preferences as measured by WTP. Our results more closely resembled the original results in Ariely et al. (2003) than the null results in Fudenberg et al. (2012), although the effects we observed appeared to be lower in magnitude than those of the original study. Our cross-study analysis showed that Ariely et al. (2003)'s study produced a significantly

larger effect than our study, whereas Fudenberg et al. (2012)'s study produced a significantly smaller one; our study and Bergman et al. (2010)'s study had similar effect sizes. These differences are larger than would be expected from sampling variation across identical experiments, implying that other factors differed across these studies.

We showed that the interactions between the anchoring number and several procedural factors are unlikely to explain the differences between prior results. However, the interaction between the anchoring number and recruiting method was marginally significant: recruited participants showed slightly stronger anchoring effects than class activity participants. This relationship remains marginally significant after screening out participants with inconsistent responses, implying that the effect was not driven by inconsistent responses (Appendix B). One possibility is that participants recruited specifically for the research study and informed that they are participating in an experiment were more likely to consider the anchoring number as informative for the valuation task. Frederick et al. (2013) did not investigate the effect of recruiting method or experimental environment, but they did argue that participants' inferences about the relevance of random number ranges provided in lab experiments are well-founded. Further investigation could help in understanding how environmental factors influence the perceived randomness of anchoring numbers, and its subsequent effect on WTP.

In addition to testing procedural differences, we also examined whether other mechanisms could explain preference anchoring in Study 3. We found that varying the anchoring number within participant, reducing the repetitiveness of the anchor did not reduce the strength of the anchoring effect. Additionally, we tested whether self-generated anchors (e.g., the amount participants previously spent on the items, or the

prices they saw in stores) dilute the effect of experimenter-given anchor on WTP. We found that anchoring can influence WTP even when people have considered possible prices before the anchor is given. Based on all three experiments and the cross-study analysis, we conclude that anchoring effects on preferential judgments are replicable and robust to a variety of procedural variations.

Although our studies did not directly test the underlying mechanisms of anchoring on preferential judgments, the results add to the ongoing theoretical discussion about anchoring efforts. In particular, the moderating effect of the percentile rank gap between anchors favors theories of anchoring where the distribution of participants' valuations must overlap with the range of anchor values used. The classic anchoring and adjustment mechanism, where participants start by considering the anchor and adjust until they find an acceptable response (Tversky & Kahneman, 1974) has this feature: an anchor well above the plausible range would get adjusted all the way down to the plausible range just as a lower anchor would. The selective accessibility account of anchoring also has this feature. Under selective accessibility, people generate anchor consistent information when evaluating the anchor as a potential response (Chapman & Johnson, 1999). Participants might generate anchor-consistent information and utilize the information when the anchors vary within a plausible range, whereas they ignore the anchors when the anchors are too extreme. Thus, the percentile rank gap of anchors could moderate anchoring effects on preferential judgments.

In contrast, other theories do not rely as much on overlap in the distributions of valuations and anchors. The basic anchoring effect where attention to the anchoring number can be responsible for the effect (Wilson et al., 1996) has already been shown to

be fragile (Brewer & Chapman, 2002), and is inconsistent with our results, especially Study 3. A more recent theory, the scale distortion theory of anchoring, posits that anchors affect the mapping of judgments to the response scale without necessarily causing representational changes of the target items (Frederick & Mochon, 2012). Thus, anchoring effects would not rely on the overlap of anchors and valuations as much as the range of the anchors themselves, which was identical (0 through 99) in the studies we analyzed. While scale distortion has been shown to play an important role in general knowledge anchoring, and could still have some role in valuation anchoring, our findings suggest that scale distortion is not the primary mechanism for valuation anchoring.

CHAPTER 3.

ESSAY 1B: PERSISTENCE OF ANCHORING EFFECTS ON PRODUCT VALUATIONS

Classic economic theories have assumed that human preference is precise, stable, and well-defined (March, 1978), while psychologists has shown that human preferences are constructed depending on given contexts and the decision makers' status at the moment they make judgments and decisions (Rakow & Newell, 2010; H. A. Simon, 1955; Slovic, 1995).

Despite of enormous evidence supporting the preference construction, its long-term effect has less been investigated, and the results are inconclusive. For example, Schonberg et al. (2014) found that participants showed increased preference for items with which irrelevant cue and motor response were associated, and the increased preference on the items retained longer than a month. Sharot, Fleming, Yu, Koster, and Dolan (2012) also showed that choice-induced preference persisted even after 2.5 to 3 years from the initial session. Izuma and Adolphs (2013) also showed that experiencing peer group's decisions influences participants preference even after four months. On the other hand, several studies on long-term effect of cognitive dissonance and social influence have shown that constructed preference does not persist at all or persist for a short term. Huang, Kendrick, and Yu (2014) examined how long people retain other's opinion, and found that participants did not retain other's opinion longer than three days. Lee and Schwarz (2010) showed that choice-induced preference can be easily washed away simply by physically washing their hands using soap. D. Simon, Krawczyk,

Bleicher, and Holyoak (2008) also showed that choice-induced preference disappears after one week.

In this article, we examined the persistence of constructed preference using a valuation anchoring task in which the source of preference construction is explicitly random and participants' judgments are consequential. Anchoring effect refers to a phenomenon that people's numeric estimates are assimilated to a previously considered standard, even when the standard is uninformative to given task (Chapman & Johnson, 1999; Tversky & Kahneman, 1974). The anchoring effect has been shown to be robust in diverse areas, such as estimating the answers of general knowledge questions, physical length and time, mathematics questions (for a recent review, see Furnham & Boo, 2011), as well as human preference (Ariely et al., 2003; Bergman et al., 2010; Chapman & Johnson, 1999; Green, Jacowitz, Kahneman, & McFadden, 1998; Johnson & Schkade, 1989; Northcraft & Neale, 1987). For example, Ariely et al. (2003) asked participants to state willingness-to-pay (WTP) for several market goods. Before answering the question, participants were asked answer whether or not they would purchase the shown items for a dollar amount which was formed by transforming their last two digits of social security number into a dollar amount. Even though the price was explicitly random and irrelevant to given task and the task was incentive compatible, participants WTP was systematically influenced by the anchoring number.

This valuation anchoring has been regarded as evidence showing preference construction (but also, see Fudenberg et al., 2012; Maniadis et al., 2014), little has been known about its long-term effect. Literature in anchoring effect on general knowledge has shown that anchoring effect can persist over time by processing knowledge that

supports anchor as a possible answer (Blankenship, Wegener, Petty, Detweiler-Bedell, & Macy, 2008; Mussweiler, 2001a). More recently, however, Frederick and Mochon (2012) showed that anchoring effect is not because of representational change rather it is driven by scale distortion, so anchoring effect might be temporal. Valuation anchoring task is, however, distinctive from other types of numeric judgment anchoring tasks in terms that participant's previous experience and preference play an important role (Fudenberg et al., 2012), and that the task is incentive compatible in which stating their true preference is the best strategy (Camerer et al., 2016). Thus, it is unclear whether constructed preference from random anchor retains long time or the effect is transient because of the nature of the task that stating true preference is the best strategy and being influenced by random anchor penalizes their utility.

In Study 1, we investigated the long-term effect of random anchor on preference in two ways. We investigated the long-term effect of random anchor by introducing a one-week time gap between comparison questions (anchor) and WTP (anchor persistence) and by repeatedly measuring WTP (preference persistence). We also measured WTP again after two to five months from the first session to examine whether or not the constructed preference from anchoring effect persists even with longer time gap. In Study 2, we investigated the preference persistence by exogenously manipulating time gap to exclude possible selection bias in Study 1, and the effect of repeatedly stating WTP on long-term anchoring effect on valuation.

Study 1. The Effect of Immediate WTP Response Elicitation

Participants

The experiment consisted of two sessions, and there was a one-week gap between the two sessions. We aimed to recruit 150 participants. Since the study was conducted with a group setting (each session had 2-10 participants), we predefined the stopping rules as follows: (1) as soon as the total number exceeds 150, or (2) the total number of recruitment is greater than 140 and the academic semester is over. A total of 145 undergraduate students were recruited in exchange for class credit, and among them, 141 participants (mean age = 20.94 years, *SD* = 3.88; male = 54, female = 71, 16 participants did not provide demographic information) completed both sessions. The study was approved by the Institutional Review Board.

Experimental Procedure

Overall procedures were similar with Ariely et al. (2003). Participants were asked to answer two questions for each of six market goods (cordless mouse, Bluetooth keyboard, normal and fancy dinner, art book, and humidifier; average price = $48.33, price range: $30-$100) across two sessions (Session 1 vs. Session 2). For each item, exploring the item, participants were asked to answer whether or not they would buy the item for a random (anchor) price (willingness-to-buy, WTB; binary decision: Yes or No). The random price was decided by asking participants to select one card from a deck of cards which ranged from 0 to 99. Therefore, the anchoring number was transparently random to the participants. After answering WTB question, the participants were asked to state the most they would be willing to pay (WTP).

The long-term effect of random anchor on WTP can be examined in two ways: (1) whether or not the effect of random anchor persists when WTP is asked one week later

from WTB question (anchor persistence), and (2) whether or not the WTP immediately asked after WTB is consistent with WTP after one week when we measure it again (preference persistence). To measure anchor persistence, participants were asked to answer WTB question for randomly selected three items, and they were asked to answer both WTB and WTP for the other three items to measure preference persistence in Session 1. At the end of Session 1, participants were asked to provide demographic information, such as age and gender. In Session 2, participants were asked to state WTP for all of the six items, and after completing the task, participants were asked to provide their best guess for the random price they picked in Session 1.

For both sessions, we used a Becker-DeGroot-Marshack (BDM) incentive compatible method (Becker et al., 1964) to make participants' decisions are consequential. The incentive compatible procedures followed as below: (1) each participant had a 5% chance to be a winner, (2) which session's answers will be used were randomly decided, (3) which item to be sold was randomly selected, and (4) which answer (WTB (only when it is decided to use the answers from Session 1) or WTP) will be used were randomly selected by rolling dice or tossing a coin. The winners were endowed $80. Participants were informed that they will be endowed money if they are selected as a winner, but we did not specify the exact amount to exclude possible reference price effect. If the purchase was based on WTB question and their answer was Yes, participants purchased the selected item for the random (anchor) price, while there was no transaction if their answer was No. If the purchased was based on WTP question, we randomly select a price from a 10 by 10 price matrix which ranged from $0 to $80 by rolling two 10-face dice, and compared the price with participant's WTP for the item

randomly selected to be sold. If their WTP was greater than the random price from the price matrix, they purchased the selected item for the random price, while there was no transaction, if there WTP was greater than the random price.

Result

Anchor persistence

For the data analysis, we regressed WTP responses on anchoring number, interacted with time gap. To account for item effect, we ran the analysis with item fixed effect, and the analysis was conducted using STATA14. We excluded responses whose WTP is greater than $150 which was 50% higher than the market price of the most expensive item, and a total of seven responses were excluded.

To examine anchor persistence, we compared the effect of random anchor on WTP responses for preference persistence in Session 1 and on WTP responses for anchor persistence in Session 2 (Table 9, Model 1). We found a significant main effect of anchor ($t = 6.84$, $p < .001$) and a significant interaction between anchor and time ($t = -2.62$, $p = .009$), but the main effect of week was not significant ($t = 0.58$, $p > .250$). The results imply that as anchoring number increased participants' WTP also increased, but the effect was significantly weaker in Session 2 than in Session 1. A separate regression analysis result for Session 2 responses showed that anchoring effect is significantly greater than zero ($b = 0.10$, $SE = 0.03$, $t = 3.22$, $p = .001$), showing that anchoring effect is significantly weaker when there was a one-week gap between WTB and WTP than when WTP directly followed WTB, but the random anchor still significantly influenced WTP.

Table 9. Regression Analysis Results for Anchor and Preference Persistence in Study 1

	Anchor Persistence		Preference Persistence	
	Model 1 Time	Model 2 Memory	Model 3 Time	Model 4 Memory
Anchor	0.16***	0.12**	0.14***	0.10*
	(0.02)	(0.04)	(0.04)	(0.04)
Time	0.69		1.01	
	(1.19)		(1.18)	
Anchor×Time	-0.06**		-0.07**	
	(0.02)		(0.02)	
Memory		1.77		2.76
		(2.15)		(2.16)
Anchor×Memory		-0.02		-0.04
		(0.04)		(0.04)
Intercept	Item FE	Item FE	Item FE	Item FE
N	842	424	836	418

*+ p < .10, * p < .05, ** p < .01, *** p < .001*
Note: Standard errors are in the parentheses. The categorical variables Time (-1: Session 1, 1: Session 2) and Memory (-1: not correctly recalled anchoring number, 1: correctly recalled anchoring number) were effect coded.

To rule out the possibility that the consistent effect was driven by memory effect, we investigated whether or not memory moderate the effect of anchor on WTP in Session 2. For the analysis, we regressed WTP of Session 2 on anchoring number with item fixed effect, interacted with memory recall (Table 9, Model 2). Among 141 participants, 98 participants correctly recalled the random number they received in Session 1, while 43 participants did not. The result showed that there was no interaction effect between anchor and memory (t = -0.41, p > .250), implying that memory did not moderate the long-term effect of anchoring on WTP when there was a one-week gap between WTB and WTP.

Preference Persistence

Next, we examined whether or not random anchor-induced preference persists over time. For the analysis, we regressed WTP responses for preference persistence on anchoring number with item fixed effect, interacted with time (Table 9, Model 3). We found a significant main effect of random anchor ($t = 6.34$, $p < .001$) and a significant interaction between anchor and time ($t = -3.26$, $p = .001$), but the main effect of memory was not significant ($t = 0.85$, $p > .250$). A separate regression analysis result for Session 2 showed a significant anchoring effect ($b = 0.07$, $SE = 0.03$, $t = 2.25$, $p = .025$). The results imply that the anchoring effect on WTP was significantly weaker in Session 2 than in Session 1 when WTP was repeatedly asked after one week, but the effect was still significantly greater than zero.

We also examined whether remembering the anchoring number influenced the anchor-induced preference. For the analysis, we regressed WTP responses of Session 2 on anchoring number with item fixed effect, interacted with memory recall (Table 9, Model 4). Similar to anchor persistence, we did not find a significant interaction effect between anchor and memory ($t = -1.10$, $p > .250$), showing that memory did not moderate the anchoring effect on WTP when WTP was measured after one week from when anchoring number was provided.

Follow-up Survey

To examine whether or not the random anchor-induced preference persists with a longer time gap, we distributed a follow-up survey after 2-5 months from the first session. Among the participants who completed the two laboratory sessions, a total of 55

participants (M_{age} = 20.78, SD = 2.46, female = 23) participated in the follow-up survey. Among them, four participants were selected as a winner at the end of the second session, while the others were not. The follow-up survey was online survey, so participants did not come to laboratory, but we also employed the same incentive compatible method with the laboratory sessions.

For the analysis, we regressed WTP on anchoring with item fixed effect at each session. Extreme response whose WTP was greater than $150, and one response was excluded. We found a decreasing trend of anchoring effect on WTP as the time gap increased, but anchoring effect was statistically significant at each session (Figure 4). In further analysis, we examined whether or not the anchoring effect was interacted with time by regressing WTP on anchor interacted with three contrasts for time variable (Session 1 vs. Session 2, Session 1 vs. Follow-up, Session 2 vs. Follow-up). Regression results showed that there was no significant interaction between anchor and time for the Session1 vs. Session 2 contrast (b = -0.04, SE = 0.03, t = -1.37, p = .170), even though there was a sharp drop between the two sessions. We found a significant interaction between anchoring and time for the Session 1 vs. Follow-up contrast (b = -0.07, SE = -0.03, t = -2.43, p = .015), but the interaction between anchor and time for the contrast between Session 2 and follow-up was not statistically significant (b = -0.02, SE = 0.02, t = -0.87, p > .250).

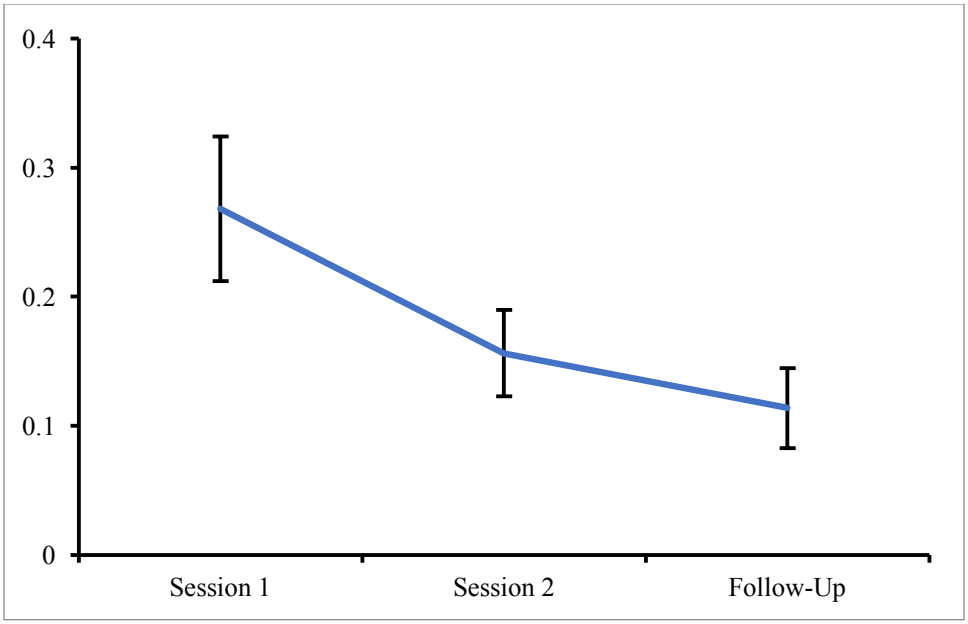

Figure 4. Coefficients of Regression Analysis at Each Session in Study 1

The results showed that anchoring effect decreases as time gap increased, but after sharp decrease in early stage, the pattern seems to be stable. The results might imply that random anchor-induced preference can persist even after longer time period (2-5 months from the initial session in this experiment). One possible limitation is a selection bias. For example, we found a significantly weaker anchoring effect in Session 2 than in Session 1 from the preference persistence analysis, but we did not find a significant difference in anchoring effect between Session 1 and Session 2 for those who completed the follow-up session. Thus, it might be possible that the long-term anchor-induced preference might be driven by participants whose WTP responses were more consistent across Session 1 and Session 2. To rule out this possibility, we replicated the findings by exogenously manipulating the time gap in Experiment 2. Indeed, we examined whether or not repeatedly answering WTP contributed to the long-term random anchor effect on WTP in Experiment 2.

Study 2. The Role of WTP Repeatition

Participants

The study consisted of three sessions. We aimed to recruit 200 participants. Since the study was a longitudinal design and was conducted with a group setting, we predefined the stopping rules as follows: (1) as soon as the total number of recruitment exceeds 200, or (2) the total number is greater than 190 and the academic semester is over. A total of 194 university undergraduate students participated in the experiment in exchange for class credit, and a total of 150 participants (M_{age} = 20.18 years, SD = 1.78, 73 women) completed all of the three sessions.

Experimental Procedure

A 2 (repeat: repeat vs. no-repeat; between-subject) × 3 (time: Session 1 vs. Session 2 vs. Session 3; within-subject) × 2 (delay of Session 3: 4-week vs. 8-week; between-subject) mixed design was used. Participants in same group were in the same condition, and the groups were randomly assigned one of the four cells (delay × repeat).

Overall procedures were similar to the design for preference persistence in Study 1, except that participants were asked to come to the laboratory three times (Session 1 vs. Session 2 vs. Session 3). In Session 1, all groups were asked to answer WTB and WTP questions for the same six market goods used in Study 1. After answering all of the questions, participants were asked to provide demographic information. In Session 2, *repeat* condition was asked to answer WTP questions for the six items, while *no-repeat* condition was not. Instead, *no-repeat* condition was asked to complete a task (e.g.,

attitude toward different donation messages) that any WTP responses were not involved. In Session 3, both *repeat* and *no-repeat* conditions were asked to answer WTP questions for the six items, but 4-week condition conducted Session 3 after 4 weeks from Session 1, while 8-week condition conducted 8 weeks after from Session 1.

We also employed the same incentive compatible method, and the rules were the same: (1) each participant had 5% chance to be a winner, (2) which sessions' answer will be used was randomly selected, (3) which item will be sold was randomly selected, and (4) which answer (WTB or WTP, if answer from Session 1 is selected) will be used was randomly selected. The reconciliation method was the same as Study 1.

Result

For the analysis, we regressed WTP on anchoring number at each session with item fixed effect. Ten extreme responses whose WTP was greater than $150 were excluded in the analysis. The regression analysis results showed that anchoring effect was significant at all the time sessions. The pattern was similar to Study 1, showing that there was a sharp decrease between Session 1 and Session 2, and later sessions showed a stable anchoring effect (Figure 5). In further analysis, we examined whether or not the anchoring effect was interacted with time by regressing WTP on anchoring number interacted with three contrasts for delay. We found a significant interaction between anchor and time for Session 1 vs. Session 2 contrast (b = -0.07, SE = 0.01, t = -4.74, p < .001), but there was no significant interaction between anchor and time for Session 2 vs. Week 4 contrast (b = -0.02, SE = 0.02, t = -0.90, p > .250) and for Session 2 vs. Week 8 contrast (b = -0.02, SE = 0.02, t = -0.80, p = > .250). These results show that anchoring

effect on WTP was significantly smaller after one week, but the effect was stable even after four and eight weeks.

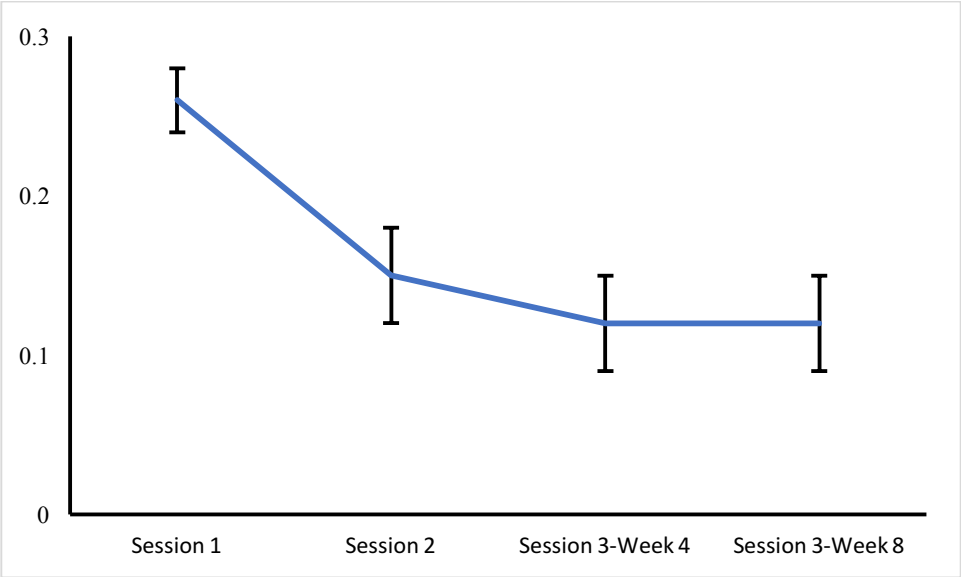

Figure 5. Coefficients of Regression Analysis at each Session in Study 2

Next, we examined whether or not repeating WTP moderated the long-term anchoring effect on WTP. For the analysis, we regressed WTP of the Session 3 on anchoring number, interacted with repeat (repeat vs. no-repeat) and delay (4-week vs. 8-week). The result showed that there was no significant anchor × repeat × delay three-way interaction (t = -0.25, p > .250) and anchor × repeat interaction (t = 1.28, p = .200), showing that repeating WTP did not influence persistent anchor-induced preference (Table 10). The results show that repeatedly answering WTP did not contribute to the long-term anchoring effect on WTP.

Table 10. Regression Results for Memory and Repeat in Experiment 2

Memory	*b* (*SE*)	Repeat	*b* (*SE*)
Anchor	0.11***	Anchor	0.11***
	(0.02)		(0.02)
Delay	-0.74	Delay	-1.21
	(1.33)		(1.21)
Anchor×Delay	-0.01	Anchor×Delay	0.01
	(0.03)		(0.02)
Memory	1.81	Repeat	0.74
	(1.33)		(1.21)
Anchor×Memory	-0.04	Anchor×Repeat	0.03
	(0.03)		(0.02)
Memory×Delay	-0.19	Repeat×Delay	-0.93
	(1.33)		(1.21)
Anchor×Memory×Delay	-0.01	Anchor×Repeat×Delay	-0.01
	(0.03)		(0.02)
Intercept	Item FE	Intercept	Item FE
N	895	*N*	895

*+ p < .10, * p < .05, ** p < .01, *** p < .001*
Note: Standard errors are in the parentheses. The categorical variables Gap (-1: Week 4, 1: Week 8) and Repeat (-1: no repeat, 1: repeat) were effect coded.

Similar to Experiment 1, we also examined the effect of memory on the long-term anchoring effect on WTP. A total 41 participants correctly remembered the anchoring number, while 109 participants did not. For the analysis, we regressed WTP on anchoring number, interacted with gap and memory (Table 10). We did not find a significant anchor × memory × delay three-way interaction (t = -0.32, p > .250) and a significant anchor × memory interaction (t = -1.22, p = .222). The results imply that correctly remembering the random anchor they received did not influence the long-term anchoring effect on WTP.

General Discussion

In this study, we examined persistence of constructed preference by showing long-term effect of random anchor on WTP. In Study 1, we examined the long-term effect of anchoring on WTP in two ways: (1) whether or not anchoring effect is significant when there is a one-week gap between WTB and WTP (anchor persistence), and (2) whether or not constructed preference from irrelevant anchoring number retains over time (preference persistence). The results showed that anchoring effect retained after one week, and that anchor-induced preference persisted after one week. In a follow-up survey distributed after 2-5 months from the first session, participants' WTP was still significantly associated with anchoring number, and the effect did not differ from that of Session 2. In Study 2, we tested the long-term effect of random anchor on WTP by exogenously manipulating the time gap (4-week or 8-week) to excluded possible selection bias in Study 1. We found that a significant anchoring effect on WTP in both 4-week and 8-week delay conditions, and there was no difference in anchoring effect between the two conditions. We also examined whether or not repeating WTP responses and remembering anchoring number moderated the long-term anchoring effect on WTP in Study 2, and found that repeating WTP response and memory did not moderate the long-term anchoring effect on WTP.

The findings have implications for studies of persistence of constructed preference and anchoring effect on valuations. Numerous judgment and decision-making studies have shown that preference is malleable and constructed by given contexts and decision maker's cognitive traits at the time they make judgments and decisions, but relatively little is known about the long-term effect of constructed preference, and the

results are inconsistent. Indeed, most of persistence of constructed preference was studied using choice-induced preference (e.g., Sharot et al., 2012; D. Simon et al., 2008) or social conformity (Huang et al., 2014; Izuma & Adolphs, 2013) where source of preference construction can be (in)directly informative, but relatively less has been known whether task irrelevant information can persist over time. This study is a first attempt to examine persistence of constructed preference using a valuation anchoring task where the anchoring number was explicitly random and participants' decisions were consequential. Our findings showed that anchor-induced preference can retain after two months, even though being susceptible to anchoring number harms their utility.

The findings also have implication for the studies of anchoring effect on valuation. Anchoring effect on factual numeric judgment tasks has been shown to be robust, and recent a Many Labs project also showed that its effect is stronger than the original study (Klein et al., 2015). Recent replication attempts of anchoring effect on valuation, however, showed inconsistent results. For example, Bergman et al. (2010) showed that anchoring effect significantly influenced WTP, while Fudenberg et al. (2012) did not find a significant anchoring effect on WTP and willingness-to-accept (WTA). Maniadis et al. (2014) also showed that anchoring effect did not influence WTA using an unpleasant hedonic experience (e.g., white noise). In this study, we found strong anchoring effect in two experiments using bigger sample sizes than previous studies. The results might imply that anchoring effect on valuation (at least for WTP) is not false positive.

There are also several limitations in this study. First, even though we tried to exclude possible selection bias in Study 2, there was a still high attrition rate (22.68%). Future research might be needed to test the findings with lower attrition rate to fully

eliminate selection bias. Second, due to the issue of Institutional Review Board approval, we endowed money for the incentive compatible method. Even though participants did not know the exact amount the winners will be endowed, but it is true that they did not use their own money for the task. Thus, this might be less consequential than using their own money. Future research might be needed to test the long-term anchoring effect on WTP with a fully incentive compatible method, if available. Third, there is a possibility that the long-term effect might be driven by consistency motivation. Even though we designed the study so that each session is incentive compatible, participants could have motivation to make their decisions consistent. We did not systematically measure how many participants mentioned, but in conversation with participants at the end of the last session, most of the participants mentioned that they did not think their previous answers when they answered WTP in later sessions, since in each session they were informed that their answer is consequential and which week's answer will be used are randomly selected. However, since there is still a possibility that their previous answer could influence their answers in later sessions, future research might be needed to test the role of consistency motivation in the long-term anchoring effect on WTP by applying longer time gap between WTB and WTP questions or by measuring whether participants considered previous WTP answers when they make WTP decisions in later sessions.

CHAPTER 4.

ESSAY 2: THE DESCRIPTION-EXPERIENCE GAP IN COMPLEX GAMBLES: EVIDENCE FOR OVERALL PROBABILITY OF WINNING HEURISTIC

The description-experience (D-E) gap refers to the inconsistent preferences observed in risky choice tasks depending on the format in which the information is presented (Camilleri & Newell, 2013; Hertwig, Barron, Weber, & Erev, 2004a; Rakow & Newell, 2010). In a seminal study, Hertwig et al. (2004a) asked two groups (decision from description versus experience) of participants to choose between a risky gain (e.g., 80% chance of gaining 8₵) or a sure gain (e.g., 6₵ for sure). When the information was described completely (i.e., magnitude and probabilities of both options were known; decision from description), only 36% of the participants selected the risky option even though it was associated with higher expected value (6.4₵ versus 6₵). When the information had to be learned through repeated sampling of each alternative (decision from experience), 88% of the participants chose the risky option instead. The difference, known as the D-E gap, has since been directly replicated using the same stimuli as the original study, though the effects (size of the gap) have been smaller (Hau, Pleskac, Kiefer, & Hertwig, 2008; Rakow, Demes, & Newell, 2008). Similarly, the D-E gap has also been shown across several other studies using different stimulus sets (for a recent review, see Camilleri & Newell, 2013; Rakow & Newell, 2010). Understanding the D-E gap is particularly important given that several real-world decisions (e.g., a doctor deciding to prescribe a certain medication to a patient while considering its potential side effects) are often better characterized by learning from experience rather than description

(Gonzalez & Dutt, 2011; Hau et al., 2008; Li, Rakow, & Newell, 2009; March & Shapira, 1987; Weber, 2006; Yechiam, Barron, & Erev, 2005).

The D-E gap has historically been attributed to the inconsistent weighting of a rare event (Hertwig & Erev, 2009). People often overweight the probability of a rare event in description condition, while they underweight the probability of the rare event in experience condition. The D-E gap has been mostly reported in situations where there is a rare event, and the magnitude of the gap often decreases as the rare event becomes less extreme (e.g., Erev et al., 2010; Rakow & B Rahim, 2010). However, recent studies have also demonstrated the D-E gap in equal probability gambling tasks in the absence of rare events (Ludvig, Madan, & Spetch, 2014; Ludvig & Spetch, 2011). For example, Ludvig and Spetch (2011) showed that participants were more risk seeking for gains in description condition than in experience condition, while they were more risk averse for losses in experience condition than in description condition. Indeed, Ludvig et al. (2014) showed that the D-E gap in equal probability gamble tasks was more prominent when the gains and losses were extreme.

Another characteristic of the D-E gap literature is that a vast majority of past studies investigating the D-E gap have used binary choice tasks in which participants choose between two alternatives, with each alternative having one or two outcomes. Very little is known about how these findings extend to more complex tasks, such as risky choice involving mixed gambles with multiple outcomes and multiple alternatives (Rakow & Newell, 2010). In one study, Hills, Noguchi, and Gibbert (2013) investigated the experience-based decision making using multiple alternatives (e.g., 2, 4, 8, 16, and 32 alternatives) with different presentation orders (few-to-many vs. many-to-few). They

showed that participants sampled more as the set size increased, but the number of samples per gamble however decreased as the set size increased. They also found a strong effect of the order of set size change, such that participants selected higher rank expected value gambles in smaller set size trials (set size 2 and 4) when the presentation order was few-to-many but not when the presentation order was many-to-few. Noguchi and Hills (2015) also found inconsistent preference for risky alternatives in an experience-based decision task depending on set size. In their study, participants preferred risky alternatives more when a set size was large (32 alternatives in the choice set) than when the set size was small (binary) for gain domain. Though these papers investigated experience-based decision making using multiple alternatives, each alternative still had a binary outcome (one gain and zero) and the decisions were not directly compared with corresponding description-based choices.

In the present study, we investigated whether people employ consistent decision strategies across description-based and experience-based decisions involving complex decision environments. There are several advantages to using non-binary mixed gambles similar to the ones used in this study (Rakow & Newell, 2010). First, they are more representative of most real-world decisions, which involve choices among different alternatives with multiple outcomes (Lopes, 1995; Lopes & Oden, 1999). Second, multiple-outcome gambles lead to the use of more diverse decision strategies across trials and individuals, providing valuable insights into the use of simplifying strategies and heuristics in decisions from experience – a key focus of this study.

It is well known that people simplify complex tasks using simplifying heuristics due to limited cognitive resources (Erev, Ert, & Yechiam, 2008; Fox & Hadar, 2006; H.

A. Simon, 1955; Tversky, 1972b; Tversky & Kahneman, 1974). In fact, it has been argued that people choose from a toolbox of heuristics based on *task environment* and *decision context* (Payne et al., 1988). However, most of this research has utilized description-based paradigms, and little is known about how the same heuristics extend to decisions from experience. Crucially, it remains unknown whether the D-E gap is restricted to differences in choice preferences between the two formats, or also extends to decision strategies. In one study, Hau et al. (2008) predicted preferences from an experience-based paradigm using a diverse array of choice models and heuristics. The results showed that the maximax heuristic (i.e., selecting an option that gives highest experienced maximum outcome) predicted choices in the task best, followed by others like the natural mean heuristic (i.e., choosing alternatives with higher sampled mean) and the lexicographic heuristic (i.e., identifying the most frequent outcome and selecting the alternative with the highest value for that outcome). In another study, Camilleri and Newell (2011) asked participants to report decision strategies that they used during the experience-based tasks. Participants' commonly reported strategy was similar to the natural mean heuristic, in addition to other strategies such as risk aversion, priority heuristic, and multiple composite strategies. Finally, Hills and Hertwig (2010) also showed that people used different decision strategies depending on the way they sample. Participants whose sampling patterns were frequently changing (i.e., zigzag) used round-wise decision strategy (i.e., counting how many rounds each option won and selecting an option which won more times), while people whose sampling patterns rarely changed showed summary-based decision strategy (i.e., comparing final mean values and selecting the alternative with higher final mean value). However, these studies do not

compare the consistency of strategies across the two formats. Here, we used three-outcome mixed gambles in an experience-based paradigm to directly compare the use of simplifying heuristics between decisions from description and experience.

Overall Probability of Winning (Pwin) Heuristic

In two independent studies, we used a modified version of the value allocation task where participants choose among three three-outcome gambles in each trial (Venkatraman, Payne, & Huettel, 2014). One heuristic that is often employed with complex mixed gambles is the overall probability of winning (Pwin) heuristic (Payne, 2005). The Pwin heuristic refers to decisions that are based on the valence of the gamble, while often ignoring the magnitude of the individual outcomes. For example, in Payne (2005), participants were asked to choose between a gamble that provided the highest positive outcome and a gamble that maximized the overall probability of winning. The results showed that participants preferred the option that maximized the overall probability of winning, even when the choice was associated with a lower expected value. The Pwin heuristic has since been reliably replicated in a number of studies involving three- and five-outcome gambles (Venkatraman, Payne, Bettman, Luce, & Huettel, 2009). It has also been shown to be associated with distinct processing strategies (Venkatraman et al., 2014) and brain regions (Venkatraman et al., 2009).

Here, we sought to investigate whether the Pwin heuristic extends to decisions from experience (DFE). Several experience-based decision making studies in the past have demonstrated a variant of the Pwin heuristic with binary gambles. For example, in one of the classic problems from Hertwig and colleagues (S: 3,1 vs. R: 32,0.1), participants show an increased preference for Gamble S (which has a higher probability

of winning) in DFE. However, in a different problem within the same study (S: 3,1 vs. R: 4,0.8), participants show an increased preference for Gamble R in DFE, which runs counter to the Pwin heuristic. One problem in these two examples is that the Pwin heuristic is confounded with the maximax heuristic and the minimax heuristic (choosing an alternative that gives the greatest minimum outcome). For example, in the first example, Gamble S gives a higher probability of winning but also has greater minimum outcome than Gamble R (minimum gain of S: 3 vs. R: 0). Thus, it is difficult to attribute whether the higher preference for Gamble S was contributed by the greater probability of winning or by the larger minimum outcome. In the latter example, Gamble S gives a higher probability of winning but also a lower greatest gain (maximum gain of S: 3 vs. R: 4). Therefore, it is not clear whether the lower preference for Gamble S is evidence against Pwin heuristic or evidence for maximax. The use of three-outcome mixed gambles in the current study enables us to isolate the Pwin heuristic from other strategies like maximax or minimax heuristics. Even in the absence of rare outcomes, Erev et al. (2008) showed that for the higher nominal magnitude condition (i.e., the magnitude of outcomes was large), participants preferred an alternative with greater number of positive outcomes over an alternative which provided a greater gain but included zero outcome in a repeated choice task where participants knew the possible outcomes and associated probabilities of alternatives. In this study, we seek to understand the role of the Pwin heuristic in a classic sampling paradigm using a three-outcome mixed gambles task.

Without any rare events for the mixed gambles used in this study, one possible prediction is that people will be more sensitive to extreme outcomes when they are learning the likelihood of an event from experience. For example, taste aversion research

has shown that animals can learn associations between extreme outcomes and the causal events, even with sparse exposures and longer gaps between the causal event and its effects (Garcia, Kimeldorf, & Koelling, 1955). The peak-and-end rule also implies that peak intensity influences people's judgments on past experienced events (Fredrickson, 2000). A similar focus on extreme outcome has also been shown to play an important role in experience-based paradigms with simple binary gambles (Ludvig et al., 2014). In that study, the authors found that the magnitude of outcomes in binary choice tasks contributed to distinct preferences in decisions from experience. Specifically, participants were more risk seeking when they experienced extreme gains or losses than when they experienced less extreme outcomes. Similarly, Payne, Samper, Bettman, and Luce (2008) showed that people considered the magnitude of biggest gain when they had to learn the attractiveness of gambles by sampling. Using a sequential paradigm similar to decisions from experience to study the boundary conditions of the unconscious thought effect, the authors found that the greatest proportion of participants simply selected the option that provides the greatest gain. Therefore, we seek to understand whether decisions from experience will lead to increased preference for choices that emphasize extreme outcomes or the effect of the overall probability of winning still significantly influences choices during experience-based decision making.

Alternatively, choices during decisions from experience may also be explained by the natural mean heuristic, with participants being more sensitive to the experienced expected value of the individual alternatives (Camilleri & Newell, 2011; Hau et al., 2008). For example, Hertwig and Pleskac (2008) argued that individuals rely on small samples in experience-based decision making since it makes decisions easier by amplifying the

difference between alternatives. The natural mean heuristic relies on such strength of evidence, leading to the choice of a gamble with the larger sample mean. It accounted for 77% of preferences in the choice sets used in Hertwig et al. (2004a), further demonstrating an important role of this strategy in experience-based decision making. However, it remains unclear if the use of natural mean heuristics extends beyond binary gambles and whether participants will continue to choose the option with higher sample mean even in more complex gambles. Therefore, in two independent studies using three-outcome mixed gambles, we examined the role of different decision strategies like natural mean, Pwin, minimax and maximax in explaining differences between decisions from description and experience, when information is different or matched across the two formats (Fox & Hadar, 2006; Hertwig & Pleskac, 2010).

Study 1.

Method

A total of 69 university undergraduate students (M_{age} = 22.28, SD = 3.61, female = 36) completed the study in exchange for class credit or $5 cash payment. Temple University's institutional review board approved the study and all participants provided informed consent.

Materials

We used a modified version of the value allocation task used in several past studies (Venkatraman et al., 2014). In each trial, participants had to choose between three different mixed gambles, presented in a 4 × 4 grid format (Figure 6). Each gamble

consisted of three outcomes (one gain, one loss and one intermediate), each with its own probability. The three alternatives were constructed by improving one of the three base outcomes respectively (Figure 6). Therefore, one alternative was always associated with the highest gain outcome (Gain maximizing or Gmax; see G1 in Figure 6), one alternative was associated with the lowest loss outcome (loss minimizing or Lmin; see G3 in Figure 6) and the third intermediate alternative (G2 in Figure 6) was associated with superior value for the intermediate outcome. Trials were further classified into two types: Pwin available or Pwin unavailable. In the Pwin available trials, the intermediate alternative was associated with a greater overall probability of winning compared to the other alternatives (G2 in Figure 6a). In the Pwin unavailable trials, there was no difference in overall probability of winning across all alternatives (G2 in Figure 6b). We used these later trials to rule out the possibility that the intermediate alternative is merely a compromise option (Venkatraman et al., 2009). In both trial types, all alternatives had equal expected value (Table 11).

Experimental Design and Procedure

We used a 2 (trial type: Pwin available vs. Pwin unavailable) × 2 (presentation format: description vs. experience) within-subject design (Camilleri & Newell, 2009; Ludvig et al., 2014). All participants solved the same problems in both presentation formats with a fixed block order. All trial types were randomized within each block, and the description format block was always presented second to prevent participants from familiarizing themselves with the structure of the problems (i.e. three outcome equal probabilities). Since the choice proportions in the description format were consistent with

previous studies using a similar paradigm (Venkatraman et al., 2014), we believe that order did not critically influence the core findings presented here.

		0.33 (P1)	0.33 (P2)	0.33 (P3)
Base		55 (O1)	-15 (O2)	-65 (O3)
G1	Gmax	75	-15	-65
G2	Pwin	55	5	-65
G3	Lmin	55	-15	-45

a. Pwin available trial

		0.33 (P1)	0.33 (P2)	0.33 (P3)
Base		65 (O1)	-30 (O2)	-85 (O3)
G1	Gmax	85	-30	-85
G2	Intermediate	65	-10	-85
G3	Lmin	65	-30	-65

b. Pwin unavailable trial

Figure 6. An example of Pwin available and Pwin unavailable trials. The three gambles (G1, G2, and G3) were developed by adding a constant (here 20) amount to the highest gain (O1 of G1, Gmax), intermediate (O2 of G2), and loss (O3 of G3, Lmin) outcomes of the base gamble. Adding a constant amount to the intermediate outcome changed the overall probability of winning in Pwin available trials (a), but not for Pwin unavailable trials (b). The row with base gamble is shown here for display purposes only and was not presented to participants. (P indicates probability and O indicates outcome)

Table 11. Stimuli and choice proportions in Study 1

	Base Gambles					Description Choice %			Experienced Choice %		
Trial	O1	O2	O3	Constant	EV	G1 Gmax	G2 Intermediate	G3 Lmin	G1 Gmax	G2 Intermediate	G3 Lmin
1	40	-10	-65	15	-6.67	30.88	48.53	20.59	20.59	51.47	27.94
2	60	-20	-60	25	10.00	32.35	47.06	20.59	26.47	48.53	25.00
3	55	-15	-65	20	-1.67	23.53	64.71	11.76	36.76	35.29	27.94
4	65	-10	-75	20	0.00	29.41	52.94	17.65	27.49	42.65	29.41
5	70	0	-90	20	0.00	30.30	51.52	18.18	40.91	48.48	10.61
6	45	-15	-65	15	-6.67	27.49	57.35	14.71	33.82	42.65	23.53
7	65	-30	-85	20	-10.00	47.06	25.00	27.94	45.59	23.53	30.88
8	50	10	-75	10	-1.67	29.41	35.29	35.29	26.47	35.29	38.24
9	80	-25	-70	15	0.00	27.54	30.43	42.03	44.93	15.94	39.13
10	60	5	-70	5	0.00	16.18	55.88	27.94	33.82	42.65	23.53

Trial 1 – 2: Practice trials, Trial 3 – 6: Pwin available trials, Trial 7 – 10: Pwin unavailable trials O1: highest gain, O2: intermediate, O3: loss outcomes. Constant is the amount added to each of the outcomes to form the three alternatives (G1, G2 and G3).

After completing two practice trials, participants completed 8 main trials (Table 11). In the description format, the gambles were presented in a 4 × 4 table consisting of three gambles with three outcomes and their (common) probabilities (Figure 6). Participants indicated their choice by pressing the appropriate button on the keyboard. There was no time constraint for responding. For the experience format, we used a sampling paradigm similar to past studies of decisions from experience (Camilleri & Newell, 2013; Hertwig et al., 2004). Three boxes (corresponding to the three gambles) were presented on the computer screen, and the three gambles were randomly allocated into those three boxes. Participants could sample one gamble at a time by pressing the corresponding button (1: left box, 2: middle box, 3: right box), and a randomly chosen outcome (based on the underlying probabilities) from that gamble was revealed within that box for 500ms. They could sample each gamble as many times as they wanted without any restriction in the sampling order. When participants felt that they had

sufficient information to make their decision, they were instructed to press the number 4. This toggled the screen to an output format, where they could indicate their choice by pressing the corresponding button for the chosen gamble (1, 2 or 3). If participants wanted to continue sampling, they were given the option to toggle back to the sampling screen again by pressing 4 again. At the end of the both sessions, participants were asked to complete a demographic questionnaire.

Results

Sampling: First, we investigated the number of samples and sampling pattern (across-gamble vs. within-gamble) depending on trial types (Pwin available vs. Pwin unavailable). A total of nine trials had zero samples, so we excluded these trials and their corresponding description trials from all subsequent analyses. Participants sampled on average 36.24 times (*Median* = 33.50, *SD* = 28.31), which is higher than the average number of samples reported in studies with binary gambles (a recent meta-analysis (Hertwig, 2015) showed that median number of samples in binary gambles was 16). In previous studies using a similar task in description condition, Pwin unavailable trials are associated with longer response times than Pwin available trials (Venkatraman et al., 2009). Consistent with these studies, we found that participants took longer for Pwin unavailable trials (*M* = 12.27, *SD* = 7.54) than Pwin available trials (*M = 10.40, SD = 6.66*) in the description condition ($t(68) = -3.05$, $p = .003$). However, there was no significant difference in the number of samples across trial types ($t(68) = -0.89$, $p > .250$, $d = 0.059$), though participants sampled slightly more in Pwin unavailable trials (*M* = 37.12, *SD* = 28.93) than Pwin available trials (*M* = 35.37, *SD* = 29.96).

Next, we examined the sampling pattern by looking at switching frequency - the ratio between the number of switches between the gambles and the maximum number of possible switches (Hills & Hertwig, 2010). Sampling pattern showed that participants sampled more within gamble (M = .34, SD = 0.36) in general. People switched across gambles more often in the Pwin unavailable trials (M = .34, SD = 0.37) than in the Pwin available trials (M = .33, SD = 0.37), but the difference in the search pattern between the two trial types was not statistically significant ($t(68) = -0.67$, $p > .250$, $d = 0.023$). These results suggest that decisions were more homogenous across trial types in the experience condition relative to the description condition.

Choice: We examined the effect of the two presentation formats (description vs. experience) on choice, by trial types (Pwin available vs. Pwin unavailable). First, we examined whether participants showed a consistent preference for the intermediate gamble in the Pwin available trials across both formats. We showed a strong preference for the intermediate Pwin alternative in the description condition (Table 12) and the proportions were consistent with prior studies using similar paradigm (Venkatraman et al., 2014). A binomial test for the choice proportion of the Pwin alternative was significantly greater than chance level (33%, p (two-tailed) < .001). Further, a cross tabulation analysis revealed that the association between presentation formats and choice was statistically significant in the Pwin available trials ($\chi^2(2, N = 540) = 11.68$, $p = .003$, Cramer's V = .15). In other words, we found that there was a decrease in the proportion of choices for the intermediate alternative in the experience condition relative to the description condition for Pwin available trials (Table 2, Pwin available columns).

Table 12. Choice proportions (%) by trial types in Study 1

Gamble	Choice Type	Pwin available		Pwin unavailable	
		Description	Experience	Description	Experience
G1	Gmax	27.54	34.90	39.83	37.92
G2	Intermediate*	57.00	41.91	36.35	28.99
G3	Lmin	15.46	23.19	33.82	33.09

**The intermediate choice represents the Pwin alternative for Pwin available trials.*

Next, we tested whether participants showed a strong preference for the intermediate gamble even in the Pwin unavailable trials, to rule out the possibility that participants' preference for the Pwin gamble (i.e., the intermediate gamble in the Pwin available condition) was merely a compromise strategy and not related to the Pwin heuristic. A binomial test indicated that the proportion of intermediate gamble choices in the description condition was at chance level (33%, p (two-tailed) > .250). An additional cross tabulation analysis revealed that there was no significant association between presentation formats and choice in the Pwin unavailable trials, ($\chi^2(2, N = 546) = 4.61$, $p = .100$, Cramer's $V = .09$). These results indicated that people did not prefer the intermediate gambles in these trials, and there was no difference in choice pattern across the two presentation formats, when the intermediate gamble was not associated with an increase in overall probability of winning (Table 12, Pwin unavailable column).

Indeed, further multilevel logistic regression analysis with random intercept for the Pwin available trials also showed a significant main effect of presentation format ($\chi^2(1) = 12.82$, $p < .001$). This indicates that participants had a strong preference for the intermediate alternative only when the intermediate alternative had a higher overall probability of winning, and this preference for the intermediate alternative was

significantly smaller in the experience condition than in the description condition. No such differences were observed between the formats for the Pwin unavailable trials ($\chi^2(1) = 3.37$, $p = .066$).

Although the choice-based analysis indicates a difference in preferences across the two formats with the three-outcome mixed gambles used in this study (similar to the traditional D-E gap observed in binary gambles), it provides little insights into the differences in the underlying decision strategies. For example, it is probable that participants' true experienced probabilities for the various gambles are significantly different from the predefined probability distributions, based on their sampling history and experienced outcomes. This could in turn lead to different probability profiles for the alternatives across the two formats. In other words, the intermediate alternatives (G2) in the Pwin available trials have two positive values, while both the extreme alternatives (G1 and G3) have only one positive value. However, if a participant experiences the positive outcomes more when sampling the extreme alternatives (especially in trials where participants sampled too little), these alternatives might be perceived as the Pwin option for those trials based on the experienced samples. Similarly, even though the expected values were set to be equal across the different alternatives, this might no longer be the case based on the experienced probabilities in the decisions from experience. As expected, we found that the experienced overall probability of winning and expected values were not identical with the original predefined values for the experience condition (Appendix C). Therefore, we conducted a follow-up study, where we modified the probabilities in description format to match those sampled by the participants in the experienced format. We predicted that if the differences in choice across formats were

purely related to sampling errors, then these differences should be eliminated when the probabilities in description condition are matched to those in experience, and preferences should be more consistent across formats.

Follow-up Study with Matched Problems Across Formats

A total of 66 additional participants[3] participated in a follow-up study. The procedures were largely similar to the main study. First, all participants completed 12 problems in the study, with 6 problems in Pwin-available condition and another 6 in Pwin-unavailable condition (Stimuli are in Appendix D). For each participant, half of the description trials were formed based on their sampling history in the experience format (matched condition), while the other half of the description trials were formed from predefined information (original condition, similar to main study). Therefore, there were 3 problems of each trial type in the matched and original conditions. Second, participants completed the experience format first, followed by description format, consistent with the main study. This also allowed us to easily match the probabilities in description to those experienced in the corresponding experience trial. If a particular outcome was never experienced in the experience condition, we used a negligible probability of 0.01 for those outcomes in the description condition in the matched trials to maintain consistency in presentation format across individuals (Appendix E). We were primarily interested in whether preferences across description and experience formats are different, when the information is matched across the formats. Trials in both formats were randomized.

A total of 25 responses that had zero samples were excluded from this analysis. First of all, we did not find any differences in the number of samples ($M_{\text{original}} = 26.94$,

[3] Age information is not available due to a computer recording error, but all students were undergraduates recruited through the subject pool at Fox School of Business, Temple University.

SD = 23.98 vs. $M_{matched}$ = 26.98, *SD* = 22.77; $t(65) = -0.03$, $p > .500$) and in the search pattern ($M_{original}$ = .34, SD = 0.28 vs. $M_{matched}$ =.31, *SD* = 0.27; $t(65) = 1.46$, $p = .148$), indicating that there were no behavioral differences between the original and the matched conditions during sampling. While we replicated the previously shown inconsistent preferences across formats in the original condition (Table 13; cross tabulation analysis: $\chi^2(2, N = 770) = 9.55$, $p = .008$; multilevel logistic regression: $\chi^2(1) = 5.48$, $p = .019$), we found relatively consistent preferences across formats in the matched condition (Table 3; cross tabulation analysis: $\chi^2(2, N = 764) = 4.53$, $p = .104$; multilevel logistic regression: $\chi^2(1) = 3.11$, $p = .078$). We also computed the proportion of trials where participants chose the same alternatives for the same trial across the two formats. We found significantly more consistent preferences across formats in the matched condition (*M* = .46) than in the original condition (*M* = .38; $t(65) = -2.08$, $p = .041$). In summary, matching the information across formats in a yoked experimental design diminished the differences in preference across formats (i.e., reduced the D-E gap). However, preferences were still inconsistent in more than 50% of the trials, suggesting that the gap is not completely eliminated even in the matched condition. Therefore, we next sought to explore whether these differences can be explained further by explicit differences in decision strategies across the two formats.

Table 13. Choice proportions (%) in the follow-up study across all problems. The gambles are not mapped to specific choice types since the probabilities are based on the experienced samples in the matched condition.

	Original		Matched	
	Description	Experience	Description	Experience
G1	36.36	35.32	30.89	37.70
G2	42.86	34.81	39.27	33.25
G3	20.78	29.87	29.84	29.06

Decision Strategies in Decisions from Experience

To further investigate the decision strategies underlying decisions from experience, we conducted two additional analyses, pooling the data across all trials in decisions from experience. We focused our analyses on the experienced outcomes and probabilities rather than the predefined ones. We also combined the data across both the main study as well as the follow-up study[4]. In the first analysis, we predicted participants' choice based on four common decision heuristic strategies (maximax, minimax, natural mean and Pwin), since these heuristics make differential predictions across the different alternatives in this study (experienced extreme gains and losses, expected value, and overall probability of winning). In the second analysis, we sought to further investigate the relative importance of each of these variables (extreme gain and loss values, expected values, and the overall probability of winning) in predicting choice across all trials using a conditional logistic regression analyses.

The prediction accuracies are summarized in Figure 7. Using a one-way ANOVA with each individual's choice prediction accuracies from the four decision strategies as a dependent variable, we found a significant main effect of decision strategy ($F(3, 134) = 13.63$, $p < .001$, $\eta_p^2 = .092$). Post-hoc analyses using Tukey HSD test showed that both the Pwin strategy (54% correct, $SD = 0.25$) and the natural mean strategy (55% correct, $SD = 0.22$) predicted choices better than the minimax strategy (42% correct, $SD = 0.21$) and the maximax strategy (48% correct, $SD = 0.27$). The comparison between the Pwin strategy and the natural mean strategy did not reveal any significant differences in

[4] We did not find a significant interaction between decision strategies and study (Study 1 vs. follow-up) for the prediction accuracies ($F(3, 399) = 0.17$, $p > .500$) and interactions between the four predictors and study code conditional in a multiple logistic regression analysis (z's < 1.00, p's > .450). In other words, the findings are largely consistent for each individual study.

prediction accuracy. In fact, we found a very strong correlation between experienced overall probability of winning and experienced expected value ($r(133) = .80$), and 42% of the choices were predicted accurately by both the Pwin and natural mean strategies. Additional one-sample *t*-test results showed that all of the four decision strategies predicted choices better than chance (chance = 33%; maximax: $t(134) = 6.54$, $p < .001$, minimax: $t(134) = 4.69$, $p < .001$, Pwin: $t(134) = 10.15$, $p < .001$), and natural mean: $t(134) = 11.61$, $p < .001$).

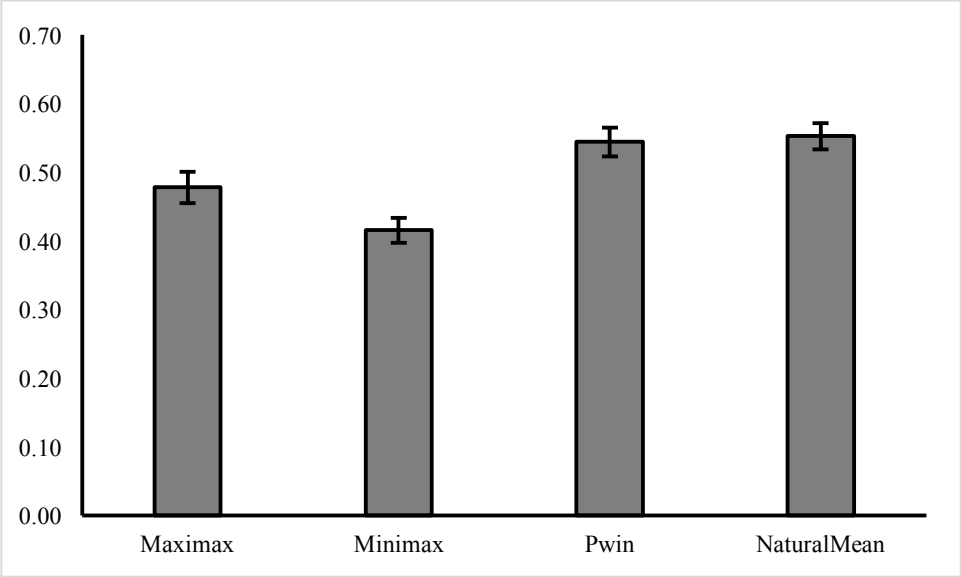

Figure 7. The prediction accuracies for the experience format from Maximax, Minimax, Pwin, and Natural Mean strategies in Study 1.

For the conditional logistic regression analysis using standardized values of the four predictors (Table 14), we found that all four predictors positively influenced choice when modeled separately, and experienced expected value (ExpEV) explained the greatest amount of variance of the dependent variable (Table 14, Models 1-4). Considering all four predictors together (Table 14, Model 5), we found that ExpEV, experienced overall probability of winning (ExpOP), and experienced maximum gain

(MaxGain) positively influenced choices. Further linear comparisons showed that the effect of ExpEV was stronger than that of ExpOP (ExpOP vs. ExpEV: b = -0.35, SE = 0.15, z = -2.30, p = .021) and MaxGain (MaxGain vs. ExpEV: b = -0.33, SE = 0.16, z = -2.07, p = .038), but there was no difference between MaxGain and ExpOP (MaxGain vs. ExpOP: b = 0.01, SE = 0.09, z = 0.15, p > .500).

Table 14. Conditional logistic regression results predicting choice across all trials in the experience format in Study 1. Models 1-4 indicate simple conditional logistic regression results using experienced maximum gain (MaxGain), minimum loss (MinLoss), overall probability of winning (ExpOP), and expected value (ExpEV) estimated from participants' sampling history as independent variables respectively. Model 5 indicates the multiple conditional logistic regression using all four variables simultaneously.

	(1)	(2)	(3)	(4)	(5)
MaxGain	0.723***				0.269***
	(0.060)				(0.084)
MinLoss		0.338***			-0.316***
		(0.074)			(0.075)
ExpOP			0.618***		0.255**
			(0.067)		(0.101)
ExpEV				0.722***	0.602***
				(0.075)	(0.099)
N	3,705	3,705	3,705	3,705	3,705
pseudo R2	.074	.026	.093	.115	.134

Standard errors in parentheses
** $p < .01$, *** $p < .001$*

Discussion

In summary, we found inconsistent preferences across description and experience formats based on the predefined outcomes and probabilities. Participants showed a higher preference for the intermediate alternative in the description format relative to the experience format, when it was associated with higher probability of winning, consistent with the D-E gap. However, when we accounted for sampling error by matching the

probabilities in the description condition to the experienced condition, we found that the magnitude of D-E gap reduced (Fox & Hadar, 2006; Hadar & Fox, 2009; Rakow et al., 2008), but was not completely eliminated. Therefore, we next sought to understand whether the differences across formats could be explained by differences in decision strategies. Although the Pwin heuristic significantly predicted choices, the natural mean heuristic also predicted choices as accurately as the Pwin heuristic did in the experience condition. Critically, the experienced expected value was a stronger predictor of choices than the experienced overall probability of winning in a multiple conditional logistic regression model that included these variables simultaneously.

Previous studies suggest that participants sacrifice expected value in favor of the Pwin heuristic in decisions from description (Payne, 2005; Venkatraman et al., 2009), suggesting that the two formats may favor the use of different strategies for the same problems. However, the core problems in the description condition in this study were constructed in such a way that there were no changes in expected values across the alternatives (probabilities were always equal), precluding a more direct dissociation of strategies across formats. Therefore, we sought to replicate our findings in an independent study with greater variability across the problems.

Study 2.

Method and Materials

A total of 47 undergraduate students (M_{age}: 21.75, *SD*: 2.98, Female = 31) participated in the study in exchange for class credit. Temple University's institutional review board approved the study and all participants provided informed consent. Overall

procedures were similar to Study 1. However, to provide additional variance in the problem types, we included problems where the expected value was not equal across the alternatives. Across a total of 14 problems (two practice trials and 12 main trials), we varied the expected value of the intermediate alternative in addition to the availability of Pwin alternative (see Table 15). Similar to Study 1, participants were presented these problems in two formats – experience and description. The description condition always followed experience, to prevent participants from familiarizing themselves with the structure of the problems, and only included the original problems that were not matched to the experienced samples. None of the participants from Study 1 participated in Study 2.

Results

Sampling: A total of five trials had zero samples, so we excluded these trials and their corresponding description trials from all subsequent analyses. Participants sampled on average 48.23 times (*Median* = 36, *SD* = 40.87), which is higher than the average number of samples reported in Study 1 and in studies with binary gambles. Consistent with Study 1, sampling pattern showed that participants sampled more within gamble (*M* = .20, *SD* = 0.27).

Choice: The choice proportions for each problem across decision formats are summarized in Table 5. As seen from the table, preferences varied across the different problem types. Participants showed a strong preference for the alternative that maximized the overall probability of winning when such an alternative was available in the mix (mean choice share: .51 for Trials 3, 4, and from 7 to 12). Similar to Study 1, they

switched to other strategies when there were no alternatives that maximized the overall probability of winning.

Table 15. Stimuli and choice proportions in Study 2

Trial		O1	O2	O3	EV	Choice Type	Description (%)	Experience (%)
1	prob.	0.33	0.33	0.33				
	G1	75	-15	-65	-1.67	Gmax	31.11	31.11
	G2	55	5	-65	-1.67	Pwin	57.78	51.11
	G3	55	-15	-45	-1.67	Lmin	11.11	17.78
2	prob.	0.33	0.33	0.33				
	G1	85	-10	-75	0.00	Gmax	8.51	23.40
	G2	65	10	-75	0.00	Pwin	70.21	51.06
	G3	65	-10	-55	0.00	Lmin	21.28	25.53
3	prob.	0.33	0.33	0.33				
	G1	90	0	-90	0.00	Gmax	19.15	25.53
	G2	70	20	-90	0.00	Pwin	57.45	44.68
	G3	70	0	-70	0.00	Lmin	23.40	29.79
4	prob.	0.33	0.33	0.33				
	G1	60	-15	-65	-6.67	Gmax	19.15	31.91
	G2	45	0	-65	-6.67	Pwin	61.70	38.30
	G3	45	-15	-50	-6.67	Lmin	19.15	29.79
5	prob.	0.33	0.33	0.33				
	G1	85	-30	-85	-10.00	Gmax	29.79	38.30
	G2	65	-10	-85	-10.00	-†	38.30	29.79
	G3	65	-30	-65	-10.00	Lmin	31.91	31.91
6	prob.	0.33	0.33	0.33				
	G1	60	10	-75	-1.67	Gmax	20.00	24.44
	G2	50	20	-75	-1.67	-†	40.00	31.11
	G3	50	10	-65	-1.67	Lmin	40.00	44.44
7	Prob.	0.25	0.25	0.50				
	G1	95	-10	-65	-11.25	Gmax	8.51	31.91
	G2	75	10	-65	-11.25	Pwin	61.70	53.19
	G3	85	-10	-55	-8.75	Lmin, EV	29.79	14.89
8	Prob.	0.33	0.33	0.33				
	G1	95	-20	-90	-4.95	Gmax	12.77	23.40
	G2	70	5	-90	-4.95	Pwin	65.96	51.06
	G3	85	-20	-75	-3.30	Lmin, EV	21.28	25.53
9	Prob.	0.33	0.33	0.33				
	G1	60	-5	-65	-3.30	Gmax	17.02	27.66
	G2	45	10	-65	-3.30	Pwin	42.55	40.43
	G3	55	-5	-50	0.00	Lmin, EV	40.43	31.91
10	Prob.	0.44	0.28	0.28				
	G1	80	-5	-50	19.80	Gmax, EV	21.28	40.43
	G2	75	10	-60	19.00	Pmax	68.09	36.17
	G3	75	-5	-45	19.00	Lmin	10.64	23.40
11	Prob.	0.33	0.33	0.33				
	G1	90	-15	-75	0.00	Gmax, EV	27.66	38.30
	G2	75	5	-90	-3.30	Pmax	61.70	42.55
	G3	75	-15	-70	-3.30	Lmin	10.64	19.15

Table 15. Stimuli and choice proportions in Study 2, continued

Trial		O1	O2	O3	EV	Choice Type	Description (%)	Experience (%)
12	Prob.	0.33	0.33	0.33				
	G1	55	-5	-55	-1.65	Gmax, EV	27.66	34.04
	G2	45	10	-65	-3.30	Pwin	53.19	36.17
	G3	45	-5	-50	-3.30	Lmin	19.15	29.79
13	Prob.	0.25	0.50	0.25				
	G1	85	5	-65	7.50	Gmax	21.28	36.17
	G2	70	15	-60	10.00	EV	48.94	44.68
	G3	65	5	-45	7.50	Lmin	29.79	19.15
14	Prob.	0.33	0.33	0.33				
	G1	60	-15	-70	-8.25	Gmax	10.64	25.53
	G2	55	-5	-65	-4.95	EV	68.09	38.30
	G3	50	-15	-60	-8.25	Lmin	21.28	36.17

Trial 1 and 2 were practice trials

† G2 in Trial 5 and 6 are the same as the intermediate alternative in the Pwin unavailable trials in Study 1

Similar to Study 1, we also found that choice preferences were different across the two formats. Though experienced EV was a strong predictor of choices in the experience condition in Study 1, we could not test the role of EV in description condition since all problems were matched for expected value. Here, we examined whether participants showed a higher preference for an alternative that maximized expected value in the description format, when an EV alternative was in the mix (Trials 7 to 13). The average preference for EV alternative was .30, which was not significantly different from chance (chance = 33%; $t(46) = -1.24$, $p = .222$). This suggests that maximizing expected value is not a dominant strategy in the description-based decision making, consistent with previous studies using similar paradigms (Payne, 2005; Venkatraman et al., 2009).

To replicate the effect of experienced overall probability of winning and expected value on choice in the experience condition, we conducted the same two additional analyses as Study 1. First, we predicted choices using the four decision strategies (maximax, minimax, natural mean, and Pwin strategies). The prediction accuracies are summarized in Figure 8. The results showed that the Pwin strategy and the natural mean

strategy predicted choices better than the maximax and the minimax strategies. Using a one-way ANOVA with each individual's choice prediction accuracies from the three decision strategies as a dependent variable, we found a significant main effect of decision strategy ($F(3, 46) = 20.82$, $p = .006$, $\eta_p^2 = .312$). Post-hoc analyses using Tukey HSD test showed that the Pwin strategy (54% correct, $SD = 0.20$) and the natural mean strategy (54% correct, $SD = 0.18$) predicted choices better than maximax (36% correct, $SD = 0.18$) and minimax strategies (35% correct, $SD = .16$). Additional one-sample t-test results showed that the predictions from the maximax ($t(46) = 1.01$, $p > .250$) and the minimax ($t(46) = 0.68$, $p > .500$) strategies were not significantly different from chance (33%), but the Pwin strategy ($t(46) = 7.33$, $p < .001$) and the natural mean strategy were ($t(46) = 8.04$, $p < .001$). However, the prediction accuracies between the Pwin strategy and the natural mean strategy were not different. Similar to Study 1, we still found a very strong correlation between experienced probability of winning and experienced expected value ($r(45) = .77$). Consequently, we found that 37% of the choices were accurately and commonly predicted by the both Pwin and natural mean strategies.

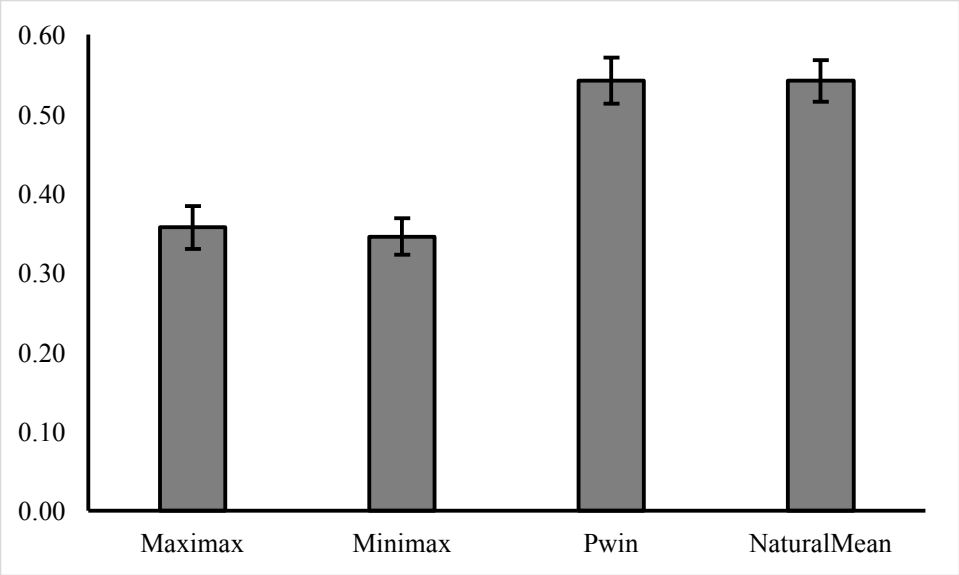

Figure 8. The prediction accuracies for the experience format from Maximax, Minimax, Pwin, and Natural Mean strategies in Study 2

Second, we ran a conditional logistic regression analysis to investigate the relative importance of extreme values, probability of winning, and expected value based on experienced values, using standardized values of the four predictors (Table 16). We found that all of the four predictors positively influenced choice (Table 16, Models 1-4). In terms of the variance explained by the four predictors, experienced expected value had the greatest explanatory power. Furthermore, when we considered all the four predictors together, the effect of experienced overall probability of winning and expected value were significant predictors of choice, but greater than maximum gain and minimum loss were not (Table 16, Model 5). Linear comparisons between predictors revealed that the effect of experienced overall probability of winning and expected value were stronger predictors than maximum gain (ExpOP vs. MaxGain: $b = 0.31$, $SE = 0.11$, $z = 2.67$, $p = .008$; ExpEV vs. MaxGain: $b = 0.73$, $SE = 0.21$, $z = 3.55$, $p < .001$) and minimum loss (ExpOP vs. MinLoss: $b = 0.40$, $SE = 0.11$, $z = 3.59$, $p < .001$; ExpEV vs. MinLoss: $b =$

0.82, $SE = 0.19$, $z = 4.28$, $p < .001$). Further linear comparison between the effect of experienced overall probability of winning and expected value showed that the effect of experienced expected value was greater than that of experienced overall probability of winning (ExpOP vs. ExpEV: $b = -0.42$, $SE = 0.19$, $z = -2.23$, $p = .026$).

Table 16. Conditional logistic regression results for the experience format in Study 2

	(1)	(2)	(3)	(4)	(5)
MaxGain	0.307***				0.020
	(0.077)				(0.117)
MinLoss		0.242**			-0.070
		(0.076)			(0.103)
ExpOP			0.643***		0.325**
			(0.065)		(0.092)
ExpEV				0.923***	0.748***
				(0.084)	(0.120)
N	1,647	1,647	1,647	1,647	1,647
Pseudo-R^2	.015	.009	.094	.137	.152

Standard errors in parentheses
** $p < .01$, *** $p < .001$

Conclusion and Discussion

Across two independent studies, we investigated differences in preferences and underlying decision strategies across two presentation formats (description vs. experience) using a complex risky choice task involving decisions between three three-outcome mixed gambles. We found inconsistent preferences across formats when only considering the predefined outcomes and probabilities. Participants showed a stronger preference for the intermediate alternatives in the description relative to the experience format, only when these gambles were associated with a greater overall probability of winning. When we matched the probability information in the description format to those experienced in the corresponding trials in the experienced format, we found that the differences between

formats diminished, but were not eliminated. Lastly, choices in the experience format were predicted significantly by both the natural mean and the Pwin heuristic strategies, when accounting for the experienced probabilities and outcomes. Therefore, participants still demonstrate the use of Pwin strategy even when making decisions from experience, which might explain some of the similarities in the matched conditions. Strikingly however, participants' decisions in the experience format were influenced by experienced expected value, and the conditional logistic regression results showed that the effect of expected value in the experience format was stronger than that of overall probability of winning. Yet, there was no corresponding preference for alternatives associated with increased expected value in the description format. These results suggest that the natural mean heuristic and expected value maximization play an important role in decisions from experience for complex gambles, similar to prior findings with binary gambles (Hau et al., 2008). Critically, the differences in choice preferences across description and experience formats may be explained by the use of systematically different decision strategies across these formats.

Our findings extend our understanding of the use of decision strategies in risky choice across presentation formats. First, this study investigated the description-based and experience-based decision making outside of canonical binary gambles, which have been extensively used in previous studies (Rakow & Newell, 2010). Even though there are numerous studies investigating experience-based decisions using multiple alternative-multiple outcome tasks, they were either heavily focused on binary outcome gambles with multiple alternatives or binary alternative gambles with multiple outcomes, or they did not directly compare preferences between experience- and description-based formats

(Hills et al., 2013; Newell, Rakow, Yechiam, & Sambur, 2015; Noguchi & Hills, 2015; Teodorescu & Erev, 2014; Wulff, Hills, & Hertwig, 2015; Yechiam, Rakow, & Newell, 2015). Here, we examined risk preferences in description-based and experience-based decision making using a three-outcome mixed gamble task and using a within-subject design to directly compare different decision strategies across the two presentation formats. Second, unlike the commonly used framework for D-E gap that involves alternatives with rare outcomes, we demonstrate preference reversals across the two formats using three-outcome mixed gamble task with equal probability and with slight variations from the equal probability. Some recent studies have demonstrated similar D-E gap in equal probability binary gamble tasks (Ludvig et al., 2014; Ludvig & Spetch, 2011), but in this study, we examined how the valence of outcomes in a choice set influences decisions in the description format and in the experience format without involving any rare event. Lastly, we try to relate any differences across the formats to differences in the underlying decision strategies across the two formats.

We found that both the natural mean and the Pwin heuristics were significant predictors of choice across both studies in decisions from experience. Participants essentially treated the gambles across the Pwin available and Pwin unavailable types as homogenous in the experience condition, but not description condition where they switched away from the Pwin heuristic when a Pwin alternative was not available. This is also evident from the fact that participants took longer for Pwin unavailable trials relative to Pwin available trials in the description condition, but there were no differences in the number of samples across problem types in the experience condition. Critically, the experienced expected value was the strongest predictor of choice in decisions from

experience. However, expected value did not significantly predict choice in decisions from description when it was explicitly modulated in Study 2, arguing again in favor of explicit differences in decision strategies across the two formats using very similar tasks. In summary, our findings suggest that presentation format plays a crucial role in shaping decision strategies. However, the current study has several limitations that need to be explored further in subsequent studies. First, we could not systematically dissociate the role of expected value and overall probability of winning in predicting choices in the experience condition, since these variables were highly correlated in both studies. For example, the correlations between experienced overall probability of winning and experienced expected value were really high ($r(67) = .75$ in Study 1, $r(64) = .91$ in the follow-up study of Study 1, and $r(45) = .77$ in Study 2). Even though we tried to dissociate the effect of overall probability of winning and expected value in Study 2 by employing diverse sets of stimuli, they were still highly correlated. Therefore, although the current studies used relatively more complex gambles compared to the binary gambles in previous studies, we might need to extend it further to include five-outcome mixed gambles in future studies to successfully dissociate the roles of experienced EV and probability of winning in the experience condition. These gambles may also make it easier to explicitly separate EV choices from Gmax, Lmin and Pwin alternatives in the description condition (Venkatraman et al., 2014), allowing us to evaluate differences in the decision strategies between experience and description conditions more robustly.

CHAPTER 5.

ESSAY 3. CHOOSING VERSUS REJECTING:

THE EFFECT OF DECISION MODE ON SUBSEQUENT PREFERENCES

Preferential choice decisions can be made largely in two paths. One might make a choice decision by choosing an alternative which they prefer the most among the available alternatives in a choice set, while another might decide by rejecting less attractive alternatives in the choice set (Shafir, 1993). For example, an individual purchasing a portable charger searches items in Amazon and put several items in the cart. Then, the person can purchase a portable charger by choosing the most attractive one or by rejecting less attractive ones until he or she has one leftover. Previous studies have shown that these *choose* and *reject* decision frames affect consumer judgment and decision-making, leading to a higher preference for hedonic items in rejection (Dhar & Wertenbroch, 2000), larger choice set in rejection (Huber et al., 1987; Levin, Jasper, & Forbes, 1998; Nagpal, Lei, & Khare, 2015), larger product option size in rejection (Park, Jun, & MacInnis, 2000), preference reversals (Shafir, 1993), higher cognitive effort and elaboration in rejection (Laran & Wilcox, 2011; Sokolova & Krishna, 2016), and different information processing patterns and neural mechanisms (J. Chen & Proctor, 2017; Foo et al., 2014).

This essay investigates the underlying cognitive processes of the *choose* vs. *reject* decisions and their effect on subsequent consumer preferences using a 2-part binary choice task. Even though previous *choose* vs. *reject* preference reversal literature has shown that people show inconsistent preferences depending on task framing in a single

decision context, it has not yet been understood how these two decision frames affect consumer preferences in subsequent preferential choices. This research question is important since in majority cases, consumers confront the same choice or purchase decision situations repeatedly (e.g., purchasing a shampoo or a car), and they often employ different decision strategies even under the same decision problem (Venkatraman et al., 2009; Venkatraman et al., 2014). Current research investigates whether the *choose* vs. *reject* decision strategies people employed in a previous choice symmetrically influence subsequent preference for the previously considered items.

Across four experiments, using a multi-method approach that includes behavioral, eye-tracking, and computational modeling, I found that *choose* frame increased subsequent preferences, while *reject* frame neither increased nor decreased subsequent preferences. The results of an eye-tracking study and a drift diffusion modeling suggested that the asymmetric effect of *choose* vs. *reject* decision frames on subsequent preferences might be contributed by more deliberation in *reject* frame, while the more selective process in *choose* frame.

Explanations of Choosing vs. Rejecting Decision Processes

The compatibility hypothesis

In a seminal study, Shafir (1993) asked participants to imagine that they were part of the jury in a sole-custody case. They were given descriptions of two parents: Parent A had average income, average health, average working hours, reasonable rapport with the child, and relatively stable social life, whereas Parent B had above-average income, very close relationship with the child, extremely active social life, lots of work-related travel,

and minor health problems. He asked one group of participants to indicate which parent they wanted to award sole custody of the child (*choose* frame), while he asked the other group to indicate which parent they wanted to deny sole custody of the child (*reject* frame). Surprisingly, 64% of the participants in the *choose* frame awarded sole custody of the child to Parent B. On the other hand, 55% of the participants in the *reject* frame denied sole custody to Parent B, showing that an alternative (here, Parent B) can be both better and worse depending on the *choose* vs. *reject* decision frame.

The *choose* vs. *reject* preference reversal was attributed to the selective focus on the compatibility of attributes to the given task frame (Shafir, 1993). For example, Parent A has average values for all of the five attributes (income, health, working hour, rapport with the child, and social life), while Parent B has more extreme values for those attributes (both in favorable and unfavorable directions). In the *choose* frame, participants are looking for the reasons why a parent should be awarded the sole custody, while in the *reject* frame, they are looking for the reasons why a parent should be denied the sole custody. Parent B can easily provide justifiable reasons both to award (above average income, very close relationship with the child) and to deny sole custody (minor health problem, extremely active social life, a lot of work-related travel). This selective confirmatory information processing in each decision frame leads to the *choose* vs. *reject* preference reversal.

Nagpal and Krishnamurthy (2008) showed supporting evidence that participants in the *choose* frame made faster decision for an attractive items pair than for an unattractive items pair, while those in the *reject* frame made faster decision for an unattractive items pair than for an attractive items pair. Moreover, participants felt less

difficulty, lower effort, and higher memory recall of attributes in the compatible conditions (*choose*-attractive items pair vs. *reject*-unattractive items pair). More recently, Perfecto, Galak, Simmons, and Nelson (2017) also showed that people were more confident and predicted consensus for their decisions from others in the compatible conditions (e.g., *choose*-attractive faces pairs and *reject*-unattractive faces pairs) than in the incompatible condition (e.g., *choose*-unattractive faces fairs and *reject*-attractive faces pairs). Meloy and Russo (2004) investigated the underlying processes using a think-aloud protocol and found that task-compatible attributes (positive attributes in *choose* vs. negative attributes in *reject*) were weighted more during the decision processes, showing supporting evidence of the compatibility hypothesis.

This compatibility hypothesis posits that people make choose decisions by comparing positive aspects of alternatives, while they make reject decisions by comparing negative aspects of the alternatives. Therefore, based on this theory, the activated positive features in *choose* frame will increase subsequent attractiveness and preference for the chosen item, while the activated negative features in *reject* frame will decrease subsequent attractiveness and preference for the non-rejected item. Additionally, if people focus on task-compatible information, there should be no difference in response time between *choose* vs. *reject* decision frames.

Commitment Hypothesis

Ganzach (1995) attempted to replicate the Shafir (1993)'s findings using a job candidates selection task with variations in attribute levels of job candidates. He showed that participants preferred candidates who had higher variance across attributes (enriched

alternatives) in the *reject* frame than in *choose* frame, in contrast to the findings from Shafir (1993). Ganzach explained his results using Commitment hypothesis, which posits that people mostly make *choose* decisions in daily life and hence they commit more in the *choose* frame than in the *reject* frame. The greater commitment in the *choose* frame results in processing both positive and negative information more than in *reject* frame. However, negative aspects of an alternative are weighted more than positive aspects, and this leads to a lower preference for the enriched alternatives in *choose* frame than in *reject* frame. Since the negative aspects are weighted more in *choose* frame than in *reject* frame, the *choose* frame will decrease subsequent preference for the considered items. Additionally, decisions in the *choose* frame will be associated with longer response time than those in the *reject* frame, due to increased commitment and deliberation.

Accentuation Model

Wedell (1997) attempted to reconcile the inconsistent results of Ganzach (1995) and Shafir (1993). He pooled the data from both studies and compared the correlation between overall preference for the enriched alternatives and the percentage of participants preferring the enriched alternatives in the *choose* vs. *reject* decision frames. The results showed that the overall preference for the enriched alternatives was different in the two previous studies. While only one experiment among seven experiments in Shafir (1993) showed lower than 50% overall preference for the enriched alternative, all studies in Ganzach (1995) showed lower than 50% overall preference for the enriched alternatives. Wedell (1997) argued that people are indeed more committed in *choose* frame than in *reject* frame as the commitment hypothesis posits, and the higher

commitment works in a way to accentuate the differences between alternatives. For example, in Ganzach (1995), the overall preference for the enriched alternatives was lower than 50%. The negative aspects of alternatives are accentuated and this results in a lower preference for the enriched alternatives in the *choose* frame. On the other hand, in Shafir (1993), the overall preference for the enriched alternatives was greater than 50%. The positive aspects of alternatives are accentuated, and this leads to a higher preference for the enriched alternative in the *choose* frame. Using a logistic regression model, he found a higher discriminant parameter value (i.e., a steeper regression coefficient) in the *choose* frame than in the *reject* frame. Critically, he also found that, depending on the baseline preference of the enriched alternatives, different attributes were accentuated. More recently, J. Chen and Proctor (2017) successfully replicated the findings of the accentuation model.

One additional claim of accentuation model is that the lower commitment in *rejection* is contributed by higher cognitive demands in *rejection*. Preferential choice decisions in our daily lives are mainly based on choose decisions, so choosing is a default decision mode. On the other hand, rejection is a less common task for people, requiring inhibition of the default decision mode. Similar to default-intervention dual process theory (e.g., Evans, 2007), rejection is a difficult task for people and requires great deals of cognitive resources. This processing difficulty leads to more elaboration and deliberative processes (J. Chen & Proctor, 2017; Laran & Wilcox, 2011; Sokolova & Krishna, 2016). Sokolova and Krishna (2016) tested the effect of *choose* vs. *reject* decision frames on the classic gain vs. loss framing effect using the Asian Disease problem. The results showed that the framing effect was weaker in the *reject* frame (gain

frame: 48% vs. loss frame: 43%) than in the *choose* frame (gain frame: 76% vs. loss frame: 42%), indicating that people were more deliberate in *reject* frame than in *choose* frame. J. Chen and Proctor (2017) also showed that people in the *reject* frame recalled more negative attributes of alternatives than those in the *choose* frame, while there was no difference in recalling positive attributes of the alternative between the two decision frame conditions. Indeed, both studies showed that people spent a longer time in the *reject* frame than in the *choose* frame.

Based on the accentuation model, the *choose* frame accentuates the difference between alternatives, making difference in positive aspects between items more salient. The accentuated positive features of the preferred item will increase subsequent preferences in the *choose* frame. This is also supported by choice-induced preference literature, which has shown that choice not only reveals people's preferences, but also constructs preferences (for a review, Ariely & Norton, 2008), and often its effect persists long time (Sharot et al., 2012). For example, Sharot, Velasquez, and Dolan (2010) investigated pre- and post-choice attractiveness ratings changes and found that people showed an increased preference for the chosen items, while they showed a decreased preference for the non-chosen items after making choice decisions. Greenberg and Spiller (2015) also investigated post-choice preference changes in *whether or not* decision frame and in *which one* decision frame. The authors found higher post-choice preference changes in the *which one* frame than in the *whether or not* frame. Thus, the accentuation model predicts that the *choose* frame will increase preference for the chosen items in subsequent preferential choice decisions.

On the other hand, the accentuation model posits that rejection involves more deliberative processes, so the *reject* frame will not affect subsequent preferences. This was also partly supported by Sokolova and Krishna (2016)'s findings that people were less influenced by outcome framing in the *reject* frame than in the *choose* frame. Similarly, the effect of *reject* frame on subsequent preferences will be also weaker than that of the *choose* frame, and *reject* decisions will take longer time than *choose* decisions. Even though the choice-induced preference changes have been studied widely, how rejection changes post-choice preferences has not yet been understood well. In Study 3, employing the choice-induced preference experiment paradigm, I investigate the post-choice preference changes depending on the *choose* vs. *reject* decision frames. Based on the accentuation model, the results will show a greater post-choice preference changes in *choose* frame than in *reject* frame.

The role of attention on subsequent choice

Attention not only reveals our preferences but also shapes preferences (Krajbich, Armel, & Rangel, 2010; Krajbich & Rangel, 2011; Shimojo, Simion, Shimojo, & Scheier, 2003; Venkatraman et al., 2014). Shimojo et al. (2003)'s gaze cascade model argues that gaze helps in the formation of preferences during the decision processes. Using eye-tracking in a binary choice task involving faces, the authors found that people allocated similar amounts of attention to both faces in earlier decision stages, but they allocated more attention to the preferred faces in later decision stages. Krajbich et al. (2010) introduced the attention factor in the classic drift diffusion model and demonstrated that fixations drive evidence accumulation. These findings indicate that in addition to

revealing a decision maker's preference, attention can also shape preferences during the decision processes.

It has also shown that attention influences subsequent preferences. Attention inhibition literature has shown that people show decreased emotional responses for ignored images than attended images in post-choice evaluations (Fenske & Raymond, 2006; Raymond, Fenske, & Tavassoli, 2003). Janiszewski, Kuo, and Tavassoli (2012) investigated the effect of selective attention on subsequent consumer preferences and found that participants chose attentive items more when the attended items were presented together with a foil item in a subsequent trial. The findings indicate that the more attention in *choose* frame not only influences choice decisions *per se* but also could have long-term effects.

Study 4 investigates whether attention is differentially allocated in *choose* vs. *reject* decision frames, and how the different allocation patterns affect subsequent preferences for the item. Specifically, I use eye-tracking to understand the decision processes in the two frames, and how they relate to the predictions of the three different theories presented above. The compatibility hypothesis predicts that people in *choose* frame will focus more on the preferred item (chosen item) than on the non-preferred item (non-chosen item), while those in *reject* frame will focus more on the non-preferred item (rejected item) than the preferred item (non-rejected item). However, there will be no difference in the amount of time spent on the preferred item in the *choose* frame and the non-preferred item in the *reject* frame. The commitment hypothesis predicts that people will focus on both preferred and non-preferred items more in the *choose* frame than in the *reject* frame, since people are more committed in choosing than in rejecting. The

accentuation model predicts that people in the *choose* frame will focus more on the preferred item than on the non-preferred item, and the difference in the amount of attention on preferred and non-preferred items will be greater in the *choose* frame than in the *reject* frame.

Study 1. The Effect of Decision Frames on Subsequent Preferential Choices

Study 1 sought to investigate the effect of *choose* vs. *reject* decision frames on subsequent preferences using a 2-part free choice paradigm (Egan, Santos, & Bloom, 2007; Janiszewski et al., 2012). A binary choice task enables us to directly compare preferences for items across the two decision frames. For example, when deciding between eating an apple or an orange, choosing the apple or rejecting the orange indicate that the apple is preferred to the orange. Further, I wanted to see if the relative preference for the apple varies as a function of the two decision frames in an independent and subsequent decision involving another fruit (e.g. mango).

Method

A total of 72 university undergraduate students (M_{age} = 19.64 years, SD = 1.52, female = 26) participated in the experiment in exchange for class credit. All studies were approved by the Institutional Review Board at the university where the data was collected. Participants were randomly assigned to either *choose* condition or *reject* condition. Both groups were asked to make a series of binary decisions across two phases (Phase 1 and Phase 2) for 48 different item categories. Each item category consisted of three similar items. In each trial, two of these items were randomly selected and presented as decision

alternatives in Phase 1, and the remaining third item was used as an alternative in Phase 2. In Phase 1, participants in the *choose* condition were asked to choose one of the two alternatives that they liked, while those in the *reject* condition were asked to reject one of the two alternatives that they disliked. In Phase 2, participants in the *choose* condition were asked to make a second choice between the chosen item from Phase 1 and the remaining item from the same category (new item), while those in the *reject* condition were asked to choose between the non-rejected alternative from Phase 1 and the new item. In Phase 2 therefore, both groups of participants were always asked to choose between one of the two items (the preferred item from Phase 1 vs. new item). I was specifically interested in whether the relative preference for the preferred item varied as a function of whether it was chosen or not-rejected in Phase 1.

Results

Response Time: Using a 2 (phase: Phase1 vs. Phase2; within-subject) × 2 (decision frame: choose vs. reject; between-subject) mixed ANOVA on mean response times across all 48 trials, I found a significant main effect of phase ($F(1, 70) = 210.37$, $p < .001$, $\eta_p^2 = 0.75$), showing that people spent longer time in Phase 1 than in Phase 2. The main effect of decision frame was not significant ($F(1, 70) = 0.38$, $p = .538$, $\eta_p^2 = 0.03$), but the interaction between phase and decision frame was significant ($F(1, 70) = 6.21$, $p = .015$, $\eta_p^2 = 0.08$; Figure 9). Further contrasts revealed that participants in the *reject* condition spent longer time than those in the *choose* condition in Phase 1 ($M_{\text{choose}} = 4.14$ sec., SD = 1.24 vs. $M_{\text{reject}} = 4.61$ sec., SD = 0.31; $F = 8.06$, $p = .006$), but there was no difference between the two decision frame conditions in Phase 2 ($M_{\text{choose}} = 2.73$ sec., SD = 0.94 vs.

M_{reject} = 2.62 sec., SD = 0.18; F = 0.447, p = .495). The longer response time in the *reject* condition was consistent with a previous study (J. Chen & Proctor, 2017; Foo et al., 2014; Perfecto et al., 2017), implying that rejection task is more difficulty and involves greater effort than *choose* task (Sokolova & Krishna, 2016). The non-significant difference in RT in Phase 2 indicates that decision latencies in Phase 2 were not affected by the decision frames employed in Phase 1.

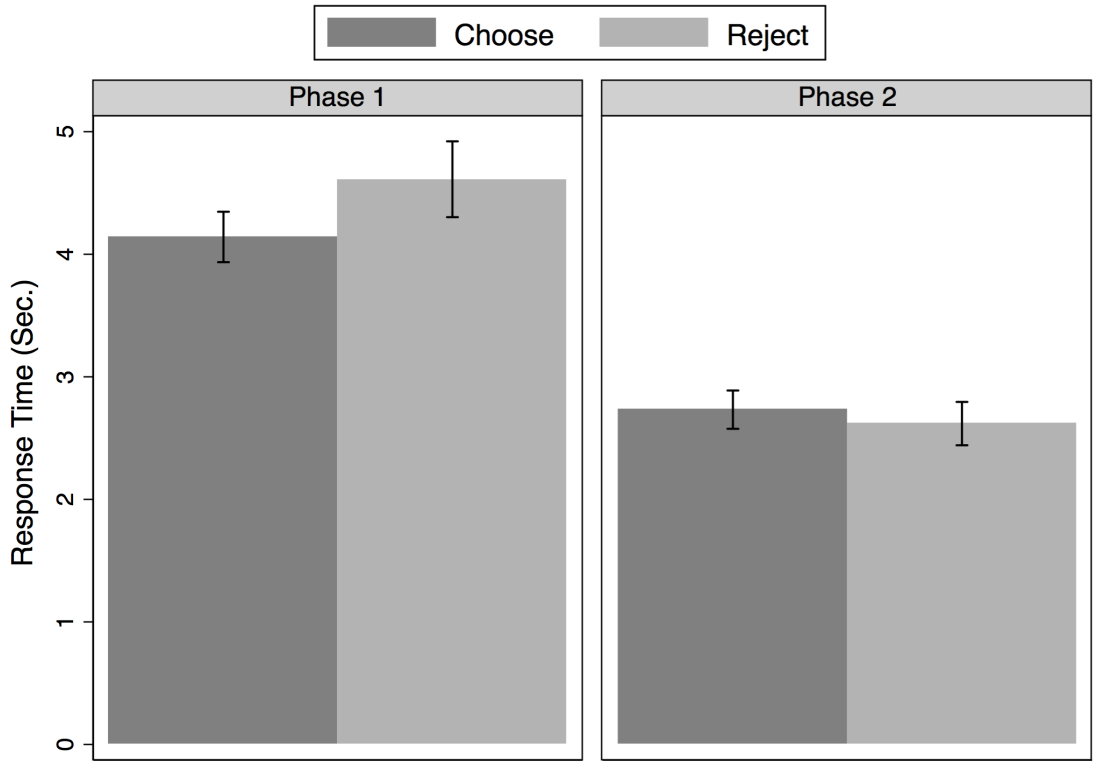

Figure 9. Mean response time across two phases in Study 1. The error bars indicate standard errors.

Choice Share: To examine the effect of *choose* vs. *reject* decision frames on subsequent preferences, the choice shares of the preferred items (chosen items in the *choose* condition or not-rejected items in the *reject* condition) in Phase 2 were compared as a

function of the two decision frames in Phase 1. An independent-sample *t*-test result showed that the *choose* frame resulted in a higher preference for the preferred items than the *reject* frame (M_{choose} = 0.70, SD = 0.08 vs. M_{reject} = 0.64, SD = 0.13; $t(70) = 2.37$, $p = .021$, $d = 0.56$; Figure 10). Further analysis was conducted to test whether the choice share of the preferred items was different from the chance-level ($p = .66$; M. K. Chen & Risen, 2010) in the two decision frame conditions. One-sample *t*-test result for each condition showed that the *choose* frame led to the selection of the preferred items significantly more than the chance level ($t(35) = 3.13$, $p = .004$, $d = 0.52$), while the *reject* condition did not ($t(35) = -0.82$, $p = .416$, $d = 0.15$). The findings suggest that while the *choose* decision frame increased subsequent preferences for the chosen items, the *reject* decision frame did not influence preferences for the not-rejected items.

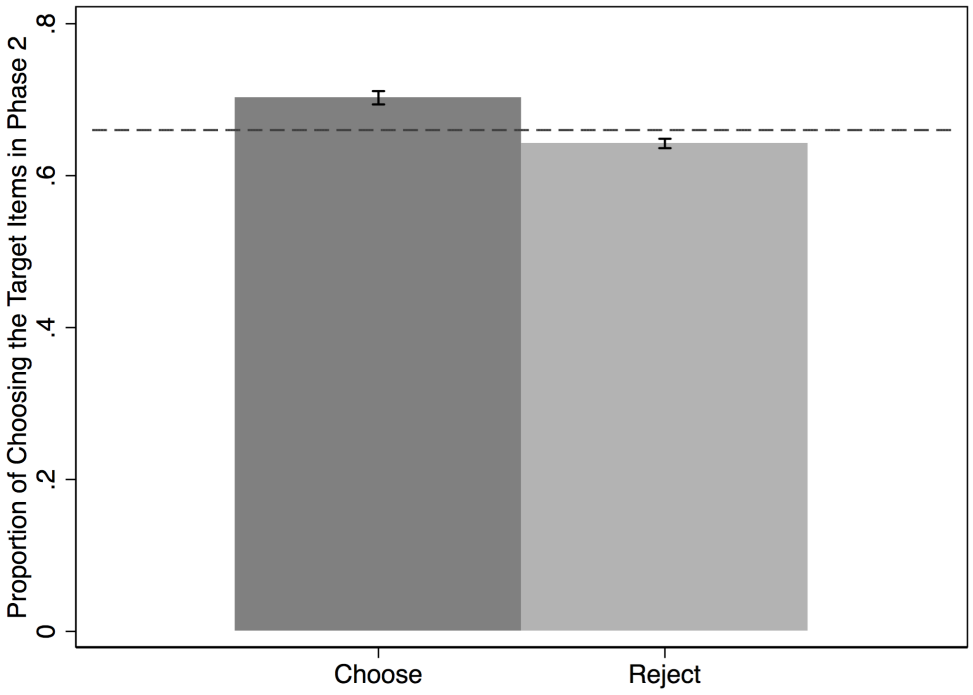

Figure 10. Mean choice proportions of the target items in Phase 2. The dash line indicates the chance level, and the error bars indicate standard errors.

Discussion

Study 1 investigated the effect of *choose* vs. *reject* decision frames on subsequent preferences using a 2-part binary choice task. Participants in the *reject* condition spent longer in making their decisions than those in the *choose* condition. This implies that rejecting a less attractive alternative requires more cognitive effort than choosing a more attractive alternative, consistent with the accentuation model (J. Chen & Proctor, 2017; Sokolova & Krishna, 2016). The results also showed that there was an asymmetric effect of the two decision frames on subsequent preferences. Specifically, the *choose* frame increases subsequent preferences, while the *reject* frame neither decreases nor increases subsequent preferences, supporting the prediction from the accentuation model.

In the following study, the effect of *choose* vs. *reject* decision frames on subsequent preferences are tested more directly by paring preferred items from both decision frames. In Study 1, participants made choice decisions in Phase 2 between a preferred item from Phase 1 and a new item. However, people may process the new items differently in the *choose* and *reject* decision conditions. To rule out this possibility, in Study 2, participants make choice decisions between a chosen item and a non-rejected item pair.

Study 2. Direct Comparison of Two Preferred Items

Method

A total of 52 university undergraduate students (M_{age} = 21.06 years, SD = 3.78, female = 32) were recruited for the experiment in exchange for class credit. Overall experimental procedures were similar to Study 1, except for a few small changes. A new

item was added to each of the 48 categories from Study 1, such that each category had four different items now. Among the four items in each category, two items were randomly selected and used for the *choose* task, while the remaining two items were used for the *reject* task. This enabled participants to make both *choose* and *reject* decisions for each item category in Phase 1. In Phase 2, participants were provided two different types of items pairs blocks: preferred item pair block (chosen vs. non-rejected item pairs from Phase 1) and non-preferred item pairs block (non-chosen vs. rejected items pairs from Phase 1). The preferred items pairs block first was always presented first since the main research interest was to examine the relative preference for the preferred items from the two decision frames. Consistent with Study 1, participants were always asked to choose one of the two items they liked more in Phase 2.

Results

Response Time: For the analysis, the mean values of each participant were compared. A paired-sample *t*-test result showed that participants spent longer time in the *reject* trials than in the *choose* trials (M_{choose} = 4.66 sec., SD = 2.44 vs. M_{reject} = 4.85 sec., SD = 2.41; $t(51) = -3.01$, $p = .004$, $d = -0.42$; Figure 11) in Phase 1 consistent with Study 1. In Phase 2, participants spent longer time in the preferred items pair trials than in the non-preferred item pair trials ($M_{\text{preferred}}$ = 2.36 sec., SD = 1.11 vs. $M_{\text{non-preferred}}$ = 2.14 sec., SD = 0.85; $t(51) = 2.84$, $p = .006$, $d = 0.39$). The longer response time for the preferred items pairs trials is inconsistent with previous studies which show that choosing between preferred items is typically faster than choosing between non-preferred items (Nagpal & Krishnamurthy, 2008; Perfecto et al., 2017). However, the results may also be explained

by the effect of order in this study - the preferred item pairs block was always presented first followed by the non-preferred items pairs block.

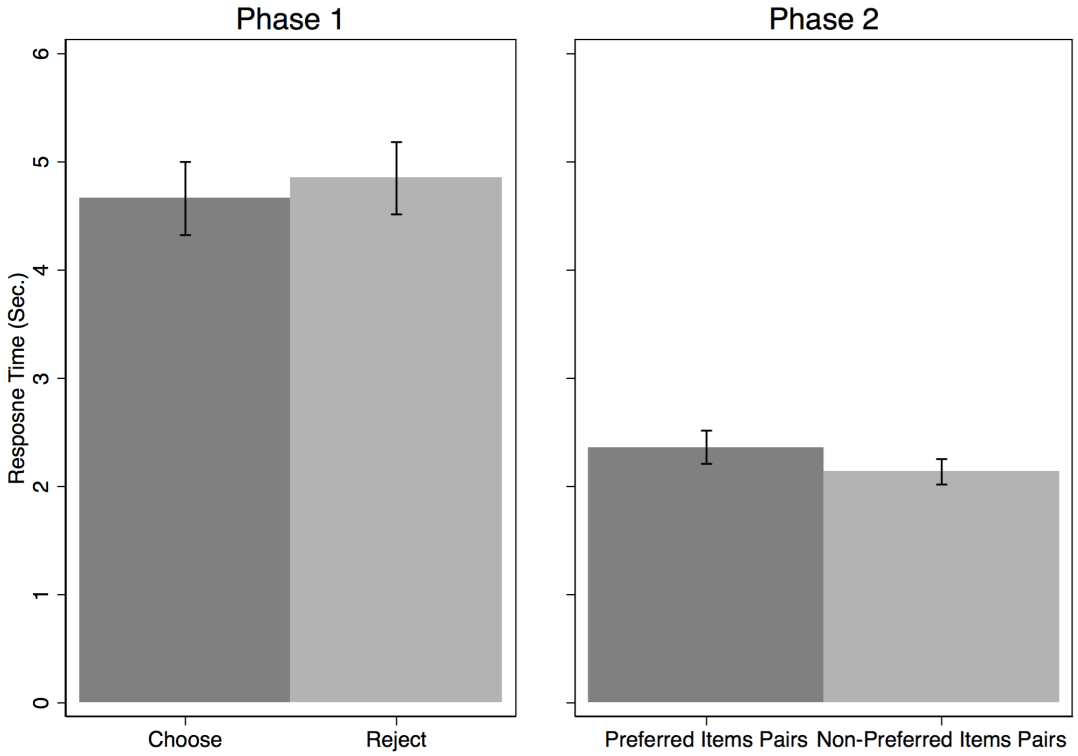

Figure 11. Mean response time across two phases in Study 2. The error bars indicate standard errors.

Choice Share: To examine relative preference between chosen items and non-rejected items, I computed the choice share of selecting chosen item in Phase 2 and tested whether this choice share was different from the chance level (p = .50). One-sample t-test result showed that participants preferred the chosen items to the non-rejected items, and the choice share was significantly above the chance level (M = 0.55, SD = 0.07; $t(51)$ = 4.67, $p < .001$, d = 0.65; Figure 12). I also compared whether the choice share of the rejected items was different from the chance level. However, the proportion of choosing the

rejected items in Phase 2 was not different from the chance level (M = 0.48, SD = 0.08; $t(51)$ = -1.38, p = .173, d = -0.19). These results are consistent with Study 1 indicating that the *choose* frame increases preference for the chosen items, while the *reject* frame does not affect preference for the rejected items in subsequent preferential choices.

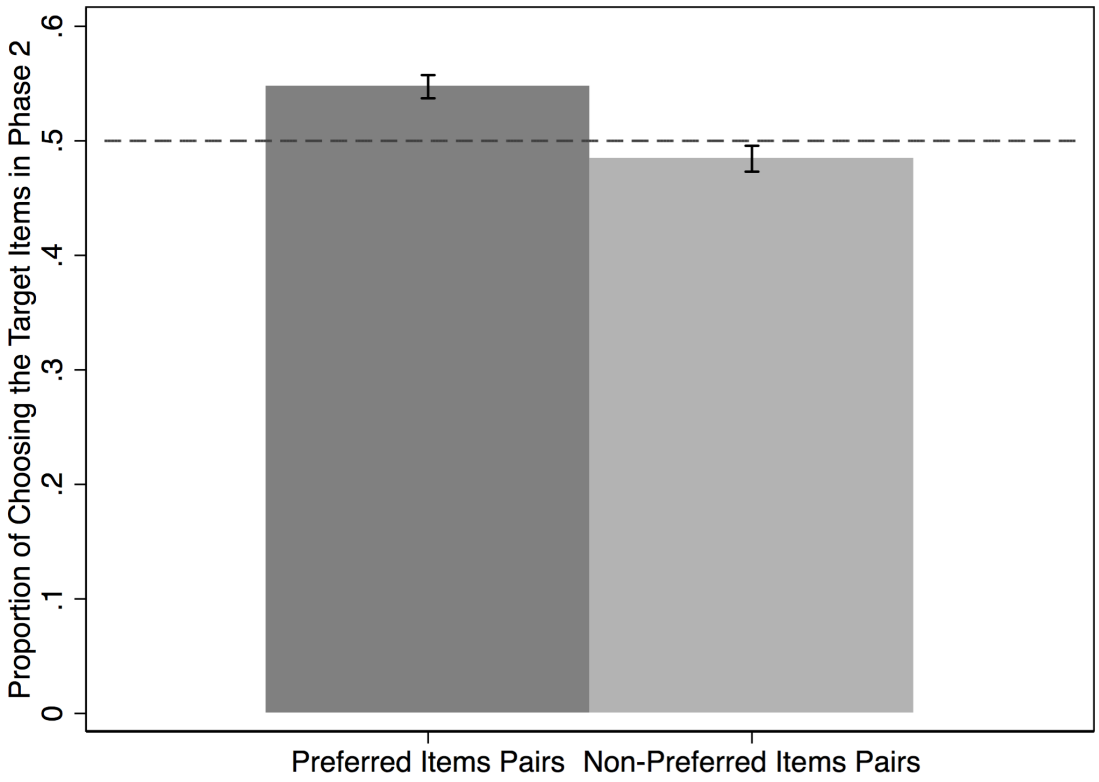

Figure 12. Mean choice proportions of the target items in phase 2. The dash line indicates the chance level, and the error bars indicate standard errors.

Discussion

Study 2 investigated the effect of *choose* vs. *reject* decision frames on subsequent preferences more directly by pairing preferred items from both decision frames. Consistent with the results of Study 1, participants preferred the chosen items to the non-rejected items in Phase 2, showing that *choose* decision frame does increase subsequent preferences. Study 2 also examined the effect of decision frames on non-preferred items

pairs. The results showed that participants did not discriminate between the non-chosen items and the rejected items, implying that the *reject* frame does not decrease subsequent preferences. The response time results were also consistent with those of Study 1. Participants spent a longer time in the *reject* frame than in *choose* frame.

The results of choice and response time analyses supported the accentuation model. The compatibility hypothesis predicts that participants would show a higher preference for the chosen items than for the non-rejected items and a lower preference for the rejected than for the non-chosen items. The commitment hypothesis predicts that participants would show a lower preference for the chosen items and non-chosen items than for the rejected and non-rejected items. However, the results of Study 2 did not support either explanation. Instead, as predicted by the accentuation model, the *choose* frame increased subsequent preferences, while the *reject* frame neither increased or decreased subsequent preferences. The longer response time in the *reject* frame than in the *choose* frame also indicates that more deliberation in the *reject* frame than in the *choose* frame, as predicted by the accentuation model.

Study 3. Post Decision Preference Changes

The previous two experiments showed that the *choose* frame increased subsequent preferences, while the *reject* frame neither increased nor decreased subsequent preferences. Indeed, these findings were robust to experimental design (significant results both in within-subject and between-subject design studies). Study 3 examined whether the asymmetric effect of decision frames on subsequent preferences was contributed by

the difference in post-choice preference changes. The choice-induced preference literature has shown that people show increased preference for the items they chose, while they show decreased preference for the non-chosen items (e.g., Sharot et al., 2010). Both the compatibility hypothesis and the accentuation model predict that post-choice preference will increase for the preferred items but will decrease for the non-preferred items in both decision frames. However, the compatibility hypothesis predicts a similar amount of post-choice preference change in both decision frames, while the accentuation model predicts no post-choice preference change or a smaller change in the *reject* frame than in the *choose* frame.

Furthermore, to better understand the underlying processes in *choose* and *reject* decision frames, I analyzed the response time data using a drift-diffusion model. The drift diffusion model assumes that in making judgments and decisions, people accumulate evidence with a random Gaussian noise and make decisions when the accumulation process reaches to a boundary (Figure 13). The drift diffusion model classifies response time data into several segments, including non-decision time (time for encoding stimuli and motor response; *T* in the Figure 13) and its variation, drift rate (representing how fast and easily one can extract information and accumulate evidence; *V* in the Figure 13) and its variation, boundaries (representing speed-accuracy tradeoff; *a* in the Figure 13), and bias toward either boundary (*Z* in the Figure 13) and its variation (for a recent review, see Ratcliff, Smith, Brown, & McKoon, 2016). If rejection is simply an inverse of choose decision, participants will show longer non-decision time in the *reject* frame than in the *choose* frame. Similarly, if participants can extract information more easily in *choose* frame than in *reject* frame, the drift rate will be greater in the *choose* frame than in the

reject frame. However, if participants are more deliberate in the *reject* frame than in the *choose* frame, the boundary will be greater in the reject frame than in the choose frame (Philiastides & Ratcliff, 2013).

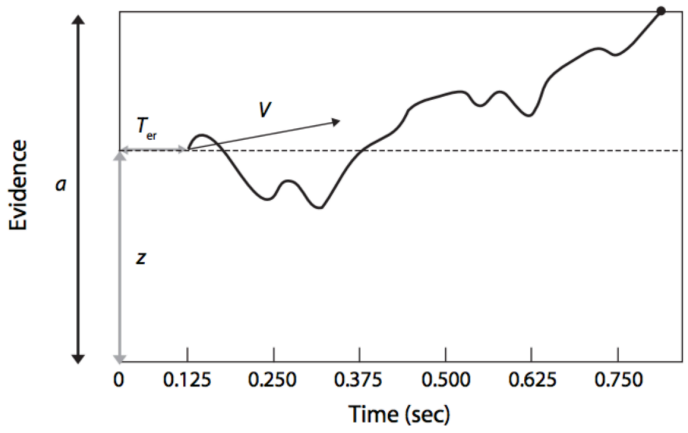

Figure 13. An illustration of diffusion processes. The illustration is from Vandekerckhove and Tuerlinckx (2008).

Method

A total of 106 university undergraduate students were recruited for the experiment in exchange for class credits (M_{age} = 21.38 years, SD = 3.97, 71 female). The study consisted of three phases. In Phase 1 (first rating), participants were asked to rate the attractiveness of 40 items using an 8-point Likert scale (1 = not attractive at all, 8 = very attractive). In Phase 2 (decision phase), the 40 items were rank ordered based on ratings from Phase 1, and paired with similarly attractive items. Participants were randomly assigned to either choose condition (N = 53) or reject condition (N = 53) and made decisions between the pairs of items on each trial. In the last phase (second rating), participants were asked to rate the attractiveness of all 40 items again using the same 8-point Likert scale as Phase 1.

Results

Response Time: The results of a 2 (decision frame: choose vs. reject; between-subject) × 2 (phase: first rating vs. decision vs. second rating; within-subject) mixed ANOVA found a significant main effect of phase ($F(2, 208) = 137.70$, $p < .001$, $\eta_p^2 = 0.57$). Participants were faster for the second rating than the first rating and the decision phase (Figure 14). The results also showed a significant decision frame × phase interaction ($F(2, 208) = 11.01$, $p < .001$, $\eta_p^2 = 0.10$). Further contrasts revealed that there was no difference in response time between decision frames in the first rating ($M_{choose} = 3.62$ sec., SD = 1.16 vs. $M_{reject}t = 3.70$ sec., SD = 1.28; $F = 0.26$, $p = .610$) and in the second rating ($M_{choose} = 2.17$ sec., SD = 0.75 vs. $M_{reject} = 1.97$ sec., SD = 0.64; $F = 1.81$, $p = .180$). However, the response time in the decision phase was significantly different between the two decision frame conditions ($M_{choose} = 3.13$ sec., SD = 1.27 vs. $M_{reject} = 3.89$ sec., SD = 1.53; $F = 26.01$, $p < .001$), showing that people in the *reject* condition spent a longer time than those in the *choose* frame in making their decisions. The response time results in Phase 2 were consistent with the results of Study 1 and Study 2. The main effect of decision frame was not significant ($F(1, 104) = 1.30$, $p = .256$, $\eta_p^2 = 0.03$).

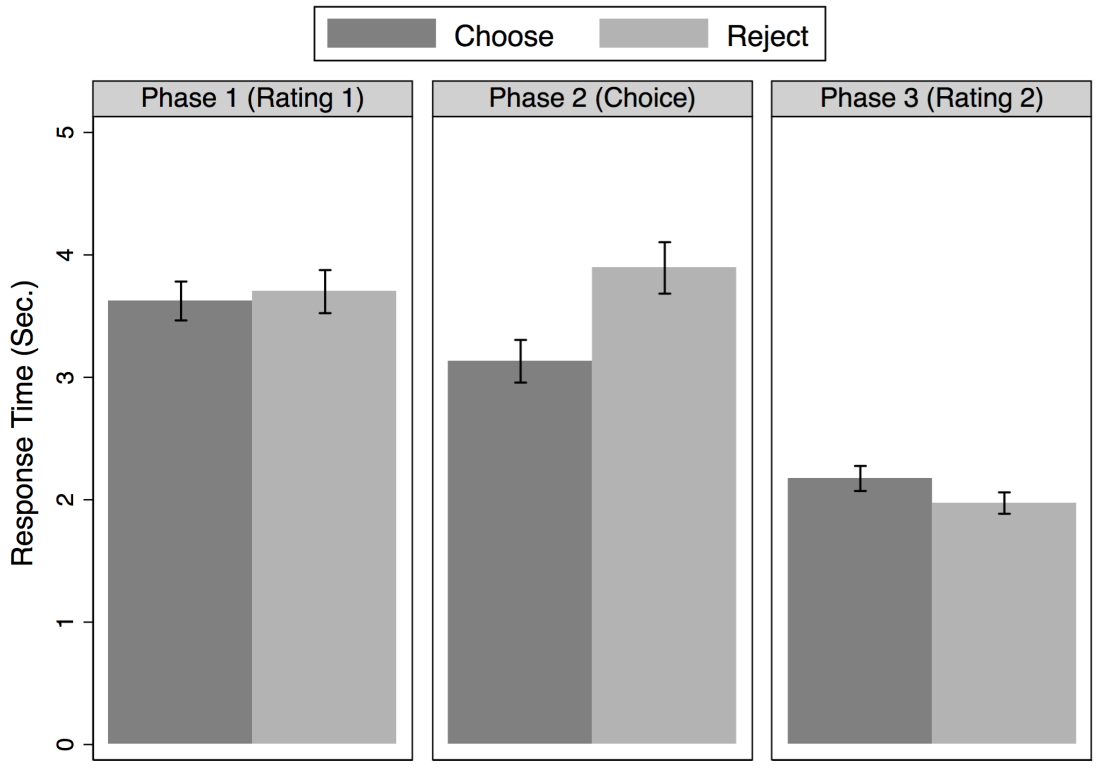

Figure 14. Mean response time across the three sessions in Study 3. Error bars indicate standard errors.

Rating Changes: The main dependent variable is the post-decision attractiveness rating changes. The post-choice rating changes were calculated by subtracting the first ratings from the second ratings after demeaning the attractiveness ratings in each session separately at individual subject level (Sharot et al., 2012). For the analysis, I regressed the rating changes on decision frame (choose vs. reject; between-subjects) and preference (preferred vs. non-preferred items; within-subjects) using a multilevel linear regression model. To account for the dependency between items paired together in the decision phase, I introduced random intercepts for subject and trial of the decision phase (Phase 2) in the model.

The results showed a significant main effect of preference (b = 0.98, SE = 0.06, z = 16.66, p < .001) and a marginally significant interaction between decision frame and preference (b = -0.16, SE = 0.08, z = -1.89, p = .059). Participants showed increased preference for the preferred items than the non-preferred items in both decision frames, but the rating changes were greater in the *choose* frame ($M_{\text{preferred}}$ = 0.49, SD = 1.51 vs. $M_{\text{non-preferred}}$ = -0.49, SD = 1.57) than in the *reject* frame ($M_{\text{preferred}}$ = -0.41, SD = 1.43 vs. $M_{\text{non-preferred}}$ = -0.41, SD = 1.50). This pattern indicates that post-decision preference changes were smaller in the *reject* frame than in the *choose* frame as predicted by the accentuation model. However, the main effect of decision frame was not significant (b = 0.08, SE = 0.07, z = 1.21, p = .227).

To rule out a possible impact of the difference in attractiveness ratings in Phase 2, I repeated the same analysis by limiting only to trials where items were paired with the same attractiveness ratings in the decision phase, and the results were identical. The main effect of preference (b = 1.06, SE = 0.06, z = 16.32, p < .001) and the interaction between decision frame and preference were statistically significant (b = -0.22, SE = 0.09, z = -2.44, p = .015).

Drift-Diffusion Model: To fit the model, the responses in trials where the paired items had the same attractiveness ratings in the decision phase were used. The model fitting was conducted using HDDM package embedded in Python 3.5 (Wiecki, Sofer, & Frank, 2013). The posterior distributions of the three parameters (non-decision time, drift rate, and boundaries) are presented in Figure 15. The results showed that there was no difference in non-decision time between the two decision frames (*the proportion of choose* > *reject* was .477), indicating that rejection is not a simple reverse of choosing.

There was no difference in drift rates between the two decision frames (*the proportion of choose > reject* was .850). The indifferent drift rates indicate that difficulties in accumulating evidence are not different between the *choose* frame and the *reject* frame (Philiastides & Ratcliff, 2013). However, there were different boundaries between the *choose* and the *reject* decision frames (*the proportion of choose > reject* was less than .001). The boundary in the *reject* frame was greater than that in the *choose* frame, indicating that participants in the *reject* frame were more deliberative than those in the *choose* frame (Voss, Nagler, & Lerche, 2013).

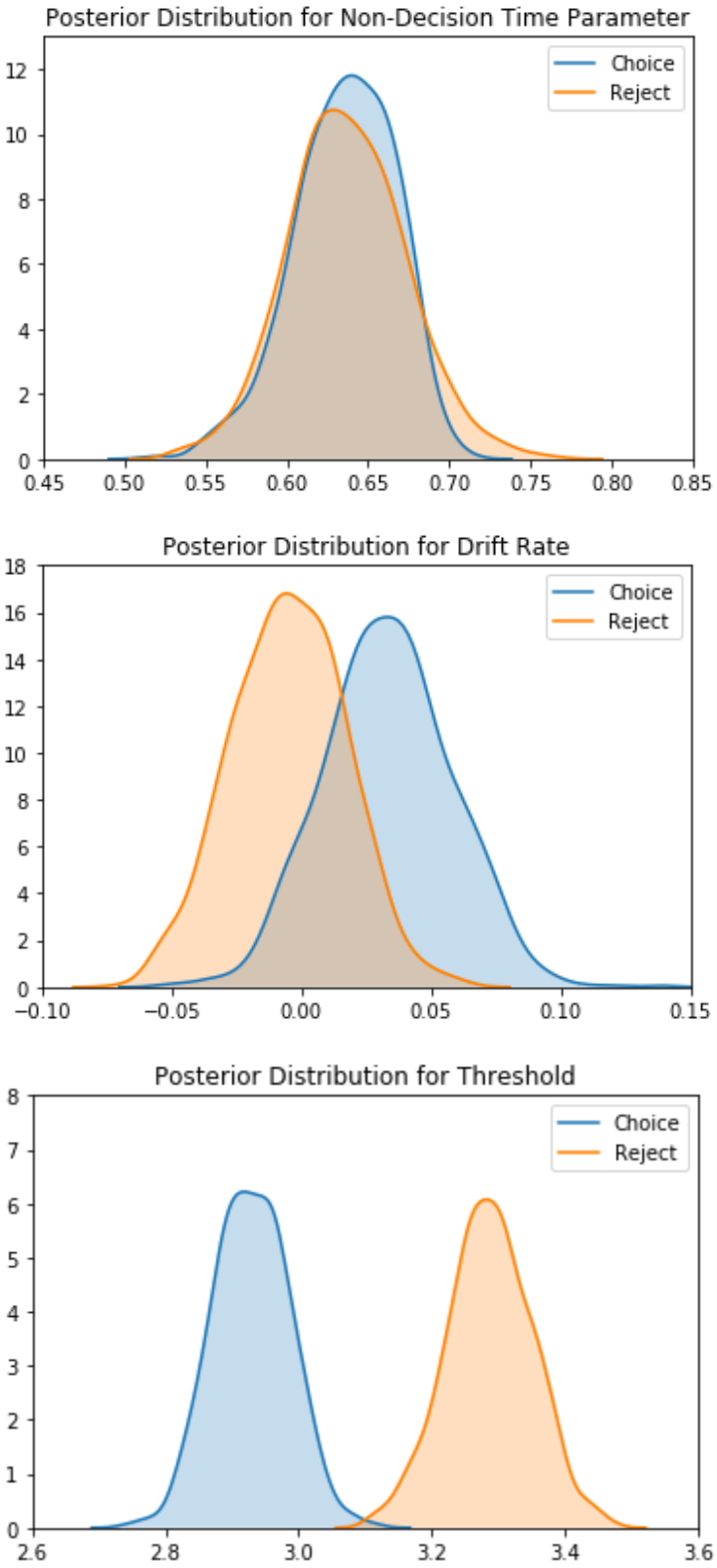

Figure 15. Posterior distributions of the non-decision time parameter (upper), drift rate (middle), and boundaries (threshold; bottom).

Discussion

Study 3 investigated post-choice preference changes in the *choose* and *reject* decision frames. Consistent with the choice-induced preference literature (D. Simon et al., 2008; D Simon, Krawczyk, & Holyoak, 2004), the results showed that preference for the preferred items increased, while preference for the non-preferred items decreased. Post-choice preference changes in the *reject* frame also showed the same pattern, but its magnitude was smaller than that in the *choose* frame. The accentuation model predicts that the difference between alternatives is accentuated more in the *choose* frame than in the *reject* frame. Consistent with the claim of the accentuation model, choice-induced preference was greater than reject-induced preference.

The results of a further analysis using a drift diffusion model also provided supporting evidence of the accentuation model. There was no difference in non-decision time and drift rate between the two decision frames, showing that people in the *reject* frame did not simply reverse their decisions after making choose decisions and that difficulties in processing features of alternative were not different in the two decision frames. However, the boundary in the *reject* frame was larger than that in the *choose* frame, implying that people were more deliberate in rejection. The overall results of post-choice preference change and the drift diffusion modeling indicate that people accentuate positive aspects in the *choose* frame, resulting in higher preference for the preferred items in subsequent choice decisions. However, the deliberate processes in the *reject* frame do not influence both preferred and non-preferred items in subsequent choice decisions.

Study 4. The role of attention on the choose vs. reject decision frame

Study 4 sought to investigate whether and how attention mediates the effect of *choose* vs. *reject* decision frames on subsequent preferences. It has been shown that people focus more on preferred items. Focusing on such an item for a long time could increase preference for that item in *choose* tasks (Krajbich et al., 2010; Krajbich & Rangel, 2011; Shimojo et al., 2003). Moreover, Janiszewski et al. (2012) tested the effect of attention on subsequent preferential choice decisions using market goods and found that participants preferred the attended items more in a subsequent choice task when the attended items were paired with foil items.

However, it is not clear how attention forms and predicts consumer preferences in *reject* tasks. To my knowledge, there was only one study that investigated gaze bias in the *choose* frame and the *reject* frame. Mitsuda and Glaholt (2014) examined how people allocate attention to alternatives in the *choose* frame and in the *reject* frame using images of faces and scenes. The authors found that participants showed higher gaze bias (i.e., allocating more attention to selected items (chosen or rejected items) than non-chosen items (non-chosen or non-rejected items) in the *choose* frame than in the *reject* frame. The findings are consistent with the accentuation model. Since individuals are more deliberate in the reject frame, so the gaze allocation between two items will be similar in the reject frame. According to the accentuation model and the Mitsuda and Glaholt (2014)'s findings, I predict that participants will focus more on preferred items than non-preferred items in the *choose* frame, while they will focus more on non-preferred items than preferred items in the *reject* frame. However, the difference in attention allocation

between preferred and non-preferred items will be greater in the *choose* frame than in the *reject* frame.

Methods

A total of 90 (M_{age} = 20.47 years, SD = 1.86, 51 female) university undergraduate students were recruited in exchange for class credit, and they were randomly assigned to either *choose* frame condition (N = 45) or *reject* frame condition (N = 45). Overall procedures were similar to Study 1, with slight variations. Participants completed the task while sitting in front of Tobii T60XL eye-tracker. Participants were informed not to move their head or lean on the chair too much. A 9-point calibration was used for the eye-tracker, and the experimenter monitored the eye-tracking signal during the task, while seated in the experimental room. A total of 267 trials which had lower than 70% valid eye-tracking signal were excluded from the analysis, including the exclusion of one participant's whole responses.

Results

Response Time: A 2 (frame: choose vs. reject; between-subject) × 2 (phase: phase 1 vs. phase 2; within-subject) mixed ANOVA results showed only a significant main effect of phase ($F(1, 87) = 63.45$, $p < .001$), showing that people spent a shorter time in Phase 2 than in Phase 1. Critically, unlike the findings reported by previous studies, there was no difference in response time between the two decision frames (M_{choose} = 5.12 sec., SD = 2.99 vs. M_{reject} = 5.31 sec., SD = 2.59; $t(87) = -0.33$, $p = .742$, $d = 0.07$). The non-significant difference in Phase 1 might be related to an overall increased engagement and

commitment in the eye-tracking study than in behavioral studies, since participants were passively monitored during the task. This is supported by the overall increase in response time in this study, relative to Study 1 (the mean response time in the *choose* frame and in the *reject* frame were 4.14 seconds and 4.61 seconds, respectively).

Choice Share: Participants in the *choose* frame showed slightly higher preference for the preferred items in subsequent choices than those in the *reject* frame, but the difference was not significant (M_{choose} = 0.68, SD = 0.13 vs. M_{reject} = 0.67, SD = 0.09; $t(87) = 0.59$, $p = .557$, $d = 0.09$). Indeed, the choice share of the preferred items in both frame conditions did not differ from the chance level (*choose* frame: $t(44) = 1.09$, $p = .281$, $d = 0.15$ vs. *reject* frame: $t(43) = 0.54$, $p = .593$, $d = 0.11$). The non-significant effect of decision frame on the subsequent preference choices might be because of the increased response time in Phase 1. The results of a further multilevel regression analysis showed that longer response time in the *choose* frame decreased the probability of choosing the target items in Phase 2 ($b = -0.04$, $SE = 0.01$, $z = -3.46$, $p = .001$), while longer response time in the *reject* condition did not influence the probability of choosing the target items in Phase 2 ($b = -0.02$, $SE = 0.01$, $z = -1.62$, $p = .106$). The effect of response time on subsequent choice was also same as in Study 1 (*choose*: $b = -0.08$, $SE = 0.02$, $z = -3.88$, $p < .001$ vs. *reject*: $b = -0.02$, $SE = 0.02$, $z = -1.36$, $p = .175$). These results seem to support the idea that the non-significant behavioral findings might be due to the overall longer decision time in Study 4.

Eye-tracking

Duration: First, I investigated whether the amount of time spent on the preferred items (i.e., chosen or non-rejected) and non-preferred (i.e., non-chosen or rejected) items was different in the two decision frames. For the analysis, standardized durations were used, and the standardization was conducted by dividing the amount of time spent on preferred and non-preferred items by response time in each trail. A 2 (decision frame: choose vs. reject; between-subject) × 2 (item preference: preferred items vs. non-preferred items; within-subject) mixed ANOVA results did not show significant main effects of decision frame ($F(1, 87) = 1.45$, $p = .233$, $\eta_p^2 = 0.02$) and item preference ($F(1, 87) = 1.86$, $p = .176$, $\eta_p^2 = 0.02$). However, the interaction between decision frame and item preference was significant ($F(1, 87) = 104.14$, $p < .001$, $\eta_p^2 = 0.54$; Figure 16). Further contrasts showed that participants spent longer time on the preferred items than on the non-preferred items in the *choose* frame ($M_{\text{preferred}} = 0.51$, SD = 0.03 vs. $M_{\text{non-preferred}} = 0.44$, SD = 0.05; $F = 67.54$, $p < .001$), while they spent longer time on the non-preferred items than on the preferred items in the *reject* frame ($M_{\text{preferred}} = 0.44$, SD = 0.04 vs. $M_{\text{non-preferred}} = 0.49$, SD = 0.04; $F = 38.57$, $p < .001$). A further analysis tested whether the gaze bias (i.e., longer duration on the preferred items than on the non-preferred items in the *choose* frame and the reverse in the *reject* frame) in the *choose* frame was different from that in the *reject* frame. As predicted by the accentuation model, the gaze bias was greater in the choose frame ($M = 0.07$) than in the reject frame ($M = 0.05$), and the difference was statistically significant ($b = -0.12$, $SE = 0.01$, $t = -10.20$, $p < .001$).

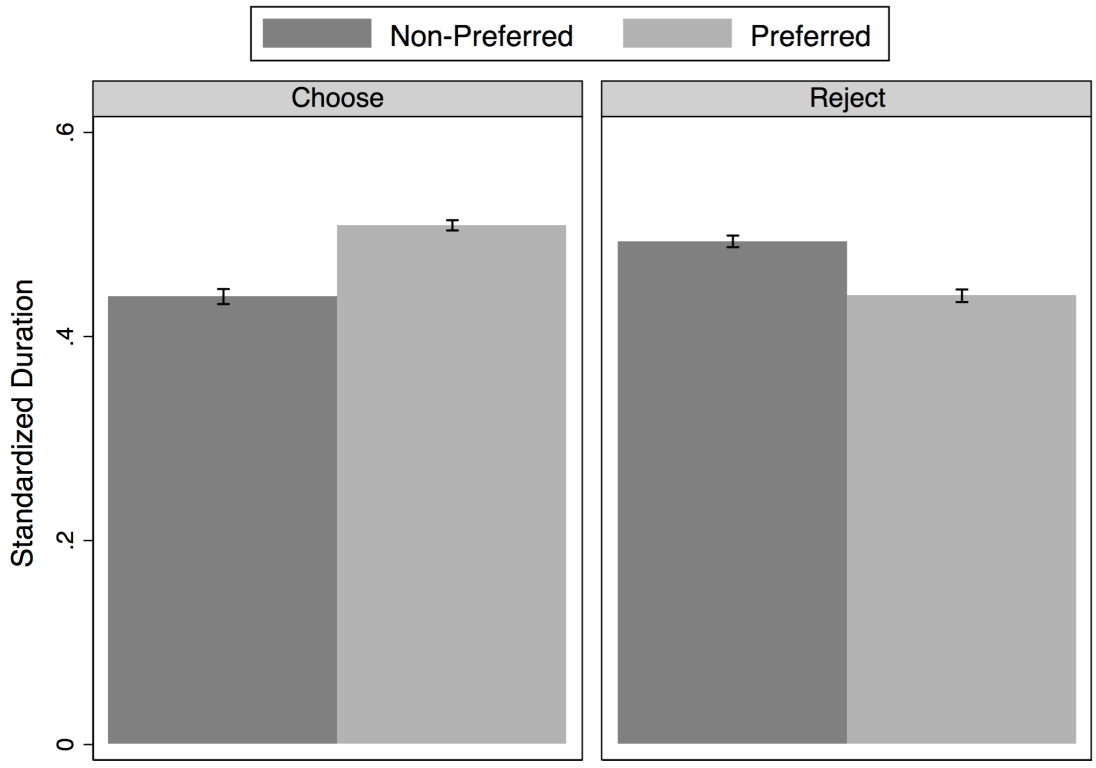

Figure 16. Mean standardized duration on preferred and non-preferred items in Study 4. the error bars indicate standard errors.

Next, I investigated whether the attention towards the preferred items varied consistently across the entire duration of the decision epoch. For the analysis, the eye-tracking data in each trial were binned into three bins, and the standardized durations were cumulated across bins. I created these bins by dividing each decision epoch into three equal segments. A 2 (decision frame: choose vs. reject; between-subject) × 2 (item preference: preferred vs. non-preferred; within-subject) × 3 (bin: first vs. second vs. third; within-subject) mixed ANOVA results showed a significant three-way interaction ($F(2, 174) = 54.57$, $p < .001$, $\eta_p^2 = 0.39$). To understand the three-way interaction better, I ran a 2 (item preference: preferred vs. non-preferred; within-subject) × 3 (bin: first vs. second

vs. third; within-subject) repeated-measure ANOVA in each decision frame condition. Both the *choose* frame condition and the *reject* frame conditions showed significant interaction between item preference and bin (*choose* frame: $F(2, 86) = 16.84$, $p < .001$, $\eta_p^2 = 0.28$ vs. *reject* frame: $F(2, 88) = 38.83$, $p < .001$, $\eta_p^2 = 0.47$), but the patterns diverged. Participants did not show difference in the amount of time spent on both preferred and non-preferred items in the first bin (*choose* frame: $F = 1.94$, $p = .167$ vs. *reject* frame: $F = 1.83$, $p = .179$) and in the second bin (*choose* frame: $F = 1.66$, $p = .201$ vs. *reject* frame: $F = 0.40$, $p = .529$), but the pattern diverged at the last bin (*choose* frame: $F = 71.83$, $p < .001$ vs. *reject* frame: $F = 41.20$, $p < .001$). The *choose* frame condition showed increased attention to the preferred items than the non-preferred items, while the *reject* frame condition showed increased attention to non-preferred items than the preferred items for the last bin (Figure 17).

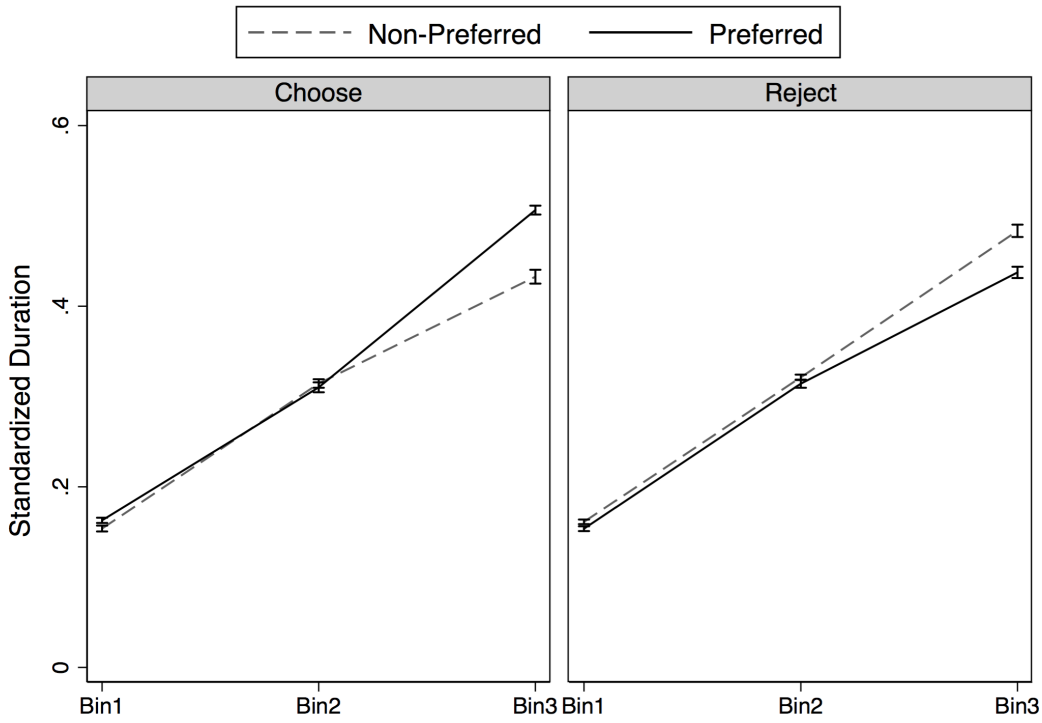

Figure 17. Mean standardized duration focused on preferred and non-preferred items across three time bins in Study 4. Error bars indicate standard errors.

Prediction of Choice in Phase 2: Finally, I investigated whether the amount of time spent on the preferred items in Phase 1 influenced the probability of choosing the preferred items in Phase 2. For this analysis, I regressed participants' choices in Phase 2 (0: not selecting preferred items vs. 1: selecting preferred items) on the amount of time spent on the preferred items, interacted with the two decision frames using a multilevel logistic regression model. To account for dependency within a participant, random intercepts for each participant were introduced.

The results did not show significant main effects of decision frame (b = -0.04, SE = 0.10, z = -0.35, p = .727) and the amount of time spent on the preferred items (b = 0.44, SE = 0.37, z = 1.19, p = .233). However, the decision frame × the amount of time on the

preferred items was significant (b = -1.33, SE = 0.52, z = -2.54, p = .011; Figure 18). A separate multilevel logistic regression in each condition results showed that longer attention on the preferred items did not influence subsequent choices in the *choose* frame, (b = 0.47, SE = 0.37, z = 1.27, p = .205), while longer duration on the preferred items decreased subsequent preferences in the *reject* frame (marginal effect = -0.89, SE = 0.37, z = -2.43, p = .015). The negative effect of duration time in the *reject* frame seems to support the accentuation model in that short duration induces a higher preference for the preferred items in subsequent choices. However, the non-significant effect in the *choose* frame was somewhat inconsistent with the prediction from the accentuation model.

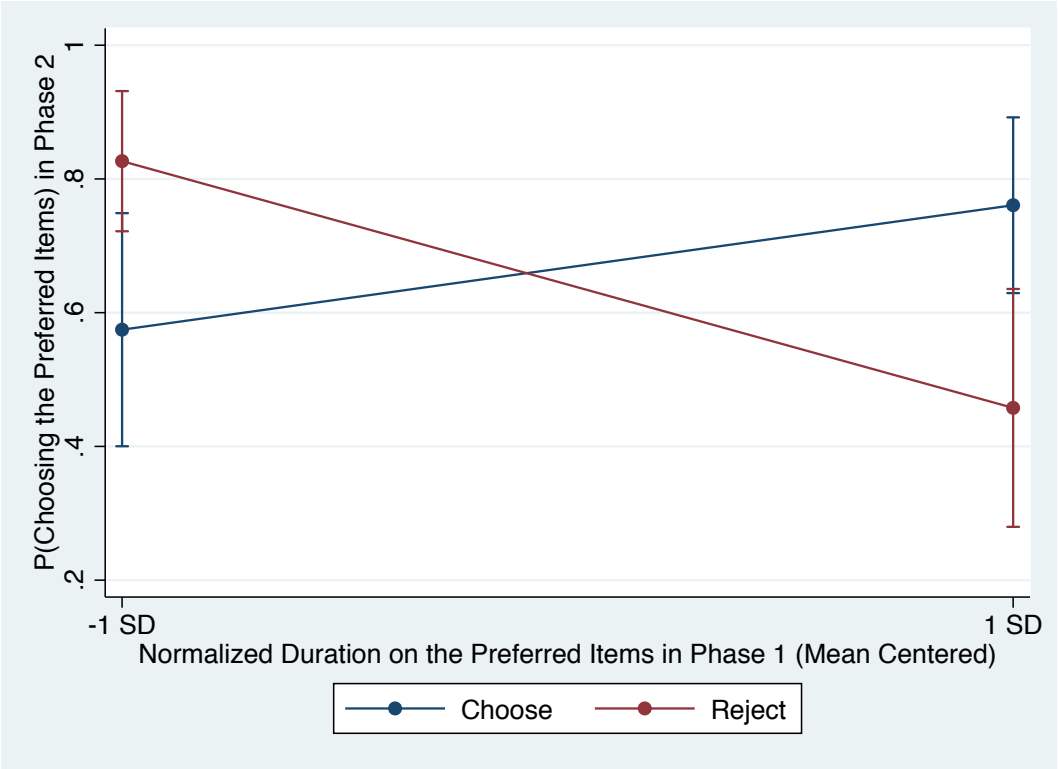

Figure 18. Predicted probability of choosing the preferred items (chosen or not-rejected items) in Phase 2. Durations were mean centered within each condition, and the x-axis indicates above and below 1SD from the mean. Error bars indicate 95% confidence interval of predicted marginal effects.

Discussion

Study 4 investigated the effect of *choose* vs. *reject* decision frames on subsequent preferences using an eye-tracking method. Even though the experimental protocol was identical with that in Study 1, I could not replicate some of the findings from Study 1. One possible explanation is that participants were more conscious in Study 4 than in Study 1 since they were being monitored by the experimenter in the same experiment room. This was partly supported by the generally longer response time in Study 4 than in Study 1. The monitoring was inevitable for ensuring high validity of the eye-tracking signal, but future research may be needed to test the findings in a more natural setting.

The gaze duration analysis results showed that participants spent longer time on the preferred items than on the non-preferred items in the *choose* frame, while the pattern in the *reject* frame was a reverse of that in the *choose* frame. Similar to the gaze cascade model, there was no difference in gaze allocation between preferred and non-preferred items in both decision frame conditions in early decision stages, while significant gaze bias emerged in the later stage of the decision stages (Mitsuda & Glaholt, 2014; Shimojo et al., 2003). However, the overall gaze bias was greater in the *choose* frame than in the *reject* frame. The smaller gaze bias in the *reject* frame than in the *choose* frame is consistent with the predictions of the accentuation model. The accentuation model predicted that the deliberate processes in the *reject* frame would eliminate or diminish the gaze bias. As predicted, the difference in the amount of time spent on the preferred and non-preferred items were smaller in the *reject* frame than in the *choose* frame.

I also tested whether the amount of time spent on the preferred items asymmetrically influenced subsequent preferential choice decisions. The results showed

that longer duration on the preferred items in Phase 1 in the *reject* frame decreased subsequent preferences. This might be because a shorter duration on the preferred items indicates less deliberate processes, resulting in increased preference for the preferred items in the subsequent preference choices. The effect of gaze duration on the preferred items on subsequent choices in the *choose* frame, on the other hand, was positive but was not statistically significant. One possibility is that the weaker behavioral findings in the current study. Another possible reason might be that the longer response time in this eye-tracking study could influence the gaze duration. In a follow-up analysis, I analyzed the effect of response time on the amount of time spent on the preferred items interacted with the two decision frames in Phase 1 using a multilevel linear regression. The results showed that longer response time decreased the amount of time spent on the preferred items (b = -0.005, SE = 0.001, z = -6.64, $p < .001$) and the effect was interacted with the decision frames (b = 0.005, SE = 0.001, z = 4.56, $p < .001$). In a further analysis, I ran a multilevel logistic regression in both decision frame conditions separately and found that the effect of response time on the gaze duration in the *choose* frame was negative (b = -0.04, SE = 0.01, z = -3.46, p = .001), while there was no effect of response time on the gaze duration in the *reject* frame (b = -0.02, SE = 0.01, z = -1.62, p = .106). The overall longer response time in Study 4 affected the gaze duration on the preferred items in the *choose* frame but not in the *reject* frame. A future study may be needed to test whether the eye-tracking results can be replicated in a more natural setting where participants are less conscious about the task.

GENERAL DISCUSSION

Consumers often make the same choice decisions repeatedly and employ different decision strategies even in the same decision contexts (Payne et al., 1988; Venkatraman et al., 2014). People can make choice decisions by simply *choosing* what they like among the alternatives in a choice set or by *rejecting* unattractive alternatives until they have the final one (Shafir, 1993; Tversky, 1972a). This essay investigated how choosing an attractive alternative people like and rejecting unattractive alternatives influence consumer preferences for the chosen items and non-chosen items when those alternatives are available in subsequent choice contexts using behavioral, eye-tracking and computational model methods.

Study 1 and Study 2 investigated the effect of decision framing on subsequent preferential choice decisions using a 2-part binary choice task. In Study 1, participants showed an increased preference for the chosen items when they were paired with new items in subsequent choices, while they showed the chance level preference for the non-rejected items. In Study 2 where the preferred items and non-preferred items were directly paired together, participants showed higher preference for the items from the *choose* frame than those from the *reject* frame in the preferred items pairs, but they showed indifference preference for the items from the two decision frames in the non-preferred items pairs. The findings indicate that the process of choosing increases preference for the chosen items, while the process of rejecting does not influence preference for the chosen and non-chosen items in subsequent preferential choice decisions.

The behavioral results seem to be supported the best by the accentuation model among the three previous theories of the *choose* vs. *reject* preference reversal. The accentuation model posits that people are more deliberate in the *reject* frame than in the *choose* frame since rejecting is a more difficult task than choosing (J. Chen & Proctor, 2017; Laran & Wilcox, 2011; Sokolova & Krishna, 2016; Wedell, 1997). Thus, it was expected that the effect of decision framing on subsequent preferences would be greater in the *choose* frame than in the *reject* frame. This claim was also supported by the results of Study 3 and Study 4. In Study 3, participants showed a greater difference in attractiveness rating changes in the *choose* frame than in the *reject* frame, indicating a greater impact of decision framing on subsequent preferential choices in the *choose* frame than in the *reject* frame. In Study 4, participants also allocated more attention to the selected items (chosen or rejected items) than to the non-selected items (non-chosen or non-rejected items), showing the gaze bias in both decision frame conditions (Mitsuda & Glaholt, 2014; Shimojo et al., 2003). However, the gaze bias was smaller in the *reject* frame than in the *choose* frame. The smaller gaze bias may indicate that participants in the *reject* frame processed the two items more comparatively, while those in the *choose* frame processed the items more selectively. A further analysis also showed that participants in the *reject* frame switched the gaze fixations across items more than those in the *choose* frame (M_{choose} = 3.21, SD = 1.20 vs. M_{reject} = 3.72, SD = 1.03; $t(87) = -2.12$, $p = .037$, $d = .45$).

Even though the gaze allocation pattern was consistent with the prediction from the accentuation model, the link between the gaze duration in Phase 1 and the choice decisions in Phase 2 was not strong. The longer gaze duration on the preferred items

decreased the probability of choosing the preferred items in the subsequent preferential choices in the *reject* frame, while the gaze duration did not influence subsequent choices in the *choose* frame. This might be because increased response time in the eye-tracking session compared to the behavioral study. Further analysis also showed that the longer response time decreased the gaze duration on the preferred items only in the *choose* frame condition but not in the *reject* frame condition. Future research may be required to test the effect of gaze duration in Phase 1 on subsequent preferential choice decisions by using a more natural setting or by manipulating given decision time.

The current study also has several limitations. The current study did not differentiate the effect of different types of market goods. Previous studies have shown that people process different aspects of alternatives in the *choose* and *reject* decision frames (Dhar & Wertenbroch, 2000; Laran & Wilcox, 2011; Nagpal et al., 2015). For example, Dhar and Wertenbroch (2000) showed that people seek more hedonic aspects of items in the *reject* frame than in the *choose* frame. Thus, it is possible that people employ different strategies in purchasing utilitarian products and hedonic products. One might seek maximum utility in purchasing hedonic goods, while one might seek a just decent level item in purchasing utilitarian goods. Laran and Wilcox (2011) also showed that people explore preference inconsistent alternatives more in the *reject* frame than in the *choose* frame. Future research may be needed to investigate how the decision frame, goals, and the types of items interact and affect subsequent preferences. Also, the current study focused on a binary choice task only. However, in many real consumer choice settings, there are many more available alternatives in the choice contexts. It will not be easy to clearly categorize decision strategies into either *choose* or *reject* decision

strategies in multi-alternative tasks. However, in future research, it may be possible to test the effect of *choose* vs. *reject* framing on subsequent preferences in a multi-alternative task context by examining the effect of decision strategies similar to the *choose* frame (e.g., *take-the-best*) or the *reject* frame (e.g., *elimination-by-aspect*) on subsequent preferences.

CHAPTER 6.

CONCLUSIONS

My dissertation aimed to understand the effect of procedural preference elicitation methods on consumer preferences. The three essays investigated whether having consumers think about a random price for their willingness-to-pay (WTP) influences their preferential judgments, how learning risk levels and payoffs from a description and from repeated experience affects consumer risk preference, and the effect of decision frames on subsequent preferential choice decisions. Whereas numerous studies have shown robust findings that description of task environments (e.g., different framing of outcomes or features of alternatives) influence consumer preferences (e.g., Ganzach & Karsahi, 1995; Gendall et al., 2006; Gourville, 1998; Khan & Dhar, 2010), the effect of procedural changes on consumer preference has been studied to a relatively lesser extent, and recently, several procedural manipulation studies have been shown weak or null effects (Klein et al., 2015; Lee & Schwarz, 2010; Pashler et al., 2012; Pashler et al., 2013; Shanks et al., 2015). To address this issue, my dissertation investigated the effect of procedural changes on consumer preference in three different domains using a multi-method approach. I believe the findings of this dissertation provide important theoretical and practical implications on consumer preference and judgment and decision-making behavior.

One of the theoretical contributions of the current dissertation is that the overall findings of this dissertation show the robustness of procedural manipulations on consumer preference. Recent replication attempts of the anchoring effect on WTP have

shown inconsistent results (Bergman et al., 2010; Maniadis et al., 2014). CHAPTER 2 attempted to reconcile the inconsistent replication results by testing different manipulations used in previous replication studies and by conducting a meta-analysis. The results show that a price anchor can sway consumers' WTP, even though the price anchor is explicitly random to the WTP decisions. Indeed, the anchoring effect on WTP is robust to various experimental procedures. CHAPTER 4 shows that people rely more on overall probability of winning in the description format, while they rely on both expected value and the overall probability of winning in a risky decision task when only two task types (e.g., Pwin available and unavailable trials) are employed or various task types are employed. CHAPTER 5 shows that the *choose* frame increases subsequent preferences, while the *reject* frame does not both in a between-subject design and in a within-subject design. The overall behavioral findings in this dissertation indicate that procedural manipulations significantly influence consumer preference, at least in these three domains. The findings were also supported by a multi-method approach that includes behavioral, eye-tracking, and computational modeling methods.

The findings of current dissertation also provide implications for human preference. One of the theoretical claims on preference reversals is that the inconsistent preference resulting from different descriptions or procedures indicates short-term distortions of consumer preference (i.e., response bias) or changes in underlying preference (Bettman, Luce, & Payne, 2008; Simonson, 2008). Using a multi-session study, I tested the long-term anchoring effect on WTP in CHAPTER 3 and the effect of *choose* vs. *reject* decision framing on subsequent preferential choices in CHAPTER 5. The long-term anchoring effect results show that anchoring not only influences WTP

when WTP is elicited immediately after anchoring procedures, but also influences WTP even after 2-5 months when the anchor is first provided. The decision framing results also show that decision strategies utilized in the previous decision stage still affects subsequent preference decisions. The persistent effect of procedural manipulations on consumer preference implies that preference constructed by procedural manipulations is not a simple response bias.

The findings of current dissertation also provide practical implications. Marketers and policy makers want to influence consumer preference and trigger specific behaviors. It has been widely shown that different descriptions of task environment could influence consumer preference, and three essays in this dissertation show that designing decision procedures can significantly influence consumer preference in the domain of judgments, choices, and risky decisions. Thus, designing decision processes can have powerful impact on consumer behavior and is one factor that marketers and policy makers should concern when they are setting marketing strategies. Studies in nudging and choice architecture show the importance of designing decision processes. A classic example of a nudge in cafeteria shows that placing healthy food first can increase consumption of healthy food (Thaler & Sunstein, 1999). Smart default literature also shows that having people think about the best alternative help them to make better decisions (Goldstein, Johnson, Herrmann, & Heitmann, 2008; Johnson, Hassin, Baker, Bajger, & Treuer, 2013). Marketers and policy makers not only need to consider how to design description of messages but also need to consider how to design decision processes that consumers may take.

The findings also imply that procedural manipulations not only influence consumer preference at the moment the manipulations are provided but also can have long-term effects even when the manipulations do not continue. Marketers and policy makers want to develop marketing strategies which can have long-term effects on consumer preference. The results of this dissertation show that procedure manipulations could influence consumer preference even after several months and in subsequent preferential decisions. The findings imply that procedural manipulations do not simply distort consumer preference, but they influence underlying preferences. This informs marketers and policy makers that designing decision procedures can be an effective strategy to form long-term customer preference for brands or items. Several previous studies also have shown supporting evidence. For example, Schonberg et al. (2014) showed that associating items with motor responses can impact people's preference for the associated items after one month. Sharot et al. (2012) also reported that people showed increased preference for the chosen vacation spots even 2.5 years since they made choice decisions. Future research may be needed to develop marketing strategies to induce long-term preference for brands, products, or services.

The current dissertation has several limitations. First, there are many more procedure invariance violation cases, but the current dissertation only focuses on the anchoring effect on WTP, risk preference in the description-based and experience-based formats, and the *choose* vs. *reject* decision framing. Future research may be needed to test whether other types of procedure manipulations (e.g., Query theory, behavioral priming, goal priming, etc.) also can have robust and persistent effects on consumer preference. The second limitation is that the tasks in this dissertation are still very

controlled tasks, even though I tried to create them with high ecological validity. For example, CHAPTER 2 and CHAPTER 3 used a BDM incentive compatible method to help people think the task was real, the gambling task in CHAPTER 4 was more complicated than the tasks used in the previous description vs. experience gap studies, and CHAPTER 5 used a multi-phase decision paradigm, which resembles more real-life decision contexts than a single shot binary choice task. However, those tasks are still very controlled and lack a cover story that helps participants to perceive the tasks are relevant to them. Future research can examine the findings with more realistic contexts (e.g., real auction settings for CHAPTER 2 and CHAPTER 3, insurance or warranty purchase decisions for CHAPTER 4, or incentive compatible tasks for CHAPTER 5) or with field experiments.

REFERENCES

Acquisti, A., & Gross, R. (2009). Predicting Social Security numbers from public data. *Proceedings of the National academy of sciences, 106*(27), 10975-10980.

Alevy, J. E., Landry, C. E., & List, J. A. (2015). Field Experiments on the Anchoring of Economic Valuations. *Economic Inquiry, 53*(3), 1522-1538.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics, 118*(1), 73-105. doi:Doi 10.1162/00335530360535153

Ariely, D., & Norton, M. I. (2008). How actions create–not just reveal–preferences. *Trends Cogn Sci, 12*(1), 13-16.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science, 9*(3), 226-232.

Bergman, O., Ellingsen, T., Johannesson, M., & Svensson, C. (2010). Anchoring and cognitive ability. *Economics Letters, 107*(1), 66-68.

Bettman, J. R., Luce, M. F., & Payne, J. W. (2008). Preference construction and preference stability: Putting the pillow to rest. *Journal of Consumer Psychology, 18*(3), 170-174.

Blankenship, K. L., Wegener, D. T., Petty, R. E., Detweiler-Bedell, B., & Macy, C. L. (2008). Elaboration and consequences of anchored estimates: An attitudinal perspective on numerical anchoring. *Journal of Experimental Social Psychology, 44*(6), 1465-1476.

Brewer, N. T., & Chapman, G. B. (2002). The fragile basic anchoring effect. *Journal of Behavioral Decision Making, 15*(1), 65-77.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., . . . Chan, T. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6280), 1433-1436.

Camilleri, A. R., & Newell, B. R. (2009). Within-subject preference reversals in description-and experience-based choice. *Cog Sci Society, 31*, 449-454.

Camilleri, A. R., & Newell, B. R. (2011). Description- and experience-based choice: does equivalent information equal equivalent choice? *Acta Psychologica, 136*(3), 276-284. doi:10.1016/j.actpsy.2010.11.007

Camilleri, A. R., & Newell, B. R. (2013). Mind the gap? Description, experience, and the continuum of uncertainty in risky choice. *Progress in Brain Research, 202*, 55-71. doi:10.1016/B978-0-444-62604-2.00004-6

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making, 7*(4), 223-242.

Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes, 79*(2), 115-153.

Chen, J., & Proctor, R. W. (2017). Role of accentuation in the selection/rejection task framing effect. *Journal of Experimental Psychology: General, 146*(4), 543-568.

Chen, M. K., & Risen, J. L. (2010). How choice affects and reflects preferences: revisiting the free-choice paradigm. *Journal of Personality and Social Psychology, 99*(4), 573-594.

Chernev, A. (2009). Choosing versus rejecting: the impact of goal-task compatibility on decision confidence. *Social Cognition, 27*(2), 249-260.

Dhar, R., & Wertenbroch, K. (2000). Consumer choice between hedonic and utilitarian goods. *Journal of Marketing Research, 37*(1), 60-71.

Egan, L. C., Santos, L. R., & Bloom, P. (2007). The origins of cognitive dissonance: Evidence from children and monkeys. *Psychological Science, 18*(11), 978-983.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science, 12*(5), 391-396.

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making, 18*(3), 199-212.

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

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., . . . Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making, 23*(1), 15-47.

Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making, 21*(5), 575-597.

Evans, J. S. B. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13*(4), 321-339.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191.

Fenske, M. J., & Raymond, J. E. (2006). Affective influences of selective attention. *Current Directions in Psychological Science, 15*(6), 312-316.

Foo, J. C., Haji, T., & Sakai, K. (2014). Prefrontal mechanisms in preference and non-preference-based judgments. *Neuroimage, 95*, 151-161.

Fox, C. R., & Hadar, L. (2006). Decisions from experience= sampling error+ prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making, 1*(2), 159-161.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25-42.

Frederick, S., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General, 141*(1), 124.

Frederick, S., Mochon, D., & Danilowitz, J. (2013). The role of inference in anchoring effects. *Unpublished manuscript*.

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion, 14*(4), 577-606.

Fudenberg, D., Levine, D. K., & Maniadis, Z. (2012). On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics, 4*(2), 131-145.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics, 40*(1), 35-42.

Ganzach, Y. (1995). Attribute scatter and decision outcome: Judgment versus choice. *Organizational Behavior and Human Decision Processes, 62*(1), 113-122.

Ganzach, Y., & Karsahi, N. (1995). Message framing and buying behavior: A field experiment. *Journal of Business Research, 32*(1), 11-17.

Garcia, J., Kimeldorf, D. J., & Koelling, R. A. (1955). Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science, 122*(3160), 157-158. doi:http://dx.doi.org/10.1126/science.122.3179.1089

Gendall, P., Hoek, J., Pope, T., & Young, K. (2006). Message framing effects on price discounting. *Journal of Product & Brand Management, 15*(7), 458-465.

Goldstein, D. G., Johnson, E. J., Herrmann, A., & Heitmann, M. (2008). Nudge your customers toward better choices. *Harvard Business Review, 86*(12), 99-105.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: integrating sampling and repeated decisions from experience. *Psychological Review, 118*(4), 523-551. doi:10.1037/a0024558

Gourville, J. T. (1998). Pennies-a-day: The effect of temporal reframing on transaction evaluation. *Journal of Consumer Research, 24*(4), 395-408.

Green, D., Jacowitz, K. E., Kahneman, D., & McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics, 20*(2), 85-116.

Greenberg, A. E., & Spiller, S. A. (2015). Opportunity Cost Neglect Attenuates the Effect of Choices on Preferences. *Psychological Science*, 0956797615608267.

Hadar, L., & Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making, 4*(4), 317-325.

Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PLoS One, 8*(8), e72467.

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21*(5), 493-518.

Hertwig, R. (2015). Decisions from experience. *The Wiley Blackwell handbook of judgment and decision making*, 239-267.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004a). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*(8), 534-539. doi:10.1111/j.0956-7976.2004.00715.x

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004b). Decisions from experience and the effect of rare events in risky choice. *Psychol Sci, 15*(8), 534-539. doi:10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences, 13*(12), 517-523. doi:10.1016/j.tics.2009.09.004

Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. *The probabilistic mind: Prospects for rational models of cognition*, 209-236.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: why small samples? *Cognition, 115*(2), 225-237. doi:10.1016/j.cognition.2009.12.009

Hills, T. T., & Hertwig, R. (2010). Information search in decisions from experience: do our patterns of sampling foreshadow our decisions? *Psychological Science, 21*(12), 1787-1792. doi:10.1177/0956797610387443

Hills, T. T., Noguchi, T., & Gibbert, M. (2013). Information overload or search-amplified risk? Set size and order effects on decisions from experience. *Psychonomic Bulletin & Review, 20*(5), 1023-1031. doi:10.3758/s13423-013-0422-3

Huang, Y., Kendrick, K. M., & Yu, R. (2014). Conformity to the opinions of other people lasts for no more than 3 days. *Psychological Science, 25*(7), 1388-1393.

Huber, V. L., Neale, M. A., & Northcraft, G. B. (1987). Decision bias and personnel selection strategies. *Organizational Behavior and Human Decision Processes, 40*(1), 136-147.

Izuma, K., & Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron, 78*(3), 563-573.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*(11), 1161-1166.

Janiszewski, C., Kuo, A., & Tavassoli, N. T. (2012). The influence of selective attention and inattention to products on subsequent choice. *Journal of Consumer Research, 39*(6), 1258-1274.

Johnson, E. J., Hassin, R., Baker, T., Bajger, A. T., & Treuer, G. (2013). Can consumers make affordable care affordable? The value of choice architecture. *PLoS One, 8*(12), e81521.

Johnson, E. J., & Schkade, D. A. (1989). Bias in utility assessments: Further evidence and explanations. *Management science, 35*(4), 406-424.

Jung, M. H., Perfecto, H., & Nelson, L. D. (2016). Anchoring in payment: Evaluating a judgmental heuristic in field experimental settings. *Journal of Marketing Research, 53*(3), 354-368.

Khan, U., & Dhar, R. (2010). Price-framing effects on the purchase of hedonic and utilitarian bundles. *Journal of Marketing Research, 47*(6), 1090-1099.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., . . . Brumbaugh, C. C. (2015). Investigating variation in replicability. *Social Psychology, 45*, 142-152.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2*(1), 61-76.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience, 13*(10), 1292-1298.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National academy of sciences, 108*(33), 13852-13857.

Laran, J., & Wilcox, K. (2011). Choice, rejection, and elaboration on preference-inconsistent alternatives. *Journal of Consumer Research, 38*(2), 229-241.

LeBoeuf, R. A., & Shafir, E. (2006). The long and short of it: Physical anchoring effects. *Journal of Behavioral Decision Making, 19*(4), 393-406.

LeBoeuf, R. A., & Shafir, E. (2009). Anchoring on the" here" and" now" in time and distance judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 81.

Lee, S. W., & Schwarz, N. (2010). Washing away postdecisional dissonance. *Science, 328*(5979), 709-709.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15*(3), 374-378.

Levin, I. P., Jasper, J., & Forbes, W. S. (1998). Choosing versus rejecting options at different stages of decision making. *Journal of Behavioral Decision Making, 11*(3), 193-210.

Li, S. Y., Rakow, T., & Newell, B. R. (2009). Personal experience in doctor and patient decision making: from psychology to medicine. *Journal of Evaluation in Clinical Practice, 15*(6), 993-995.

Lichtenstein, S., & Slovic, P. (1971). Reversals of Preference between Bids and Choices in Gambling Decisions. *Journal of Experimental Psychology, 89*(1), 46-55. doi:DOI 10.1037/h0031207

Lichtenstein, S., & Slovic, P. (2006). *The construction of preference*: Cambridge University Press.

Lopes, L. L. (1995). Algebra and Process in the Modeling of Risky Choice. In J. R. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Decision Making from a Cognitive Perspective* (pp. 177-220). San Diego: Academic Press.

Lopes, L. L., & Oden, G. C. (1999). The Role of Aspiration Level in Risky Choice: A Comparison of Cumulative Prospect Theory and SP/A Theory. *Journal of Mathematical Psychology, 43*(2), 286-313.

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme outcomes sway risky decisions from experience. *Journal of Behavioral Decision Making, 27*(2), 146-156.

Ludvig, E. A., & Spetch, M. L. (2011). Of black swans and tossed coins: is the description-experience gap in risky choice limited to rare events? *PLoS One, 6*(6).

Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *The American Economic Review, 104*(1), 277-290.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *The Bell Journal of Economics*, 587-608.

March, J. G., & Shapira, Z. (1987). Managerial perspectives on risk and risk taking. *Management science, 33*(11), 1404-1418.

Meloy, M. G., & Russo, J. E. (2004). Binary choice under instructions to select versus reject. *Organizational Behavior and Human Decision Processes, 93*(2), 114-128.

Mitsuda, T., & Glaholt, M. G. (2014). Gaze bias during visual preference judgements: Effects of stimulus category and decision instructions. *Visual Cognition, 22*(1), 11-29.

Mussweiler, T. (2001a). The durability of anchoring effects. *European journal of social psychology, 31*(4), 431-442.

Mussweiler, T. (2001b). Sentencing Under Uncertainty: Anchoring Effects in the Courtroom1. *Journal of Applied Social Psychology, 31*(7), 1535-1551.

Mussweiler, T., & Englich, B. (2005). Subliminal anchoring: Judgmental consequences and underlying mechanisms. *Organizational Behavior and Human Decision Processes, 98*(2), 133-143.

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35*(2), 136-164.

Nagpal, A., & Krishnamurthy, P. (2008). Attribute conflict in consumer decision making: The role of task compatibility. *Journal of Consumer Research, 34*(5), 696-705.

Nagpal, A., Lei, J., & Khare, A. (2015). To Choose or to Reject: The Effect of Decision Frame on Food Customization Decisions. *Journal of Retailing, 91*(3), 422-435.

Newell, B. R., Rakow, T., Yechiam, E., & Sambur, M. (2015). Rare disaster information can increase risk-taking. *Nature Climate Change, 6*(2), 158-161.

Noguchi, T., & Hills, T. T. (2015). Experience-based decisions favor riskier alternatives in large sets. *Journal of Behavioral Decision Making*, Advanced online publication. doi:10.1002/bdm.1893

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes, 39*(1), 84-97.

Park, C. W., Jun, S. Y., & MacInnis, D. J. (2000). Choosing what I want versus rejecting what I do not want: An application of decision framing to product option choice decisions. *Journal of Marketing Research, 37*(2), 187-202.

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS One, 7*(8), e42510.

Pashler, H., Rohrer, D., & Harris, C. R. (2013). Can the goal of honesty be primed? *Journal of Experimental Social Psychology, 49*(6), 959-964.

Payne, J. W. (2005). It is whether you win or lose: The importance of the overall probabilities of winning or losing in risky choice. *Journal of Risk and Uncertainty, 30*(1), 5-19.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534-552.

Payne, J. W., Samper, A., Bettman, J. R., & Luce, M. F. (2008). Boundary conditions on unconscious thought in complex decision making. *Psychological Science, 19*(11), 1118-1123.

Perfecto, H., Galak, J., Simmons, J. P., & Nelson, L. D. (2017). Rejecting a bad option feels like choosing a good one. *Journal of Personality and Social Psychology, 113*(5), 659.

Philiastides, M. G., & Ratcliff, R. (2013). Influence of branding on preference-based decision making. *Psychological Science, 24*(7), 1208-1215.

Rakow, T., & B Rahim, S. (2010). Developmental insights into experience-based decision making. *Journal of Behavioral Decision Making, 23*(1), 69-82.

Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes, 106*(2), 168-179.

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making, 23*(1), 1-14.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends Cogn Sci, 20*(4), 260-281.

Raymond, J. E., Fenske, M. J., & Tavassoli, N. T. (2003). Selective attention determines emotional responses to novel visual stimuli. *Psychological Science, 14*(6), 537-542.

Rohrer, D., Pashler, H., & Harris, C. R. (2015). Do subtle reminders of money change people's political views? *Journal of Experimental Psychology: General, 144*(4), e73.

Schonberg, T., Bakkour, A., Hover, A. M., Mumford, J. A., Nagar, L., Perez, J., & Poldrack, R. A. (2014). Changing value through cued approach: an automatic mechanism of behavior change. *Nature neuroscience, 17*(4), 625-630.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & cognition, 21*(4), 546-556.

Shanks, D. R., Vadillo, M. A., Riedel, B., Clymo, A., Govind, S., Hickin, N., . . . Puhlmann, L. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives? *Journal of Experimental Psychology: General, 144*(6), e142.

Sharot, T., Fleming, S. M., Yu, X., Koster, R., & Dolan, R. J. (2012). Is choice-induced preference change long lasting? *Psychological Science, 23*(10), 1123-1129.

Sharot, T., Velasquez, C. M., & Dolan, R. J. (2010). Do decisions shape preference? Evidence from blind choice. *Psychological Science, 21*(9), 1231-1235.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature neuroscience, 6*(12), 1317.

Simon, D., Krawczyk, D. C., Bleicher, A., & Holyoak, K. J. (2008). The transience of constructed preferences. *Journal of Behavioral Decision Making, 21*(1), 1-14. doi:10.1002/bdm.575

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of preferences by constraint satisfaction. *Psychological Science, 15*(5), 331-336.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 99-118.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534.

Simonson, I. (2008). Regarding inherent preferences. *Journal of Consumer Psychology, 18*(3), 191-196.

Simonson, I., & Drolet, A. (2004). Anchoring effects on consumers' willingness-to-pay and willingness-to-accept. *Journal of Consumer Research, 31*(3), 681-690.

Slovic, P. (1995). The construction of preference. *American psychologist, 50*(5), 364-371.

Smith, A. R., & Windschitl, P. D. (2011). Biased calculations: Numeric anchors influence answers to math equations. *Judgment and Decision Making, 6*(1), 139-146.

Sokolova, T., & Krishna, A. (2016). Take It or Leave It: How Choosing versus Rejecting Alternatives Affects Information Processing. *Journal of Consumer Research, 43*(4), 614-635.

Teodorescu, K., & Erev, I. (2014). On the decision to explore new alternatives: The coexistence of under-and over-exploration. *Journal of Behavioral Decision Making, 27*(2), 109-123.

Thaler, R. H., & Sunstein, C. R. (1999). *Nudge: Improving decisions about health, wealth, and happiness*: HeinOnline.

Thorsteinson, T. J. (2011). Initiating Salary Discussions With an Extreme Request: Anchoring Effects on Initial Salary Offers1. *Journal of Applied Social Psychology, 41*(7), 1774-1792.

Thorsteinson, T. J., Breier, J., Atwell, A., Hamilton, C., & Privette, M. (2008). Anchoring effects on performance judgments. *Organizational Behavior and Human Decision Processes, 107*(1), 29-40.

Tversky, A. (1972a). Elimination by aspects: A theory of choice. *Psychol Rev, 79*(4), 281.

Tversky, A. (1972b). Elimination by aspects: A theory of choice. *Psychological Review, 79*(4), 281-299.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychol Rev, 95*(3), 371.

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods, 40*(1), 61-72.

Venkatraman, V., Payne, J. W., Bettman, J. R., Luce, M. F., & Huettel, S. A. (2009). Separate neural mechanisms underlie choices and strategic preferences in risky decision making. *Neuron, 62*(4), 593-602. doi:10.1016/j.neuron.2009.04.007

Venkatraman, V., Payne, J. W., & Huettel, S. A. (2014). An overall probability of winning heuristic for complex risky decisions: Choice and eye fixation evidence. *Organizational Behavior and Human Decision Processes, 125*(2), 73-87.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J Stat Softw, 36*(3), 1-48.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology, 60*(6), 385.

Wansink, B., Kent, R. J., & Hoch, S. J. (1998). An anchoring and adjustment model of purchase quantity decisions. *Journal of Marketing Research*, 71-81.

Weber, E. U. (2006). Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic Change, 77*(1-2), 103-120.

Wedell, D. H. (1997). Another look at reasons for choosing and rejecting. *Memory & cognition, 25*(6), 873-887.

Wegener, D. T., Petty, R. E., Blankenship, K. L., & Detweiler-Bedell, B. (2010). Elaboration and numerical anchoring: Implications of attitude theories for consumer judgment and decision making. *Journal of Consumer Psychology, 20*(1), 5-16.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in neuroinformatics, 7*.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology: General, 125*(4), 387.

Wulff, D. U., Hills, T. T., & Hertwig, R. (2015). Online product reviews and the description–experience gap. *Journal of Behavioral Decision Making, 28*(3), 214-223.

Yechiam, E., Barron, G., & Erev, I. (2005). The role of personal experience in contributing to different patterns of response to rare terrorist attacks. *Journal of Conflict Resolution, 49*(3), 430-439.

Yechiam, E., Rakow, T., & Newell, B. R. (2015). Super-underweighting of rare events with repeated descriptive summaries. *Journal of Behavioral Decision Making, 28*(1), 67-75.

APPENDIX A

ESSAY 1A: SUMMARY STATISTICS BY THE COGNITIVE REFLECTION TASK AND NEED FOR COGNITION SCORES. LOW AND HIGH GROUPS WERE MEDIAN SPLITTED.

		Low Anchor		Mid Anchor		High Anchor			
	CRT	Mean	SD	Mean	SD	Mean	SD	N	Anchor x CRT Interaction
Art Book	Low	$9.38	5.33	$13.62	9.04	$17.11	10.90	98	$F(2, 186) = 0.64$
	High	$10.24	7.61	$12.11	7.23)	$19.07	9.89	94	$p = .530$ $\eta_p^2 = 0.007$
Humidifier	Low	$29.52	13.88	$35.68	19.73	$35.91	15.64	100	$F(2, 187) = 0.67$
	High	$28.25	15.79	$28.45	9.93	$32.86	11.33	93	$p = .515$ $\eta_p^2 = 0.007$
Bluetooth Keyboard	Low	$27.18	10.39	$31.83	9.62	$39.37	14.79	100	$F(2, 188) = 3.25$
	High	$34.33	14.31	$30.24	14.59	$35.00	15.58	94	$p = .041$ $\eta_p^2 = 0.033$
Cordless Mouse	Low	$16.21	8.03	$19.23	7.54	$25.52	11.30	100	$F(2, 188) = 0.54$
	High	$15.96	6.02	$19.14	7.15	$22.56	10.75	94	$p = .582$ $\eta_p^2 = 0.006$
Electric Toothbrush	Low	$23.57	17.64	$29.05	18.04	$40.81	28.76	100	$F(2, 188) = 0.54$
	High	$22.39	13.84	$32.60	19.11	$36.81	25.24	94	$p = .582$ $\eta_p^2 = 0.006$
Vacuum Cleaner	Low	$32.62	18.36	$39.88	11.76	$51.00	21.31	99	$F(2, 187) = 0.20$
	High	$35.56	18.49	$39.11	16.05	$50.74	17.61	94	$p = .819$ $\eta_p^2 = 0.002$

		Low Anchor		Mid Anchor		High Anchor			
	NFC	Mean	SD	Mean	SD	Mean	SD	N	Anchor x NFC Interaction
Art Book	Low	$10.38	6.34	$11.20	6.32	$16.91	10.62	101	$F(2, 186) = 1.24$
	High	$9.03	6.97	$14.26	9.29	$19.23	10.21	91	$p = .291$ $\eta_p^2 = 0.013$
Humidifier	Low	$31.41	16.20	$32.35	17.59	$33.39	15.43	102	$F(2, 187) = 1.29$
	High	$26.15	13.45	$32.76	15.81	$37.10	11.20	91	$p = .277$ $\eta_p^2 = 0.014$
Bluetooth Keyboard	Low	$30.61	14.57	$30.80	9.86	$36.37	14.36	102	$F(2, 188) = 0.11$
	High	$30.56	10.31	$31.21	13.98	$38.44	16.43	92	$p = .894$ $\eta_p^2 = 0.001$
Cordless Mouse	Low	$16.29	8.13	$19.66	7.50	$22.93	9.86	102	$F(2, 188) = 0.56$
	High	$15.90	6.18	$18.54	7.05	$24.94	12.02	92	$p = .573$ $\eta_p^2 = 0.006$
Electric Toothbrush	Low	$22.39	14.55	$27.95	16.73	$39.97	29.50	102	$F(2, 188) = 0.75$
	High	$23.50	16.85	$34.70	20.52	$37.79	24.84	92	$p = .473$ $\eta_p^2 = 0.008$
Vacuum Cleaner	Low	$36.57	21.43	$39.42	11.77	$52.43	21.04	101	$F(2, 187) = 0.54$
	High	$30.17	11.90	$39.53	16.53	$49.58	18.34	92	$p = .583$ $\eta_p^2 = 0.006$

APPENDIX B

ESSAY 1A: REGRESSION RESULTS FOR STUDY 1 EXCLUDING NON-VALID RESPONSES

	(1) No Item FE	(2) Item FE	(3) Anchor Type	(4) Recruit Type	(5) Both
Anchor	0.188**	0.194**	0.195**	0.196***	0.197**
	(0.03)	(0.03)	(0.04)	(0.03)	(0.06)
Anchor Type			0.872		0.895
			(1.57)		(1.58)
Anchor×Anchor Type			0.004		0.006
			(0.03)		(0.03)
Recruit Type				-0.238	-0.387
				(1.53)	(1.58)
Anchor×Recruit Type				0.044†	0.046†
				(0.03)	(0.03)
Anchor Type ×Recruit Type					0.455
					(1.58)
Anchor×Anchor Type ×Recruit Type					0.002
					(0.03)
Intercept	19.685**	Item FE	Item FE	Item FE	Item FE
	(1.78)				
N	655	655	655	655	655

Standard errors in parentheses.
*† p < .10, * p < .05, ** p < .01, *** p < .001*
Note: The inconsistent responses were defined if (1) they were willing to pay for a given random price but their WTP was lower than the random price, or (2) they were not willing to pay for a given random price but their WTP was higher than the random price. Models (2) through (5) include fixed effects for item. Anchor type is coded as -1 for SSN and 1 for RN. Recruit Type is coded as -1 for classroom and 1 for email recruiting.

APPENDIX C

ESSAY 2: MEAN OVERALL PROBABILITIES OF WINNING AND EXPECTED VALUES OF EACH ALTERNATIVE ACROSS THE PROBLEMS USED IN STUDY 1, BASED ON PREDEFINED PROBABILITY DISTRIBUTION (SET COLUMNS) AND THE EXPERIENCED SAMPLES (EXPERIENCED COLUMNS).

			Set			Experienced		
Trial	Pwin		Gmax	Intermediate	Lmin	Gmax	Intermediate	Lmin
3	Available	OP	0.33	0.66	0.33	0.36	0.64	0.39
		EV	-1.65	-1.65	-1.65	2.90	-6.00	4.31
4	Available	OP	0.33	0.66	0.33	0.33	0.68	0.28
		EV	0.00	0.00	0.00	-0.24	0.48	-5.20
5	Available	OP	0.33	0.66	0.33	0.34	0.72	0.27
		EV	0.00	0.00	0.00	2.45	8.29	-9.16
6	Available	OP	0.33	0.66	0.33	0.34	0.67	0.34
		EV	-6.60	-6.60	-6.60	-7.02	-5.57	-6.40
7	Unavailable	OP	0.33	0.33	0.33	0.35	0.36	0.38
		EV	-9.90	-9.90	-9.90	-8.12	-5.73	-4.80
8	Unavailable	OP	0.33	0.33	0.33	0.72	0.71	0.65
		EV	-1.65	-1.65	-1.65	3.46	1.81	-3.27
9	Unavailable	OP	0.33	0.33	0.33	0.31	0.30	0.36
		EV	0.00	0.00	0.00	-3.82	-4.82	3.05
10	Unavailable	OP	0.33	0.33	0.33	0.63	0.65	0.67
		EV	0.00	0.00	0.00	-2.01	-2.11	0.02

OP: overall probability of winning, EV: expected value

APPENDIX D

ESSAY 2: STIMULI AND AN EXAMPLE OF ORIGINAL FORMAT AND MATCHED FORMAT USED IN THE FOLLOW-UP STUDY OF STUDY 1.

	Base Gamble				
Trial	O1	O2	O3	Constant	EV
1	55	-15	-65	20	-8.25
2	40	-10	-65	15	-11.55
3	70	-5	-85	20	-6.6
4	65	-10	-75	20	-6.6
5	60	-20	-60	25	-6.6
6	50	-5	-40	20	1.65
7	65	-20	-80	15	-11.55
8	50	10	-75	10	-4.95
9	80	-25	-70	15	-4.95
10	60	5	-70	5	-1.65
11	40	5	-40	15	1.65
12	65	-20	-50	10	-1.65

APPENDIX E

ESSAY 2: AN EXAMPLE (TRIAL 4) OF PRESENTATION FORMANTS FOR THE ORIGINAL CONDITION AND MATCHED CONDITION FOR ONE REPRESENTATIVE SUBJECT.

G1	90 (0.33)	-5 (0.33)	-85 (0.33)
G2	70 (0.33)	15 (0.33)	-85 (0.33)
G3	70 (0.33)	-5 (0.33)	-65 (0.33)

a. Original Format

G1	90 (0.30)	-5 (0.35)	-85 (0.35)
G2	70 (0.41)	15 (0.12)	-85 (0.47)
G3	70 (0.21)	-5 (0.43)	-65 (0.36)

b. Matched Format

Note: Probabilities for each outcome can now vary across alternatives, and hence are presented in the parentheses. The probabilities in the matched format condition are based on the participant's sampling history for the same trial in their experience format condition.