RISK AVERSION AND INFORMATION ACQUISITION ACROSS REAL AND HYPOTHETICAL SETTINGS

by

MATTHEW PAUL TAYLOR

A DISSERTATION

Presented to the Department of Economics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2012

DISSERTATION APPROVAL PAGE

Student: Matthew Paul Taylor

Title: Risk Aversion and Information Acquisition Across Real and Hypothetical Settings

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

Dr. Trudy Ann Cameron	Co-chair
Dr. William T. Harbaugh	Co-chair
Dr. Van Kolpin	Member
Dr. Ulrich Mayr	Outside Member

and

Kimberly Andrews Espy	Vice President for Research & Innovation/ Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2012

DISSERTATION ABSTRACT

Matthew Paul Taylor

Doctor of Philosophy

Department of Economics

June 2012

Title: Risk Aversion and Information Acquisition Across Real and Hypothetical Settings

I collect data on subjects' information acquisition during real and hypothetical risky choices using process-tracing software called Mouselab. I also measure subjects' cognitive ability using the cognitive reflective test (CRT). On average, measured risk preferences are not significantly different across real and hypothetical settings. However, cognitive ability is inversely related to risk aversion when choices are hypothetical, but it is unrelated when the choices are real. This interaction between cognitive ability and hypothetical setting is consistent with the notion that some individuals, specifically higher-ability individuals, treat hypothetical choices as "puzzles" and may help explain why some studies find that subjects indicate that they are more tolerant of risk when they make hypothetical choices than when they make real choices. On average, subjects demonstrate a similar degree of consistency across settings, and there are also no significant differences across settings in the amount of time subjects take to make a choice, the amount of information they acquire, or how they distribute their attention.

I also find evidence to suggest that subjects acquire information in a manner consistent with the implicit calculation of expected utility. Specifically, individuals do not merely make choices "as if" they are integrating probabilities and outcomes, it appears that they actually are. Moreover, as they progress through a series of choices in a commonly used risk preference elicitation method, their information acquisition becomes progressively more consistent with integration models. Finally, on average, individuals appear to acquire information in real and hypothetical settings in similar ways.

CURRICULUM VITAE

NAME OF AUTHOR: Matthew Paul Taylor

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
Naval Postgraduate School, Monterey, California
University of California–Santa Barbara, Santa Barbara, California

DEGREES AWARDED:

Doctor of Philosophy in Economics, 2012, University of Oregon
Master of Economics, 2009, University of Oregon
Master of National Security Affairs, 2004, Naval Postgraduate School
Bachelor of Political Science, 1997, University of California–Santa Barbara

AREAS OF SPECIAL INTEREST:

Experimental Economics, Behavioral Economics, Environmental Economics, and Public Policy

PROFESSIONAL EXPERIENCE:

Intelligence Officer, United States Air Force, 1998-2007

ACKNOWLEDGEMENTS

I thank Dr. Trudy Ann Cameron for her support, guidance, and encouragement throughout this project. I am deeply indebted to Dr. William T. Harbaugh for sharing his insight and experience. I am also grateful for my discussions with Dr. Van Kolpin, who provided invaluable advice and suggestions on this project, as well as how to manage the challenges of graduate school. I thank Dr. Ulrich Mayr for providing extremely helpful suggestions and comments. All errors are, of course, my own. Danny O'Neil and Megan Ickler provided excellent research assistance. Research funding provided by the Mikesell Foundation of the University of Oregon.

For Evie, whose patience and understanding have allowed me to pursue this endeavor.

TABLE OF CONTENTS

Chapter	Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Understanding risk preferences is important to economists because many, if not most, economic decisions that individuals make are made in the context of uncertainty. If individuals are risk averse, which research suggests that most people are, appropriately measuring the impact of policy on welfare requires reliable estimates of risk preferences.

While there is general agreement that individuals are risk averse, the degree and nature of that risk aversion, as well as how risk aversion varies across individuals and in different contexts, remain questions of intense interest. However, to address these questions often requires that economists use data consisting of hypothetical choices by research subjects. Consider the issue of climate change. Many of the benefits of climate change mitigation are inherently non-market goods (e.g., reduced severity of extreme weather events, prevention of sea level rise). Further complicating the measurement of climate change preferences is the fact that actual benefits may occur far in the future. Thus, estimating the benefits of climate change mitigation generally requires the use of hypothetical choices in artificial scenarios.

There is some evidence to suggest, however, that individuals tend to choose in ways that suggest they are more tolerant of risk when faced with hypothetical choices than when they face comparable real choices. There is not a strong economic justification for why an individual would misrepresent his preferences though, given that the purely hypothetical choices have no economic consequences if they are not being interpreted as an advisory referendum.

This dissertation goes beyond the typical comparison of average risk attitudes across real and hypothetical settings. I use process-tracing software to observe how individuals acquire information about risky choices in both real and hypothetical settings. I also measure cognitive ability and estimate richer models that incorporate into the choice models a number of interactions between individual characteristics (i.e., cognitive ability and information acquisition) and the treatment consisting of real versus hypothetical settings. In addition, I examine whether individuals acquire information about the choices in a manner that is more consistent with expected utility theory or with the application of a simplifying choice heuristic, and whether this choice behavior differs across real and hypothetical settings.

Chapter II tests for differences in average risk aversion, choice consistency, and information acquisition across real and hypothetical settings. I do not find differences in average risk aversion across real or hypothetical settings, nor do I find evidence of "hypothetical bias" in a within-subject comparison. Also, there is no evidence to indicate that the consistency of subjects' choices depends on the presence of incentives. Notably, information acquisition is remarkably similar across settings. If individuals devoted less consideration to hypothetical choices than to real choices, then it is reasonable to think that they would spend less time and acquire less or different information. However, the evidence does not support this hypothesis.

Chapter III explores whether individuals tend to acquire information in sequences that are consistent with expected utility theory (and prospect theory) or with the application of a heuristic. There is virtually no evidence to suggest that subjects were using heuristics in the context of this experiment. Furthermore, there is some evidence that subjects optimized their information acquisition process as they progressed through the choices in a manner consistent with someone integrating probabilities and payoffs. These results are important because they were found in the context of a commonly-used risk preference elicitation method. Thus, they suggest that attempts to estimate risk preferences using this method are in fact doing just that.

Chapter IV incorporates individual heterogeneity into a choice model that estimates risk preferences. Significantly, the results suggest that individuals with higher cognitive ability tend to indicate that they are more tolerant of risk when they make hypothetical choices than when they make real choices. This finding is consistent with previous studies that suggest individuals may "construct" answers to hypothetical choices. Specifically, it indicates that high-ability individuals construct answers to hypothetical choices. In the context of an experiment that elicits risk preferences, it is plausible that high-ability individuals would attempt to "solve" the hypothetical choices and respond in accordance with the notion of risk neutrality.

CHAPTER II

RISK AVERSION AND INFORMATION ACQUISITION ACROSS REAL AND HYPOTHETICAL SETTINGS

Risk preferences are fundamental to decision making under risk. To evaluate the welfare impacts of policies with uncertain outcomes, we need reliable evidence about risk preferences. Frequently, however, it is problematic to use data from real choices to estimate risk preferences or to replicate in a laboratory setting the real economic choices in which we are interested. For example, environmental economists and others who consider policies involving public goods or externalities face this problem when studying risk attitudes about choices concerning fundamentally non-market goods. As a consequence, it is common for environmental economists to resort to the use of hypothetical choice scenarios to estimate preferences over specified alternatives. Hypothetical choice scenarios have also been used extensively in market research, transportation research, and health economics. In all of these areas, researches must regularly contend with non-market goods, including pre-test market products and public goods.

Laboratory experiments can be a valuable means to elicit preference information because they allow exogenous manipulation of the variables of interest (Falk and Heckman, 2009). However, the use of hypothetical scenarios is common in the laboratory setting as well because it is often prohibitively costly, impractical, or impossible to explore risk attitudes about real choices involving large stakes, such as the decision to buy a home or which medical treatment to choose. In other cases, even when the stakes are relatively small and it is possible to design experiments with real consequences, the use of real payoffs is simply impractical. For example, one recent study used hypothetical payoffs to estimate risk preferences in Afghanistan because it was considered too dangerous to carry around the relatively large amount of currency necessary to complete the experiment with real payoffs.[1]

There is some evidence, however, to suggest that individuals may respond differently to hypothetical choices involving risk than they do in similar real-choice contexts. In an influential experiment designed to evaluate objections to laboratory-based estimates of risk aversion, Holt and Laury (2002) demonstrate that individuals tend to make choices that imply that they are more tolerant of risk in a hypothetical setting than in a real setting. The Holt and Laury study also finds that this "hypothetical bias" worsens as the size of the stakes increases. Several other studies, including Holt and Laury (2005), also find evidence of hypothetical bias in risk preferences (Battalio et al., 1990; Harrison, 2006).

However, some studies find no significant difference in risk attitudes across hypothetical and real contexts (see for example Beattie and Loomes (1997); Camerer (1995); Camerer and Hogarth (1999), and Kuhberger et al. (2002)). Moreover, the question of why individuals would tend to be more tolerant of risk in hypothetical settings, rather than less tolerant, has yet to be explained. After all, the decisions people make in such purely hypothetical settings are inconsequential in monetary terms, and the stylized choices often have no externalities or potential influence on policy. Thus, hypothetical bias does not appear to have an economic justification.[2] More recently, Kang et al. (2011) used functional magnetic resonance imaging (fMRI) to show that common areas of the brain are activated when individuals make real

[1] Charles Sprenger presented preliminary results of his study at the 2011 North American Economic Science Association Conference held in Tucson, Arizona in November 2011.

[2] Murphy and Stevens (2004) discuss the lack of theoretical reasoning for hypothetical bias with respect to willingness-to-pay elicitation, but the same argument applies to risk preference elicitation. Carson and Groves (2007) consider the incentive properties of "inconsequential" survey questions in the context of valuation studies, but their conclusion that "economic theory makes no predictions" about respondent behavior is equally relevant here.

and hypothetical choices about the purchase of consumer goods, but they note that the intensity of this activity differs. These divergent experimental findings about hypothetical bias, along with the lack of a rigorous theoretical explanation, suggests the question of risk preferences across real and hypothetical contexts merits further study.

No other study has examined whether *information acquisition* behavior differs across real and hypothetical settings when subjects make choices between risky options. Wilcox (1993) explores whether subjects spend different amounts of time on decisions when there is variability in the fraction of subjects who will be paid. This mechanism, known as the random lottery mechanism (RLM), adds additional uncertainty to the experiment from the perspective of the subject because only some *subjects* will be selected at random to be paid based on their decisions.[3] Wilcox finds that subjects spend more time on higher-yielding decisions, where "higher-yielding" implies a greater probability that the subject will be selected to have at least one of her choices have real consequences. I am concerned with more than just time spent considering a decision, however. I am interested in what types of information subjects acquire and how much information they collect before making a decision.

A between-subjects comparison of average risk preferences reveals no evidence of hypothetical bias in average risk preferences in my sample. As in most studies that elicit risk preferences, however, subjects are risk averse on average. I also find that females are more risk averse than males, which is consistent with previous findings as well.

The evidence here does not support the notion that the *inconsistency* of subjects' choices increases in the hypothetical setting relative to the real setting. In fact,

[3]The uncertainty introduced by the RLM is separate from the uncertainty created by the random selection procedure (RSP), which specifies that only a subset of a *subject's decisions* will determine a subject's payoff.

individuals are surprisingly consistent given the complexity of the tasks they were asked to complete. Choice consistency does appear to be related to information acquisition behavior in expected ways: less-complete information sets are related to less-consistent choices.

Finally, individuals' information acquisition behavior is quite similar across real and hypothetical settings. Subjects do not choose to acquire less information about hypothetical choices than real choices, nor do they devote less time to their hypothetical choices. Additionally, there is no strong evidence to suggest that subjects distribute their attention differently across settings. There is a slight tendency for subjects to allocate more attention to payoffs than to probabilities in the hypothetical setting, but not to a statistically different degree.

Experimental Design

Eliciting Risk Preferences

To measure individual risk attitudes, I use a version of the Multiple Price List (MPL) format pioneered by Holt and Laury (2002), and utilized by Harrison et al. (2005, 2007) and Holt and Laury (2005). The MPL format presents ten pairs of gambles, or lotteries, and asks the subject to choose the more-preferred gamble in each pair. Columns (1) through (4) of Table 1 illustrates the conventionally ordered MPL format. The payoffs of the gambles are structured so that Gamble B is more risky than Gamble A because the differences in its payoffs are larger. In the example shown in the table, Gamble A has possible payoffs of $40 and $32 and Gamble B has higher-variance payoffs of $77 and $2. The probabilities are adjusted progressively down the list of decisions so that the difference in the expected value between Gamble A and Gamble B, calculated in columns (5) through (7) but not shown to the subject, is monotonically decreasing. A risk-neutral subject choosing solely on the basis of expected value would select the safe gamble in the first four decisions and the risky gamble in the last six decisions.

In Table 1, the rows are labeled by their Holt and Laury (HL) decision number. Depending on the decision number at which a subject switches from the safe gamble to the risky gamble, bounds for a preference parameter that measures risk attitudes can be calculated directly from these choices. Column 8 in Table 1 contains the bounds for the risk aversion parameter if we assume that utility is described by the commonly used constant relative risk aversion (CRRA) utility function, $u(w) = \frac{w^{(1-r)}}{(1-r)}$. Expositionally, the implied CRRA intervals are associated with the first risky choice that a subject chooses. For example, if a subject were to choose the safe gamble for the first five decisions and then choose the risky gamble for the remaining five decisions, then the lower bound of r can be computed by solving for r such that the subject is indifferent between the gambles in HL Decision 5:

$$0.5(\frac{40^{1-r}}{1-r}) + 0.5(\frac{32^{1-r}}{1-r}) = 0.5(\frac{77^{1-r}}{1-r}) + 0.5(\frac{2^{1-r}}{1-r}) \Rightarrow r = 0.15 \tag{2.1}$$

Similarly, the upper bound of r can by determined by solving for r such that the subject is indifferent between the gambles in HL Decision 6:

$$0.6(\frac{40^{1-r}}{1-r}) + 0.4(\frac{32^{1-r}}{1-r}) = 0.6(\frac{77^{1-r}}{1-r}) + 0.4(\frac{2^{1-r}}{1-r}) \Rightarrow r = 0.41 \tag{2.2}$$

Thus, a subject who makes his first risky choice on HL Decision 6 has a risk aversion parameter between 0.15 and 0.41.

TABLE 1. Multiple Price List

HL	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Decision	Gamble A		Gamble B					
number	p($40)	p($32)	p($77)	p($2)	EV^A	EV^B	$EV^A - EV^B$	CRRA interval
1	0.1	0.9	0.1	0.9	32.80	9.50	23.30	$r < -1.71$
2	0.2	0.8	0.2	0.9	33.60	17.00	16.60	$-1.71 < r < -0.95$
3	0.3	0.7	0.3	0.7	34.40	24.80	9.90	$-0.95 < r < -0.49$
4	0.4	0.6	0.4	0.6	35.20	32.00	3.20	$-0.49 < r < -0.15$
5	0.5	0.5	0.5	0.5	36.00	39.50	-3.50	$-0.15 < r < 0.15$
6	0.6	0.4	0.6	0.4	36.80	47.00	-10.20	$0.15 < r < 0.41$
7	0.7	0.3	0.7	0.3	37.60	54.50	-16.90	$0.41 < r < 0.68$
8	0.8	0.2	0.8	0.2	38.40	62.00	-21.60	$0.68 < r < 0.97$
9	0.9	0.1	0.9	0.1	39.20	69.50	-30.30	$0.97 < r < 1.37$
10	1	0	1	0	40.00	77.00	-37.00	$1.37 < r$

Notes: All currency units are in 2011 U.S. dollars. Subjects were not presented with the information in columns (5) through (8). The probabilities and payoffs are averages since we randomly perturbed the probabilities by up to 2 percentage points and payoffs were randomly varied between 19x, 20x, and 21x times the "low" payoffs used in Holt & Laury (2002)

Observing Information Acquisition

The method I employ to observe individual information acquisition behavior is known in decision research as "process tracing." Process tracing can be accomplished in a variety of ways, but all of these techniques have the common objective of providing a window for the researcher into the cognitive process that an individual uses to reach a decision. Early forms included manual techniques, such as the acquisition of information cards (Bettman and Jacoby, 1976; Payne, 1976). Other methods of process tracing include eye-tracking, verbal protocols, and computer-based programs.[4]

I utilize a software program called Mouselab that has been used previously by economists to explore other types of information acquisition and decision making

[4]See Arieli et al. (2011) for a recent example of an eye-tracking experiment that examines subjects' eye movements while they are considering two lotteries. Ford et al. (1989) reviews 45 experiments that use verbal protocols or "information board" techniques. See Cokely and Kelley (2009) for a recent example of a verbal protocol experiment. Verbal protocol experiments have the subjects verbalize their decision-making process. Information board techniques were more common before the computer revolution in the 1980s and required subjects to go to a board and turn over cards with information about the alternatives being considered in the choice scenario.

(Costa-Gomes et al., 2001; Gabaix et al., 2006; Johnson et al., 2008). In a way, Mouselab is a lower-tech, more expedient alternative to eye-tracking technology and provides analogous measures of attention. Mouselab has been used by psychologists as a lower-cost alternative to eye-tracking for more than two decades (Payne et al., 1988). Mouselab enables the researcher to "mask" selected information so that a subject must move the computer pointer over the attribute to reveal it. In the experiment, the probabilities and payoffs (the attributes) of each gamble are masked and require the individual to "unmask" the attributes by scrolling over the cell with that information in order to view it. An example of how each pair of gambles was presented to subjects in our experiment is shown in Figure 1. Once the mouse pointer leaves the cell, the attribute is hidden again. By requiring a subject to unmask the gambles' attributes and allowing them to view only one piece of information at a time, the process by which individuals acquire information about risky choices is observable to the researcher in this stylized context.[5]

Information Acquisition Behavior

It is useful to distinguish the information that a subject acquires in terms of its *breadth* and its *depth*. I define breadth as the fullness of an individual's information set and measure it in three ways. The first breadth measure is how many unique attributes a subject opened for a given choice. Each choice has eight cells, so this measure will range from zero to eight for each choice.[6]

[5]The experiment was presented in color. Subjects were required only to "mouseover" an attribute to reveal it and were not required to "click" on the attribute. Requiring a click is a design feature we could have employed but we believe that it would increase the differences between our design and an eye-tracking design without an obvious benefit.

[6]To account for the possibility that subjects may "accidentally" unmask a cell we include only acquisitions for which a cell is unmasked for at least 100 milliseconds (Payne et al., 1993). However, our results are not qualitatively different if we use a 50 millisecond threshold.

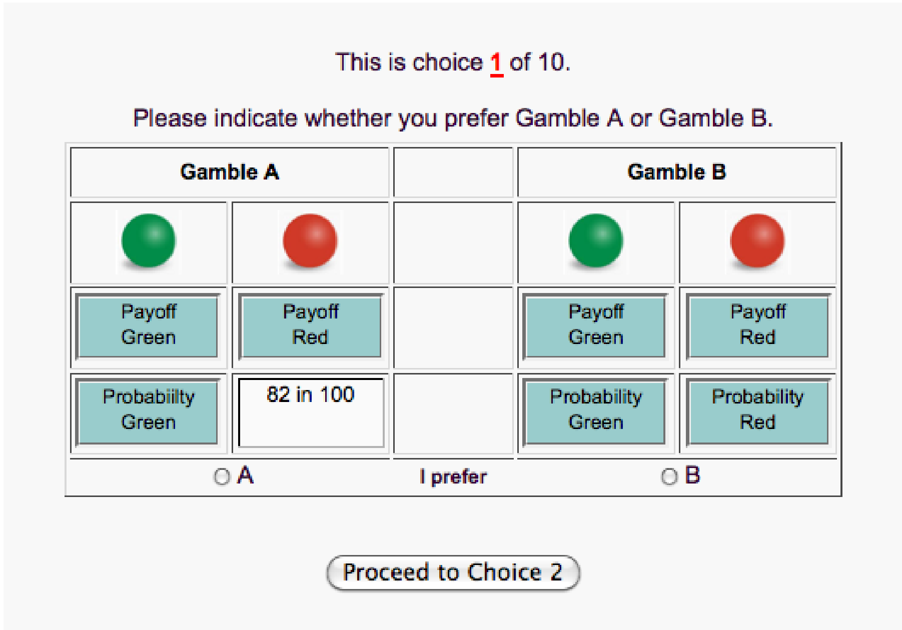


FIGURE 1. Choice example

The second measure of information set breadth is whether a subject viewed at least one probability in each gamble and all four payoffs in a given choice scenario. This measure acknowledges that some subjects will recognize the complementary nature of the probabilities in each gamble as well as the fact that they can compute expected value without opening both probabilities. This measure is called *EV-sufficient*, and it takes a value of zero or one for each choice depending on whether the subject viewed sufficient attributes to compute expected value.

The third measure of information set breadth is determined at the subject level and is an indicator for whether a subject satisfied the *EV-sufficient* measure for all ten choices in the task. In other words, it indicates whether the subject unmasked enough information on all ten choices to compute the expected value of both gambles. This measure is referred to as *EV-sufficient All Choices*.

An alternative way to think about information acquisition behavior is in terms of its depth. Whereas the breadth indicates the completeness of the information set, the depth provides a measure of how "deeply" an individual considers the information available to him before he makes a decision. The depth of a subject's information set is measured by the number of cells that she opens (including repetitions) before making a choice, as well as how much time she spends making a choice. I also explore whether subjects open probabilities and payoffs differently across settings.

There is no doubt that subjects are willing to devote at least some time to the consideration of hypothetical choices. Psychologists frequently rely exclusively on hypothetical decisions to explore questions of interest. Some of these studies have used Mouselab or other process-tracing techniques and these studies demonstrate that individuals do spend time considering hypothetical choices. The question of interest here, however, concerns whether the amount of time and effort that individuals devote to hypothetical decisions is different from the amount they devote to real decisions. Also, given the variety of information made available to the individual in our experiment, is there a difference—across real and hypothetical decisions—in the types of information subjects choose to acquire, or in how thoroughly they attend to it?

Modifications to the Multiple Price List

The goal is to observe how subjects acquire information about alternatives under uncertainty while simultaneously measuring their risk attitudes. Thus, in addition to masking the attributes of the gambles and requiring subjects to unmask them, I depart from previous studies that have used the conventional MPL design in four other important ways.

(a.) Randomized order of the choices. First, the order of the pairs of gambles is randomized and subjects are presented with only one pair of gambles at a time. Again, the conventionally ordered presentation of the MPL is shown in Table 1. Facing this

conventional order, once an individual switches from the safe gamble to the risky gamble, it would be inconsistent (according to expected utility theory) to switch back to the safe gamble in subsequent decisions. In fact, several studies simply show subjects all ten choices at the same time and ask them to identify the gamble-pair at which they would switch.[7] That structure is not imposed here. I believe it is a virtue of our design to shuffle the order in which the choices are presented as well as to present only one pair of gambles at a time. This design allows me to disentangle (a.) how individuals acquire information about more difficult choices in terms of the difference in their expected values, from (b.) learning effects that may occur as subjects become more familiar with the experimental design. In fact, if the order in which the pairs of gambles are presented is not randomized, a subject could avoid unmasking any attributes by the time she reaches the eighth or ninth gamble-pair in the conventionally ordered MPL. With experience, he could guess the values of the attributes.

(b.) Random perturbations in probabilities. Second, the probabilities are randomly perturbed, dynamically, by up to two percentage points on each gamble.[8] For instance, consider HL Decision 7 in the conventionally ordered MPL. Rather than seeing a 70 percent chance of the high payoff in both the safe and risky gambles, a subject could be presented with a 72 percent chance of the high payoff in Gamble A and a 68 percent chance of the high payoff in Gamble B. A small perturbation of the probabilities is necessary to minimize subjects' easy recognition of the similarity of the probabilities across gambles. Two percentage points is large enough to keep

[7]Harrison et al. (2007) employ this strategy, as do Tanaka et al. (2010) in a modified version of the MPL.

[8]PHP code was written so that the probability of the high payoff for each gamble was randomly determined within a specified range. This probability is then subtracted from 1 to determine the probability of the low payoff in each gamble.

the probability information valuable, while preserving the monotonically decreasing relationship in the difference in expected values of the gambles. Without these perturbations, the probabilities in both gambles in a gamble-pair would have been equal and subjects would have been able to unmask a single probability and infer the other three probabilities. It is still possible for a subject to infer the likely range of the probabilities of the other gamble after unmasking only one gamble, but not the exact probabilities. The data suggest that individuals considered the probability information worth unmasking; subjects unmasked at least one probability in each gamble in roughly 92 percent of the choices.

(c.) Random scaling of the payoffs. Third, the payoffs for each pair of gambles were randomly scaled by a multiple of 19, 20, or 21 times the "low" payoffs used in Holt and Laury (2002). In that study, the low payoffs for the safe gamble were \$2 and \$1.60 and the low payoffs for the risky gamble were \$3.85 and \$0.10. In our experiment, if the payoffs for a gamble-pair were scaled by 19, then a subject saw safe payoffs of \$38 and \$32 and risky payoffs of \$73 and \$1.90 for that particular pair of gambles. If a gamble-pair was scaled by 21, then a subject saw safe payoffs of \$42 and \$34 and risky payoffs of \$81 and \$2.10. The "20 times" scale is shown in Table 1. The scale of the payoffs is varied here because if they were the same, subjects might have begun to ignore this information after making just a few choices.

(d.) Randomization of the choice format across subjects. Finally, to address concerns about possible left-to-right and top-to-bottom bias in subjects' attention to different areas in a choice display, the presentation of the gambles is randomized along three different dimensions. I altered (i) whether the probabilities or the payoffs are presented in the top row; (ii) whether the safe gamble or the risky gamble was situated on the left; and (iii) whether the high payoff or low payoff was on the left within each gamble. These randomizations are by subject, and not across choice scenarios for

any one subject. The instructions for each subject are designed so that the tutorial material is presented in the same format as the actual choice scenarios that are later displayed for that subject.

Experimental Sessions

The experiment was completely computer-based and the subject's interaction with experimenters was minimal.[9] A typical session included about four (3.6) subjects but never more than six, and subjects were isolated by dividers. Before subjects began the experiment, they were read some instructions that explained that they would be asked to make some choices and that there were no correct choices. They were also told that they were not competing against one another.[10] The remainder of the instructions, including a tutorial that explained the choices to the subjects and tested their understanding, were all presented via the computer. Appendix A includes the full script of these instructions along with an example of how the experiment was presented to subjects.

Each subject made choices in both the real and hypothetical treatments. The type of treatment each subject received first was randomly assigned. It was not revealed to subjects at the beginning of the experiment that they would receive both real and hypothetical treatments, only that there would be two sections, and in each section

[9] The interaction between the subjects and the experimenters was limited to signing in the subjects, reading a brief set of general instructions, verifying the subject's payment, and paying the subject.

[10] They were further instructed that if they had any questions they should write them down on the note card next to their computer and then raise their hand and we would check to see if it was a question that we could answer. Having subjects write down any questions served two purposes. First, we simply wanted to minimize the distractions to the other participants. Second, because some subjects saw the real choices first and others saw the hypothetical, we wanted to reduce the possibility that a question from one subject would reveal the nature of the alternative treatment to other subjects. No one asked a question during the instruction portion of the experiment. A few students asked clarifying questions during the numeracy tests and the debriefing questionnaire. For example, some asked whether they could use a calculator (they could not).

they would be asked to make some choices. All subjects received the same instructions and the same tutorial about how the gambles would be presented and how to unmask the probability and payoff information relevant to their choices. The computer-based tutorial explained the gambles to the subjects using a vignette in which they were asked to imagine that there were two opaque jars filled with green balls and red balls. They would be shown the probability of drawing a green ball or red ball from each jar and they would also be shown the payoff associated with drawing each color from each jar. They were then instructed that they would be asked to choose the jar from which they would prefer to draw a ball.

Subjects were then informed that there would be two sections (tasks), each involving some choices. Those who were randomly selected to receive the real-choice treatment in the first task were informed that they would make ten choices, one of which would be selected to determine their payoff. These subjects were reminded that, although only one choice would be selected, each choice was equally likely to be selected so they should make each decision carefully. Those designated to complete the hypothetical choices in the first task were informed that they would make ten choices, one of which would be selected, and they would then be shown how much their payoff *would have been* had the choice been real. These subjects were then asked to consider these choices *as if* they would count toward their actual payoff.

After these initial instructions, subjects made their first set of ten choices in their randomly assigned first task. Once this first task was complete, they were notified that the results from the task would be revealed to them after they had completed the choices in their second task. Subjects then saw the instructions for whichever treatment they had not yet completed. After the subjects completed the second task, the results in terms of the real or the hypothetical payoffs from both tasks were revealed to them in the order in which the subject completed the tasks.

Once subjects completed the choice tasks and their payoffs were revealed, they were asked to complete a test that measured both numeracy and cognitive ability. Numeracy tests are designed to measure an individual's ability to understand and manipulate numeric and probabilistic information. The computational complexity of our tasks, along with the impact that numeracy could have on information acquisition behavior, made it prudent for us to elicit such a measure of numeracy. I adapted an eight-item test developed by Weller et al. (2011) that includes two of the three items in the cognitive reflective test (CRT) introduced by Frederick (2005). The third item from Frederick's CRT is included in the test as well, so I have two separate measures of ability: a numeracy score that measures a subject's numeracy, and a CRT score that measures a subject's cognitive ability.[11]

The final section of the experiment was a debriefing questionnaire that inquired whether subjects had been distracted during the experiment, whether they were liquidity constrained, the income level of their household, their education level and educational aspirations, the extent of their math and probability training, their academic major, and their gender. Table 2 provides selected summary statistics based on subjects' responses to this questionnaire.[12]

Twenty-seven experimental sessions were conducted during March and April of 2011. A total of 98 people, recruited primarily from undergraduate chemistry, economics, and environmental studies courses at the University of Oregon, participated in the experiment. Subjects earned an average total payoff of $52.68, with a maximum of $86.00 and a minimum of $6.90. Nearly all subjects finished within 40 minutes.

[11] The complete test is included in Appendix A.

[12] The complete set of summary statistics is included in Appendix A. Where appropriate, some response categories are combined into a single category.

TABLE 2. Summary statistics

Comparison of summary statistics across treatments. "Real First" indicates subjects completed the real choices in the first task; "Hyp. First" indicates subjects completed the hypothetical choices in the first task

Variables	Real first	Hyp. First	p-value
N	48	49	–
Female	.46	.35	.268
Numeracy Score	4.56	4.43	.545
CRT Score	1.42	1.12	.138
1(Distracted)	.04	.06	.667
Stats/Prob. Course	.58	.55	.751
Major			
Economics	.10	.10	.973
Environmental Science	.21	.16	.573
Science	.23	.22	.957
Business	.21	.18	.763
Other	.25	.33	.411

Notes: One subject was dropped from the analysis because she did not unmask a single attribute in seven of her ten choices in the real task. See Appendix B for the full set of summary statistics.

After dropping one subject because she did not unmask a single attribute in seven of the ten real choices, the sample includes a total of 97 subjects. Forty-eight subjects completed the real choices in the first task; the remaining forty-nine completed the hypothetical choices first. Based on two-sided t-tests, if the sample is partitioned by whether a subject completed the real or hypothetical choices first, the samples are not statistically different on dimensions commonly found to be related to risk attitudes.[13]

Approximately forty percent of the sample is female. Subjects answered 4.49 questions correctly, on average, on the six-item numeracy test. The average number of correct answers on the three-item CRT is 1.27, which is comparable to the 1.24 average score in Frederick (2005). The two scores have a correlation coefficient of 0.397. Males tended to score higher on both the numeracy test and the CRT, but not to a statistically different degree (p-value > 0.6044). Only five subjects in the entire

[13]See Table A.1 in Appendix A.

sample indicated that they were "distracted at all during this experiment." More than half of the subjects have completed a course in which at least two weeks were spent on probability and statistics. Finally, unlike many experiments that primarily use economics and psychology majors, our sample is more than one-fifth science majors, nearly one-fifth environmental studies/science majors, and another one-fifth business majors.

Risk Aversion

Similar to the findings of Holt and Laury (2002, 2005), Harrison et al. (2005), Harrison (2006), and Harrison et al. (2007), the sample is risk averse on average. Figure 2 displays the percentage of subjects who choose the safe option for each of the ten decisions in the first task. Recall that by design, a risk-neutral individual using a simple expected value criterion would choose the safe gamble in the first four decisions and then the risky gamble in the last six. This is shown in Figure 2 as a step function using a thin dotted line.

The mean number of safe choices is 6.20 in the first task. This is within the range for the average safe choices observed for subjects facing payoffs of similar magnitude in previous experiments conducted by Holt and Laury (2002, 2005), where our design is most similar to their case with "20x" payments. The subjects in their "unordered" experiment, meaning subjects completed only a single task (real or hypothetical), made an average of 6.70 safe choices.[14]

In contrast to Holt and Laury (2002) and Harrison (2006), however, I do not find differences in implied risk preferences between real and hypothetical settings. It is clear in Figure 2 that the lines tracing out the percentage of safe choices are nearly identical. If there were hypothetical bias in the form of lesser risk aversion,

[14]See page 903 of Holt and Laury (2005) for a summary of their results.

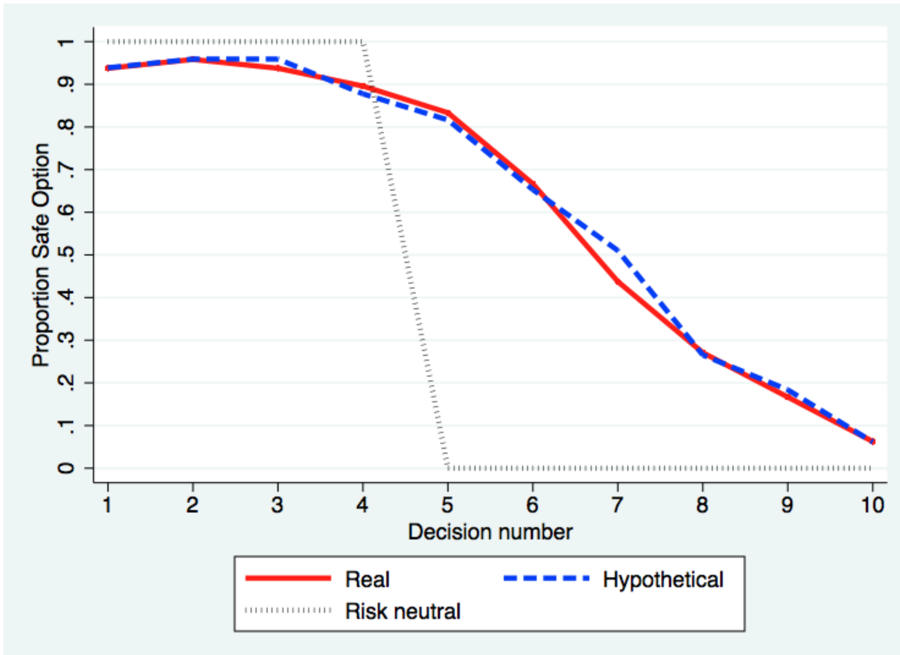


FIGURE 2. Real versus Hypothetical: Proportion choosing safe gamble in each HL Decision number in the first task

the locus of choices in the hypothetical setting (the dashed line) would be shifted to the left if subjects switch sooner to the riskier alternative. Non-parametric tests are consistent with the results implied in Figure 2. Subjects made 6.23 safe choices on average in the real setting and 6.17 in the hypothetical setting during the first task. A test of equal distributions fails to reject the null hypothesis using either a Wilcoxon rank-sum test or a Kolmogorov-Smirnov test (p=0.9765, p=0.998 respectively).[15] Moreover, although I focus on the choices in the first task because of possible order effects (Harrison et al., 2005), there is not a significant difference in the number of safe choices when all the choices from both the first and second tasks are included.

[15] I conducted an identical analysis using the decision number of the last safe choice before the first risky choice and the results are substantively unchanged.

Within-subject Risk Aversion Across Real and Hypothetical Settings

Analyzing within-subject behavior is problematic in this context because of order effects (Harrison et al. 2005). However, because my concern is that a subject may respond differently in a hypothetical setting compared to a real setting, we are interested in within-subject differences. As a crude non-parametric test, when I restrict the sample to the 63 subjects who made zero inconsistent choices in either task, there are 14 subjects who made *more* safe choices in the real setting than in the hypothetical setting (22.2 percent of the sample). This means that about one-fifth of the sample demonstrated "hypothetical bias" in the sense of reduced risk aversion in the hypothetical setting. However, another 18 percent of the sample chose *fewer* safe choices in the real setting than the hypothetical setting, which is a bias in the opposite direction. In all, 25 out of 63 subjects (39.7 percent) chose a different number of safe choices in the hypothetical setting than the real setting.

Table 3 shows the results of a logit model in which the dependent variable is equal to one if the subject chose an equal number of safe choices in both settings.[16] In this model, I include an indicator variable for whether the subject completed the hypothetical choices in the first task, an indicator variable for female, CRT score, and numeracy score. To control for the overall level of a subject's effort and attention in both tasks I include the subject's average decision duration and the average number of unique attributes she opened. Finally, I also include a pair of measures to capture how differently the subject acquired information in the real setting compared to the hypothetical setting. I compute the difference in the average amount of time the subject spent on the real choices and the hypothetical choices and the difference in the average number of unique attributes the subject acquired in the real setting and the hypothetical setting.

[16] As I discuss more thoroughly below, reversals capture the noise in a subject's choices by counting the number of inconsistent choices among the ten choices in a task. Thus, I am focused only on subjects who made fully consistent choices in both of their tasks.

TABLE 3. Within-Subject Comparison of Safe Choices

Results of logit regression where the dependent variable equals one if the subject chose the same number of safe choices in both settings.

VARIABLES	1(Equal Safe Choices)
1(Hypot. First Task)	-0.872
	(0.419)
1(Female)	1.365*
	(0.094)
CRT Score	0.001
	(0.997)
Numeracy Score	0.863***
	(0.007)
Subject's Mean Decision Duration	-0.000
	(0.308)
Duration Real - Duration Hypot.	-0.000
	(0.922)
Subject's Mean Attributes Unmasked	0.623
	(0.200)
Uniq. Attrib. Real - Uniq. Attrib. Hypot	0.210
	(0.613)
Constant	-6.926*
	(0.074)
Observations	63

Notes: Robust p-values in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3 shows that information acquisition behavior is not predictive of choice differences in real and hypothetical settings. Gender and numeracy, however, have statistically significant effects in this model. Females are more likely to choose an equal number of safe choices in both settings, as are subjects that score higher on the numeracy test. I emphasize that these results apply only to the subjects who made no reversals in either task. Restricting the sample in this way causes us to drop half of the forty female subjects who participated in this experiment.

Choice Consistency

Previous studies using the MPL design have found inconsistent choice behavior problematic rather than informative because, at a minimum, it widens the bounds of the risk aversion parameter. My interest here in the consistency of individuals' choices stems from the notion that the level of consistency conveys valuable information about whether individuals treat hypothetical choices as if they are real. It is conceivable that subjects choose less consistently because hypothetical choices have no consequences. However, I do not find that subjects are less consistent in the hypothetical setting than the real setting, on average.

There are several different ways to measure the level of a subject's consistency in a decision task. My preferred measure is one that counts the number of reversals that a subject makes. As introduced briefly in the previous section, a reversal occurs when a subject chooses the safe gamble from a gamble-pair after having picked the risky gamble in the previous choice (when the gamble-pairs are re-arranged into the conventional Holt and Laury order). Table 4 helps to illustrate this point. Based on the choices shown in the table, Subject A made only one reversal: HL Decision 7. Subject B made two reversals: the risky choice for HL Decisions 4 after choosing the safe choice for HL Decision 3, and the risky choice for HL Decision 8 after choosing the safe choice for HL Decision 7.[17]

[17] An alternative way to measure choice consistency is to count the number of safe options the subject chooses after choosing a risky gamble in an earlier choice. This measure of inconsistent behavior may be an overestimate because it counts every risky choice before the inconsistent safe choice as an inconsistent choice. In fact, it may be the case that only the single safe choice in the middle of the run of risky choices is actually the inconsistent choice. For example, suppose that Subject A made three safe choices and then appears to have switched over to the risky gambles at HL Decision 4. However, he chose the safe gamble on HL Decision 7. Clearly, this is an inconsistent choice but this measure of inconsistency essentially classifies the risky choices on HL Decisions 4, 5, and 6 as inconsistent rather than viewing the safe choice on HL Decision 7 as inconsistent. As for Subject B, if we measure inconsistency as the number of safe choices after the first risky choice then he has five inconsistent choices.

TABLE 4. Measuring inconsistency

Example of two subjects' choices. Subject A made one reversal. Subject B made two reversals.

Did subject choose the safe gamble? Yes=1										
HL Decision #	1	2	3	4	5	6	7	8	9	10
Subject A	1	1	1	0	0	0	1	0	0	0
Subject B	1	1	0	1	1	1	0	1	1	0

Given the complexity of the choice tasks, particularly given the random ordering of the choice scenarios in this experiment and the perturbation of the probabilities in each gamble—subjects demonstrated a remarkably level of consistency in these tasks. Table 5 shows that approximately 70 percent of subjects made consistent choices in the treatment they complete first, regardless of its setting.[18] Furthermore, more than 90 percent of subjects completed the first task with one or fewer reversals.

Table 5 shows that subjects did not behave any less consistently in the hypothetical setting than in the real setting. The distributions of reversals in each setting, when it was the first task presented to the subject, are nearly identical. A Wilcoxon rank-sum test fails to reject the null hypothesis that these distributions are equal (p=0.9534).

It is common practice to exclude subjects who choose the safe choice on HL Decision 10, where the high payoffs in both the safe and risky gamble have 100 percent probability. The suspicion is that making the safe choice under these conditions indicates that a subject probably does not understand the instructions. Seven subjects chose the safe option for HL Decision 10 in this experiment (four subjects who completed the real choices in the first task and three subjects who completed the hypothetical choices in the first task). These subjects are included in the present

[18]Similarly, Lévy-Garboua et al. (2011) find that 74 percent of subjects made consistent choices when gamble-pairs were presented sequentially rather than simultaneously when using payoffs half the size of those we used.

TABLE 5. Number of Reversals By Setting

# of Reversals in First Task	Real	% of Subjects	Hypothetical	% of Subjects
0	34	70.83	34	69.39
1	10	20.83	12	24.49
2	3	6.25	3	6.12
3	1	2.08	0	0

Note: Between subjects Wilcoxon p-value = .953

# of Reversals in Second Task	Hypothetical	% of Subjects	Real	% of Subjects
0	42	87.50	38	77.55
1	6	12.50	8	16.33
2	0	0	3	6.12

Note: Between subjects Wilcoxon p-value = .169

analysis, however, because in this study the attributes of the gambles are masked and a subject could have chosen the safe option because he did not have sufficient information to ascertain the unique properties of Decision 10. Either way, the results do not differ qualitatively if these subjects are excluded.

The results discussed above suggest that the level of consistency across settings is independent of whether the task is real or hypothetical. This hypothesis is explored further by modeling the number of reversals as a function of four factors that could directly impact an individual's ability to sufficiently comprehend and integrate the information in these tasks to arrive at expected values for each gamble: cognitive ability, the fullness of the subject's information set, and time spent considering a choice.[19]

The level of complexity of the experimental task was sufficiently high that someone with low cognitive ability may not have been able to understand fully the choices

[19]Previous exposure to the subject matter of probability and statistics was also included in an alternative specification but was not statistically significant. The results are not qualitatively different without it so it is not included here.

presented to them. Hence, someone with low cognitive ability or low numeracy would be expected to have more reversals. Also, subjects who did not acquire sufficient information to compute expected value would be more likely to make more reversals as well. An individual who unmasked only one attribute must have guessed about the other attribute levels, and guessing is likely to result in errors that show up as reversals.

The expected relationship between the amount of time that a subject spends on each decision and the consistency of her choices in a task is less clear. Subjects must spend a minimum amount of time simply to acquire enough information to compute expected value, so subjects may initially improve the consistency of their choices as they spend more time. Conversely, “excessive” amounts of time could indicate that a subject is confused or distracted. Based on this reasoning, decision duration is included in the model quadratically, to account for the possibility that there may be a non-linear relationship between decision duration and the observed number of reversals.

The following equation is estimated using a Poisson regression because the dependent variable, the number of reversals committed by subject i in the first task, is count data for which a large fraction of the sample equals zero. Within the stochastic structure of the Poisson model, it is assumed that: where $1(Hypot)$ is an indicator for the hypothetical setting, $1(Female)$ is an indicator for female, $Numeracy\ and\ CRT_i$ is the subject's combined numeracy and CRT score, *1(EV-sufficient All Choices)* is an indicator equal to one if the subject acquired enough information to compute expected value for every choice in the task, and $Mean\ Ln(Decision\ Duration)_i$ is the average of the natural log of decision duration for all ten choices in the task. The coefficient estimates for this specification are shown in Table 6.[20]

These parametric results are consistent with the earlier results which indicate that subjects did not make systematically more reversals in the hypothetical setting.[21] I do find, however, that information acquisition behavior is related to the overall consistency of a subject's choices in a task. Subjects who acquired an EV-sufficient information set for every decision in the task made fewer reversals. Additionally, the relationship between decision duration and the number of reversals is consistent with speculation that an increase in decision duration may lead to fewer reversals, but too much time spent could be a signal of confusion and point to more reversals.

Somewhat surprisingly, cognitive ability and numeracy are not related to the number of reversals.[22] Females tend to make more reversals than males. Why females tend to make more reversals than males, on average, is puzzling. Females score just as well as males on the ability test (i.e. the combined numeracy and CRT score)—5.79 (males) versus 5.71 (females) (t-test p-value = .8340). Females also do not differ significantly from males in terms of the proportion who have been exposed to coursework in probability and statistics (60% of males versus 51% of females, p-value=.3824).

[20]A Poisson model may be too restrictive if there is over- or under-dispersion in the data, or if there are an excessive numbers of zeros. I have explored more general specifications for count data, including negative binomial and zero-inflated models. These specifications do not reject the restrictions embodied in the Poisson, so I report only the Poisson model.

[21]An alternative specification was estimated that included interactions between the other variables in the model and the indicator for hypothetical choices. None of these interactions were significant and a likelihood ratio test failed to reject the null of equal goodness-of-fit between our model and the model that included the interactions. As a consequence, we do not include in Table 6 the alternative specification with these interactions terms.

[22]Estimating the model with CRT score and numeracy score included separately does not qualitatively change the results. Moreover, the coefficients and p-values on each score when they are included separately are nearly identical to those shown in the table.

TABLE 6. Poisson regression results where the dependent variable is the subject's total number of reversals in the first task.

VARIABLES	Inconsistency # Reversals
1(Hypot)	0.171
	(0.615)
1(Female)	0.771**
	(0.015)
Numeracy + CRT	-0.051
	(0.574)
EV-sufficient All Choices	-0.936***
	(0.007)
Mean Ln(Decision Duration)	-24.479*
	(0.058)
Mean Ln(Decision Duration)^2	1.290*
	(0.056)
Constant	115.075*
	(0.063)
Observations	97

Notes: Robust p-values in parentheses. Negative binomial and zero-inflated Poisson specifications do not reject the ordinary Poisson model. Estimating the model with CRT score and numeracy score included separately does not qualitatively change the results.
*** p<0.01, ** p<0.05, * p<0.1

Taken together these results suggest that reversals in this experiment are related to information deficiencies and gender rather than lower cognitive ability or a lack of consequentiality. Moreover, there does not appear to be a relationship between the number of reversals and the real or hypothetical setting in which the choices were made.

Breadth of Information Acquisition

Turning to individual information acquisition behavior, I find no evidence to indicate that individuals acquire less-complete information in the hypothetical setting relative to the real setting. Table 7 shows the percentage of choices where subjects viewed sufficient information to compute expected value, *EV-sufficient*, in each treatment. Subjects viewed enough information to satisfy the EV-sufficient criterion in approximately 85 percent of all choices in the task they completed first, regardless of whether the setting was real or hypothetical. These percentages are not statistically different across the two settings, based on a t-test (p-value = 0.894).

TABLE 7. Percentage of choices where subjects viewed sufficient information to compute expected value *(EV-sufficient = 1)*.

	Percent of choices		
	Real	Hypothetical	t-test p-value
First Task	85.00	84.69	.894
Second Task	77.96	62.84	$< .001$

Things are different, however, in the second task. The second row of Table 7 shows the percentage of choices that were adequately viewed to compute expected value in this second task. Subjects who completed the real choices in the second task viewed EV-sufficient information for approximately 78 percent of the choices in this task. In sharp contrast, subjects who completed the hypothetical choices in the second task viewed EV-sufficient information for less than 63 percent of the choices. There is a similar pattern in all five of the information acquisition measures. Specifically, subjects who completed the real choices in their first task and the hypothetical choices in their second task acquired less complete information and spent considerably less time on each choice in the hypothetical setting relative to the real setting.

I can only speculate as to the reasons for this pattern in information acquisition behavior between the first and second task for those who completed the real choices first and hypothetical choices second. One possible explanation may be that subjects were told that they would complete two sets of choices. Thus, subjects who completed the real task may have realized, once they received the instructions for the hypothetical choices in the second task, that their payoff had been fully determined by the first task. This explanation is unsatisfactory, however, because these same subjects actually make statistically fewer inconsistent choices in the second task than the first, even though the second task was hypothetical. In contrast, subjects who completed the real choices in the second task did not improve in terms of their consistency. Therefore, it appears that some of the differences in information acquisition in the second task may be due to an increased level of concentration, or differences in learning, during the first task. This experiment was not designed to test for these types of "learning" effects, so the results from the second task are not included in the remainder of this paper. Further experiments to resolve this issue should clearly be part of the agenda for research in this area.

The other two measures of information acquisition also do not suggest that subjects acquire less complete information in the hypothetical setting. Subjects opened every cell at least once (i.e., they acquired eight unique attributes) in approximately 83 percent of all the choices in the task they completed first, regardless of setting. Furthermore, the distributions of the number of unique attributes acquired in each setting in the first task are not statistically different (Wilcoxon p-value=0.871).

The third measure of information breadth captures the completeness of a subject's information set for the entire task. Recall that *EV-sufficient All Choices* is equal to one if the subject opened sufficient cells to compute expected value for all ten choices in a task. Only 22 of the 48 subjects who completed the real choices in the first task met this criterion. Of the 49 subjects who completed the hypothetical choices first, 23 acquired EV-sufficient information for all ten choices. Again, however, there is not a statistical difference between the two groups in the percent who viewed all ten choices (t-test p-value = 0.914).

Figure 3 shows the percentage of subjects who viewed EV-sufficient information for each choice in the order it was actually presented, i.e., whether they saw the choice first, second, and so forth. To distinguish the order that subjects completed the choices in our experiment from the conventional MPL ordering used by Holt and Laury (2002), the order in which a subject actually completed a choice is referred to as the *Position number*, as opposed to the HL Decision number. Both plots in the figure are nearly horizontal. This means subjects acquire nearly the same breadth of information for the first choice as they do for the tenth choice they make, which is consistent with the notion that they treat each choice individually and that they seem to have been engaged in the choice process through the entirety of the task. Additionally, the plots for percentage of subjects that satisfy the EV-sufficient criterion follow nearly identical paths in both the real and hypothetical settings, which is consistent with the finding above, that the breadth of information acquired is not statistically different across settings.

Figure 4 shows the percentage of subjects who viewed EV-sufficient information for each choice arranged instead by HL Decision number. Again, the plots for the real and hypothetical tasks follow similar paths. However, they also appear to be slightly higher for the decisions in the middle of the HL ordering, indicating that subjects sought more-complete information sets for these particular choices. These are the choices for which the difference in expected value is smallest and where individuals tend to switch from the safe choice to the risky choice. These choices may have been "tougher calls." This pattern is verified in the parametric analysis below.

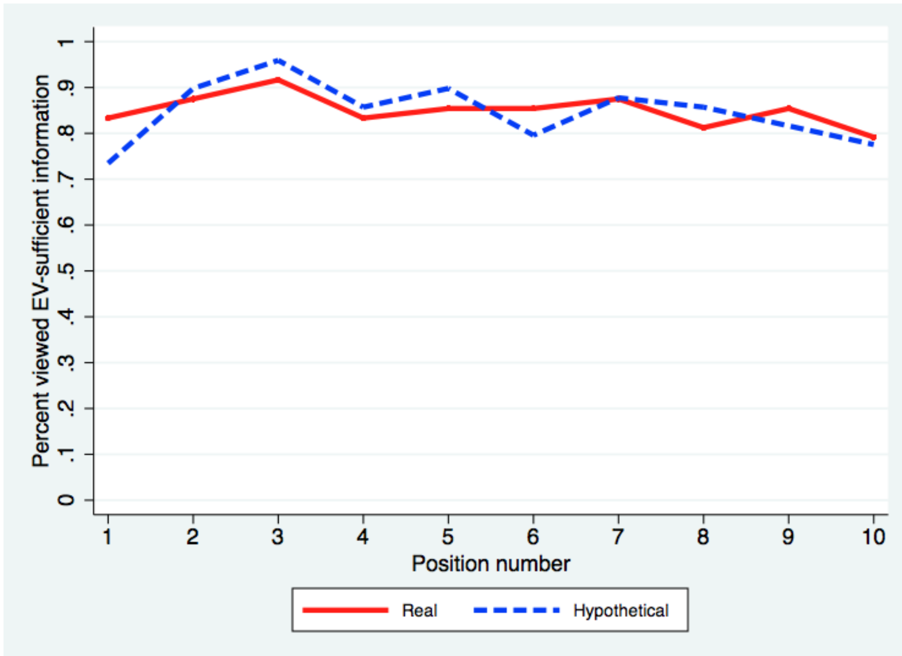


FIGURE 3. Real versus Hypothetical: Percent of subjects who viewed sufficient information to compute expected value by the position number of the choice in the first task.

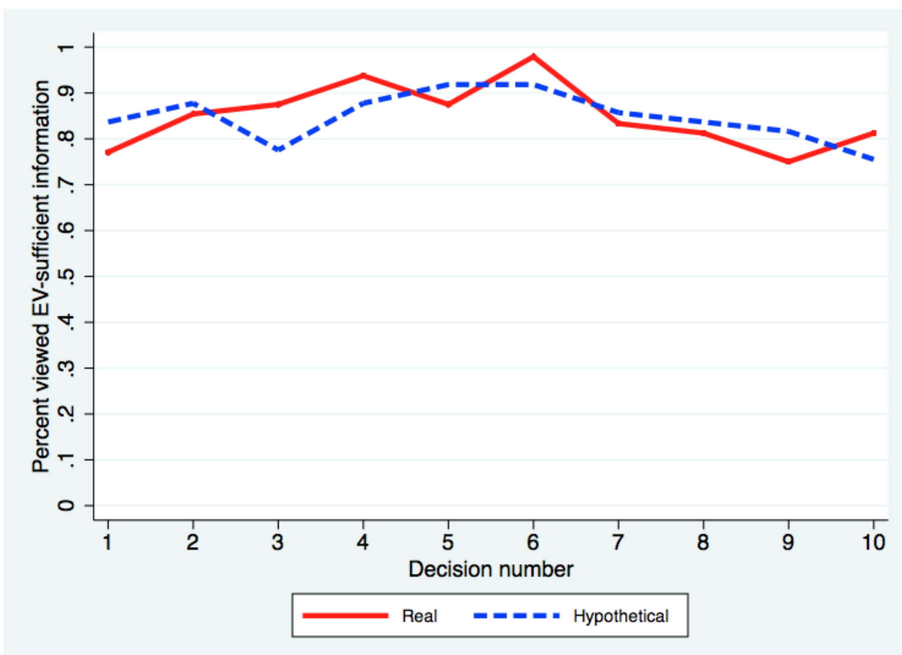


FIGURE 4. Real versus Hypothetical: Percent of subjects who viewed sufficient information to compute expected value in each decision in the first task.

Taken together, these results suggest that individuals do not acquire less information about risky decisions in hypothetical settings than they do in real settings. Moreover, based on the breadth of information that they acquire throughout our tasks, it appears that subjects remained engaged throughout the entire task, whether it was real or hypothetical.

Depth of Information Acquisition

In addition to acquiring information of comparable completeness across settings, subjects also spend at least as much time considering choices, and open as many cells, in the hypothetical choices as they do the real choices. Table 8 shows the mean and median values of *decision duration*—the amount of time a subject spends making each decision. For those who completed the real choices in the first task, the mean decision duration is 13.77 seconds. The mean decision duration for those who completed the hypothetical choices first is 13.23 seconds. There is slightly more than a half-second difference between the two means, but a t-test fails to reject the null hypothesis that the means are equal (p-value=0.425).

There is evidence, however, of a small difference in the median decision duration. Despite what would be expected if subjects paid less attention to hypothetical choices, the median decision duration is actually *greater* for the hypothetical choices than for the real ones. The median decision duration for real choices is 10.17 seconds, while that for the hypothetical choices is 11.41. A Wilcoxon rank-sum test rejects the null hypothesis that the distributions are equal (p-value = 0.011).

The results are similar for *acquisition frequency*, which is the total number of cells opened, with repetitions. The first row in Table 9 shows that the mean acquisition frequency in the first task was 16.13, regardless of the setting so that a t-test of equal means obviously fails to reject the null hypothesis (p=.998). However, the median acquisition frequency in the hypothetical setting is, again, greater than the median frequency acquisition in the real setting, and a Wilcoxon rank-sum test this time rejects the null hypothesis of equal distributions (p-value=0.003).

TABLE 8. Decision Duration Across Settings in First Task

	Decision duration (in seconds)		
	Real	Hypothetical	p-value
Mean decision duration	13.77	13.23	.425
Median decision duration	10.17	11.41	.011
N	480	490	

Notes: Decision duration is the amount of time a subject spent on a choice. p-values are reported for a t-test of the equality of means (for mean decision duration) and Wilcoxon rank-sum test of equal distributions (for median decision duration).

TABLE 9. Acquisition Frequency Across Settings in First Task

	Total acquisitions by choice		
	Real	Hypothetical	p-value
Mean acquisitions	16.13	16.13	.998
Median acquisitions	13	14	.060
N	480	490	

Notes: Mean and median acquisition frequency per choice. p-values are reported for a t-test of the equality of means (for mean acquisitions) and Wilcoxon rank-sum test of equal distributions (for median acquisitions).

Do subjects tend to change their focus between probabilities and payoffs across real and hypothetical settings? The results in Table 10 show that they do not. Subjects acquired 8.05 probabilities in the real setting and 7.96 in the hypothetical setting. Similarly, subjects who completed the real choices in the first task acquired 8.08 payoffs and subjects who completed the hypothetical choices acquired 8.18 payoffs.

Considering the columns in Table 10, subjects who completed the real choices opened probabilities and payoffs nearly the same number of times. Those who completed the hypothetical choices in the first task appear to have a slight tendency to open the payoffs more than the probabilities, but a t-test that the mean probabilities acquired and the mean payoffs acquired are equal fails to reject the null hypothesis (p-value=0.178).

TABLE 10. Probabilities and Payoffs Acquired Across Settings in First Task

	Total acquisitions by choice		
	Real	Hypothetical	p-value
Mean probabilities acquired	8.05	7.96	.779
Mean payoffs acquired	8.08	8.18	.784
N	480	490	

Notes: Mean acquisitions frequency of probabilities and payoffs acquired per choice. p-values are for t-test of null hypothesis that mean probabilities or mean payoffs are equal across settings.

Looking across the two rows of Table 10 allows a comparison of the difference in probabilities and payoffs acquired across settings. Subjects who completed the hypothetical choices first acquired 0.22 fewer probabilities than payoffs on average; subjects who completed the real choices first acquired 0.04 fewer probabilities on average. Again, however, the difference is not statistically significant. A t-test of the difference in probabilities and payoffs acquired rejects the null of equal means across settings (p-value=0.445).

Although there is no statistically significant difference in the central tendencies of the number of acquisitions of probabilities and payoffs across settings, it appears that there may be a difference in the *distributions* of probability and payoff acquisitions. Figure 5 shows quantile-quantile plots of the total probabilities acquired versus the

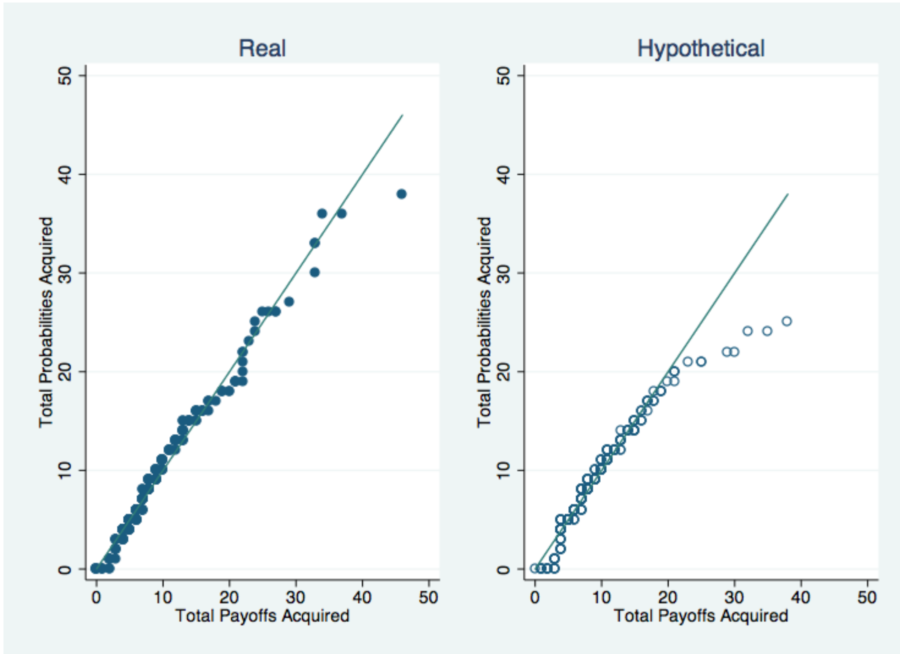


FIGURE 5. Real versus Hypothetical: Quantile-quantile plots of total probabilities acquired versus total payoffs acquired, by choice in the first task.

total payoffs acquired in the first tasks. In the plot for the real setting on the left, it is clear the distributions of probabilities and payoffs acquired are nearly identical for subjects who made real choices in the first task. The comparable plot for the subjects who completed the hypothetical choices in their first task appears on the right. The figure shows that the upper tail of the distribution for payoffs acquired is much longer than that for the distribution of probabilities acquired.

Turning to the amount of time that subjects spent on each decision, Figure 6 plots the mean amount of time spent, by HL Decision number, for the first task in each setting. Mean decision duration appears to have an inverted U-shape with respect to decision number in both settings, even though the actual position number for each choice was randomized. Given that the calculated differences in expected values between the gambles are smallest for the decisions in the middle of the conventional

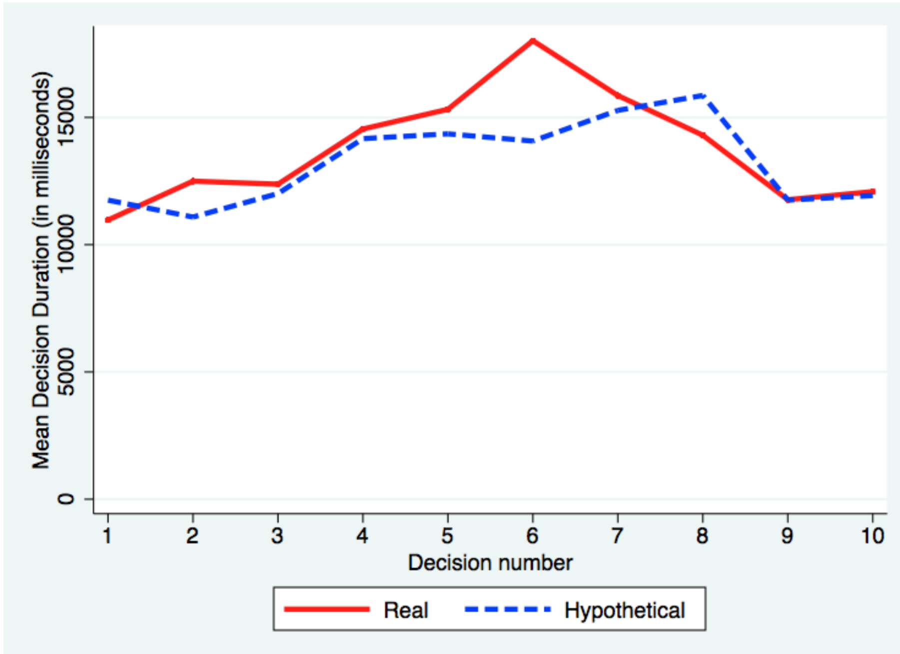


FIGURE 6. Real versus Hypothetical: *Mean* decision duration by HL Decision number in the first task

MPL order, this pattern is consistent with the notion that individuals will spend more time on decisions that are less clear-cut.[23]

Visual analysis of the median decision duration, however, suggests a possible outlier effect in the time spent on the real choices. Figure 7 seems to show that the increase in *median* decision duration over HL Decisions 1 through 6 for the real choices may not be as steep as the increase in the *mean* decision duration. In fact, median decision duration decreases for HL Decision 5. In contrast, the plot of median decision duration for the hypothetical choices looks much like the plot for the hypothetical choices in Figure 7. This may indicate that differences across real and hypothetical contexts may be caused by changes in behavior among a small group of outliers rather than by a shift in median behavior. In effect, it looks as if subjects spend more time considering the more difficult choices in both the real and hypothetical settings, but

[23]Glöckner and Herbold (2011) also find that similar expected values increase decision durations.

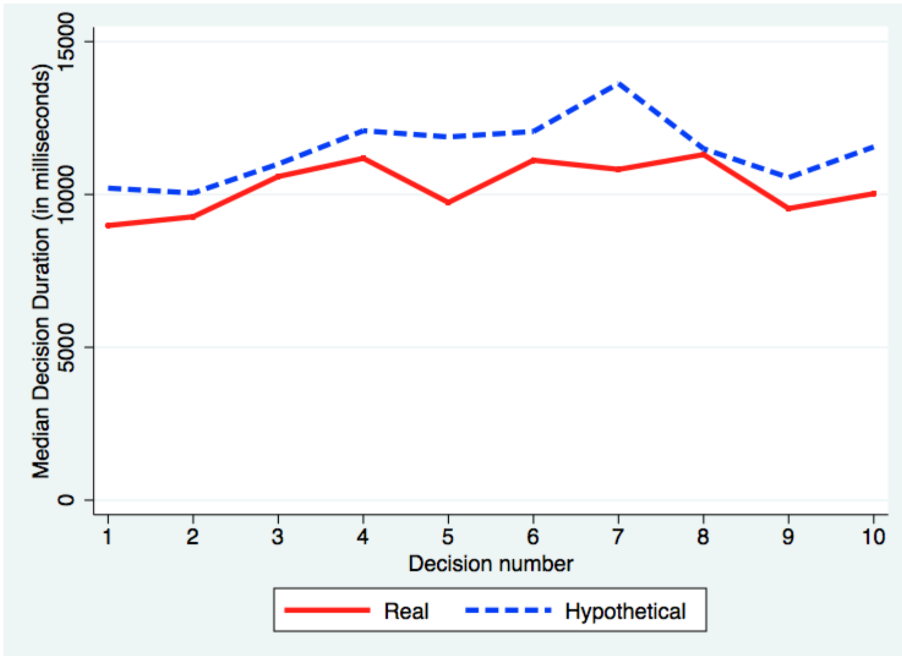


FIGURE 7. Real versus Hypothetical: *Median* decision duration by HL Decision number in the first task

the financial consequences of the choices in the real setting may induce a small fraction of subjects to undertake a considerably more deliberative decision process.

Given the graphical evidence of an apparent influence of outliers on the mean decision duration, differences in the distributions of decision durations across settings for each HL Decision number are tested using a Wilcoxon rank-sum test. The p-values for these rank-sum tests are shown by HL Decision number in Table 11 along with the median decision duration (in seconds) for the same HL Decision numbers. These tests fail to reject the null hypothesis of equal decision duration distributions across real and hypothetical settings for each HL Decision number considered separately.

Figure 7 depicts median decision durations by Position number. The figure is consistent with the expectation that individuals will spend more time on the choices they see earlier and that decision durations will decrease as subjects become familiar with the choice format, and supports the decision to randomize the choices to prevent

TABLE 11. Median Decision Duration Across Settings for each Decision in First Task, Ordered by HL Decision Number

(1) HL Decision number	(2) Real	(3) Hypothetical	(4) Wilcoxon p-value
1	8.98	10.21	.1076
2	9.27	10.05	.8853
3	10.58	10.99	.5068
4	11.18	12.09	.6862
5	9.74	11.88	.1092
6	11.12	12.06	.7075
7	10.82	13.63	.1045
8	11.31	11.50	.7895
9	9.54	10.55	.2634
10	10.03	11.56	.6598

Notes: p-values are for Wilcoxon rank-sum test of equal distributions across settings.

this behavior from confounding the analysis. It is evident that subjects reduce the amount of time spent on each choice as they complete more choices. It also appears that much of the reduction occurs within the first few choices that subjects complete, regardless of the HL Decision numbers that these choices represent.[24]

Overall, these results do not support the hypothesis that individuals put less effort or consideration into hypothetical choices. In fact, no matter which measure is used for information acquisition behavior, there is no compelling evidence to indicate that subjects devote less attention to hypothetical choices than to real choices. Neither do they focus on different types of information as they make hypothetical versus real choices.

[24] I do not include a separate figure plotting the means in this case because it is not discernibly different.

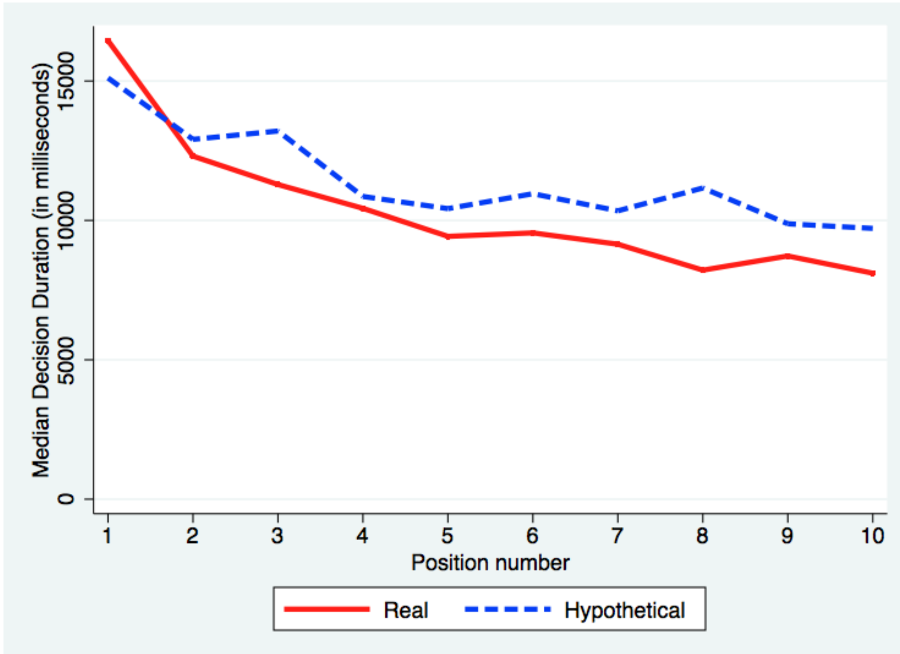


FIGURE 8. Real versus Hypothetical: *Median* time spent on each decision by Position number in which it was viewed by the subject in the first task.

Modeling Information Acquisition: Breadth and Depth

Four measures of information acquisition are considered—the number of unique attributes, whether a subject viewed sufficient information to compute expected value, decision duration, and acquisition frequency. Each measure is modeled as a function of: (1) the HL Decision number (and the HL Decision number squared), (2) the position number of the choice, (3) an indicator variable for whether the choice is hypothetical, as well as (4) individual characteristics, and interactions of these characteristics with the indicator for hypothetical choices. Specifically, four different specifications of the following form are estimated:

$$
\begin{aligned}
E[InfoAcq_{csi}] = f([\alpha_0 + \alpha_1 1(Hypot)_{csi}] &+ [\gamma_1 + \gamma_2 1(Hypot)_{csi}]Decision_{csi} \qquad (2.3)\\
&+ [\gamma_3 + \gamma_4 1(Hypot)_{csi}]Decision^2_{csi} + [\theta_1 + \theta_2 1(Hypot)_{csi}]Position_{csi}\\
&+ [\delta_{k1} + \delta_{k1} 1(Hypot)]X_i),
\end{aligned}
$$

where $InfoAcq_{csi}$ is the information acquisition measure of interest, c denotes the choice, s denotes the setting, i denotes the individual, X_i is a vector of controls for gender, and numeracy and CRT score. Standard errors are clustered at the individual level to account for within-subject error correlation. Depending upon the nature of the particular information acquisition measures (continuous, binary, or count data), we specify the model using appropriate functional form and stochastic assumptions.

Table 12 shows the results of these models using data from the first tasks only. The information acquisition variable explained by each model is specified at the top of each column. Model 1 contains the results of a Poisson specification where the dependent variable is the number of unique attributes a subject unmasked for the choice.[25] Model 2 gives the results of a logit regression to explain whether the subject unmasked sufficient attributes to compute expected value. Decision duration is modeled using OLS in Model 3. The natural logarithm of Decision duration is used as the dependent variable since durations are strictly positive. Finally, Model 4 shows the results for a negative binomial specification for the total number of acquisitions.[26]

---

[25]Negative binomial and zero-inflated Poisson specifications for the number of unique attributes unmasked were also explored but they failed to reject a Poisson specification.

[26]Although the results presented here for total acquisitions provide some evidence of over-dispersion, the coefficient estimates that result from a similar model using a ordinary Poisson specification are not qualitatively different.

TABLE 12. Information Acquisition in the First Task

Information acquisition measure used as dependent variable is shown at the top of each column.

VARIABLES	Breadth measures (1) #Unique Attributes (Poisson)[a]	Breadth measures (2) 1(EV-sufficient) (Logit)	Depth measures (3) Ln(Duration) (OLS)	Depth measures (4) #Acquisitions (Negative Binomial)
1(Hypot)	-0.011	-0.231	0.210	0.006
	(0.912)	(0.869)	(0.504)	(0.981)
HL Decision number	0.027**	0.508***	0.127***	0.180***
	(0.031)	(0.003)	(0.000)	(0.000)
(HL Decision #)×1(Hypot)	-0.010	-0.161	-0.021	-0.076
	(0.513)	(0.431)	(0.676)	(0.109)
HL Decision number^2	-0.003**	-0.049***	-0.011***	-0.016***
	(0.014)	(0.001)	(0.000)	(0.000)
(HL Decision #)^2×1(Hypot)	0.001	0.014	0.001	0.006
	(0.646)	(0.451)	(0.752)	(0.145)
Position number	0.000	-0.045	-0.061***	-0.047***
	(0.962)	(0.486)	(0.000)	(0.000)
(Position #) ×1(Hypot)	0.000	0.002	0.021**	0.031***
	(0.950)	(0.978)	(0.045)	(0.009)
1(Female)	-0.047	-0.660	-0.217**	-0.220**
	(0.154)	(0.194)	(0.049)	(0.040)
1(Female)×1(Hypot)	0.069	0.933	0.185	0.100
	(0.134)	(0.251)	(0.164)	(0.433)
Numeracy + CRT	0.010	0.208*	0.046	0.044
	(0.143)	(0.069)	(0.252)	(0.173)
(Numeracy + CRT)×1(Hypot)	0.004	0.035	-0.054	-0.005
	(0.751)	(0.853)	(0.260)	(0.893)
Constant	1.925***	0.275	9.194***	2.471***
	(0.000)	(0.775)	(0.000)	(0.000)
α	—	—	—	-0.192**
	—	—	—	(0.018)
Log-likelihood	-1994.67	-394.38	—	-3322.78
R-squared	.002[b]	.04[b]	0.12	—
Observations	970	970	970	970

Notes: Standard error estimates clustered on subject. [a]Negative binomial or zero-inflated Poisson specifications fail to reject the ordinary Poisson model. [b]Pseudo R-Squared. Estimating the model with CRT score and numeracy score included separately does not qualitatively change the results.
*** p<0.01, ** p<0.05, * p<0.1

The statistical significance of the coefficients on the linear and squared terms in the HL Decision number variable provide additional support for the conclusion that subjects acquire more-complete information sets and spend more time on the decisions that are more difficult (i.e., in the middle of the HL Decision list, where the difference in the expected value of the gambles is smaller). Furthermore, in Models 1 and 2, which explain information breadth, the coefficient on Position number is not statistically significant, but Position number is statistically significant in Models 3 and 4 where the specifications model information depth. Once again the econometric analysis supports the inference, based on simple visual inspection of Figures 3 and 8, that the completeness of the information set is not related to when in the sequence of ten choices, a particular choice was presented to the subject. However, the decision duration and total number of cells opened (i.e., acquisition frequency) both decrease as subjects advance through the task. it is noteworthy, though, that the decrease in duration and number of cells opened is actually less in the hypothetical setting than the real setting. This runs counter to what would be expected if subjects took hypothetical choices less seriously.

There is also some evidence to suggest that gender and ability are related to selected information acquisition behaviors. Females tend to spend less time and open fewer cells than males (i.e., the acquire information of less depth), but they do not differ significantly in terms of the breadth of their information sets. Ability (numeracy and CRT combined) appears to be slightly related to the breadth of the information sets acquired by an individual, but not to decision durations or total acquisitions.

TABLE 13. Quantile Regression Information Acquisition in the First Task

Information acquisition measure used as dependent variable is shown at the top of each column.

VARIABLES	(1) Ln(Duration)	(2) # Acquisitions
1(Hypot)	0.341	3.535
	(0.269)	(0.412)
HL Decision number	0.103***	1.265**
	(0.003)	(0.015)
(HL Decision #)×1(Hypot)	-0.034	-0.731
	(0.484)	(0.368)
HL Decision number^2	-0.008***	-0.103**
	(0.007)	(0.026)
(HL Decision #)^2×1(Hypot)	0.003	0.050
	(0.556)	(0.496)
Position number	-0.062**	-0.519***
	(0.000)	(0.000)
Position×1(Hypot)	0.023*	0.312*
	(0.065)	(0.097)
1(Female)	-0.109	-1.968
	(0.291)	(0.166)
1(Female)×1(Hypot)	0.059	1.451
	(0.647)	(0.429)
CRT Score	0.045	-0.081
	(0.431)	(0.928)
CRT×1(Hypot)	-0.067	0.931
	(0.338)	(0.382)
Numeracy Score	-0.008	0.692
	(0.864)	(0.283)
Numeracy×1(Hypot)	-0.055	-0.856
	(0.405)	(0.307)
Constant	9.355***	11.022***
	(0.000)	(0.000)
Observations	970	970

Notes: Bootstrapped standard error estimates clustered on subject.
*** p<0.01, ** p<0.05, * p<0.1

A comparison of Figures 6 and 7 on pages 37 and 38, respectively, indicates that mean and median information acquisition behavior may differ in the case of the information depth (i.e., decision duration and acquisition frequency). In light of this, I use a quantile regression to model decision duration and acquisition frequency as a function of the same set of regressors describe in Equation (2.3). The results of the quantile regression specifications are shown in Table 13.

As it turns out, median information acquisition behavior does not differ significantly from mean behavior in terms of its relationship with the set of regressors in the model. Again, it is evident that individuals increase the amount of time spent and acquire more cells as the difference in expected gambles narrows. Moreover, the median decision duration and acquisition frequency decreased as subjects progressed through the choices, presumably because their proficiency within this choice context increases with practice.

Notably, however, the coefficient estimate on $1(Female)$ is not statistically significant in either specification in Table 13. In contrast, columns (3) and (4) in Table 12 show a statistically significant relationship between the indicator variable for female and decision duration and acquisition frequency. Although the coefficient estimate is negative in all of the specifications, the lack of statistical significance of the estimates in the quantile regressions implies that the statistical significance of the estimates in Table 12 is likely the result of a more heavily skewed distribution of decision duration for females.

Discussion and Conclusion

I conducted an experiment to test whether information acquisition differs across real and hypothetical settings in the context of a commonly-used risk preference elicitation method. Some previous evidence exists to suggests that, on average, individuals respond differently to hypothetical versus real choices about risk. Although the potential payoffs in this experiment were relatively large and the research design increased the amount of subject effort required in the tasks compared to earlier studies of risky choices, there is no evidence of “hypothetical bias” in average risk preferences in this sample.

In both settings, subjects generally acquire sufficient information to make consistent choices. Furthermore, the level of consistency is the same for the real and hypothetical choices.

Additionally, there are no significant differences across settings in the amount of time subjects spend on a choice, or the completeness of the information sets that they choose to acquire. It does appear that subjects spend more time on choices as the difference in the expected values of the alternatives narrows. This result is interesting because in terms of the expected payoff, the cost of making an “error” is lower when the expected values of the alternatives are similar, yet we observe individuals spending more time on these choices rather than less time. Thus, it does not appear that the cost of making an error is an important factor in how much time and effort individuals devote to choices between risky options.

The apparent increase in consideration for choices for which the difference in expected value is relatively small also has implications for the debate about whether individuals actually compute expected value or apply a heuristic. If an individual were to use a heuristic to decide between the risky alternatives in a choice, the time spent on a choice would be independent of the difference in expected values. Conversely, if someone was computing expected values, then we would expect her to spend more time on choices as the alternatives become more alike and that is exactly what we observe here. In the following chapter, I explore the question of whether individuals acquire information in a manner more consistent with expected utility theory or the application of a simplifying heuristic.

CHAPTER III

ELICITING RISK PREFERENCES? INFORMATION ACQUISITION IN A COMMONLY USED METHOD OF ELICITING RISK PREFERENCES

An expected utility (EU) framework underpins most economic theories that seek to explain decision making under risk. Expected utility theory assumes that the utility of a risky set of outcomes is the probability-weighted average of the outcome utilities, and that individuals make decisions by comparing these probability-weighted average utilities, or expected utilities. Observed individual decision making behavior that is inconsistent with fundamental assumptions of EU theory has led many to doubt the generalizability of the theory (e.g., the Allais' paradox (1953)).[1]

Some efforts to account for such "anomalies" in decision making relative to the theory attempt to maintain, but modify, the assumption that outcomes are weighted in some way by the probabilities. Prospect theory, for example, incorporates insights from psychology while maintaining a framework that integrates transformed probabilities and outcomes; in particular, a value function for each possible outcome is multiplied by a decision weight (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Models which maintain the assumption that individuals integrate probabilistic information and outcomes when making decisions can be broadly classified as *integration* models.

In contrast to integration models, another class of models hypothesizes that the complexity of many decisions leads individuals to resort to "rules of thumb" or

[1]See Camerer (1995) for a catalog of behavior inconsistent with expected utility. The Allais paradox is a violation of the independence axiom of expected utility theory that occurs as a result of the certainty effect. The certainty effect is a "preference for security" in the neighborhood of certainty (Allais, 2008). Also see Kahneman and Tversky (1979) for a discussion of the certainty effect and the Allais paradox.

heuristics to simplify the decision process (Newell and Simon, 1972; Kahneman and Tversky, 1982; Payne et al., 1993; Gigerenzer et al., 1999; Gigerenzer and Gaissmaier, 2011). Gigerenzer and Gaissmaier (2011) define heuristics as "strategies that ignore information to make decisions faster, more frugally, and/or more accurately than more complex methods" (p. 454). Some common heuristics discussed in the decision-making literature include: the recognition heuristic, tallying rules, take-the-best, and fast-and-frugal trees. Economists, however, may be more familiar with the heuristics highlighted in Kahneman and Tversky (1982): representativeness, availability, and anchoring.[2]

Despite the conflicting theories about how individuals actually process information when making a decision, economists have made surprisingly few attempts to explore whether individuals appear to be processing information in a manner consistent with EU theory, or any integration theory for that matter. As noted by Willemsen and Johnson (2010), decision making research has tended to progress by focusing on choice outcomes rather than the process by which people arrive at decisions. Recently, however, economists have exploited technologies that make it possible to observe individual information acquisition behavior to test theories of decision making (Costa-Gomes et al., 2001; Gabaix et al., 2006; Johnson et al., 2008; Arieli et al., 2011), and a few studies have explored whether individuals acquire information in a manner that is more consistent with an integrative decision process or the application of a heuristic (Johnson et al., 2008; Arieli et al., 2011; Glöckner and Herbold, 2011).

Previous studies that examine whether individuals acquire information as if they were integrating information or applying a heuristic have reached contradictory conclusions. Johnson et al. (2008) and Glöckner and Herbold (2011) both find

[2]See Gigerenzer and Gaissmaier (2011) for a discussion of the first set of heuristics as well as a summary of the evolution of research focused on heuristics. See Kahneman and Tversky (1982) for a summary of the latter set of heuristics.

information acquisition patterns consistent with integration models, while Arieli et al. (2011) do not. The conflicting results do not appear to be caused by differences in how information acquisition was observed. Glöckner and Herbold (2011) and Arieli et al. (2011) use eye-tracking technology, while Johnson et al. (2008) uses process-tracing software called Mouselab, which is the method used in the present study and is described below. Of these studies, Arieli et al. (2011) is the only study that examined information acquisition using gambles that had only one positive payoff (i.e., one of the possible outcomes in both options was zero, so these outcomes and their corresponding probabilities were not shown). Thus, each gamble required that subjects acquire only two pieces of information rather than the four pieces of information presented to subjects in Johnson et al. (2008), Glöckner and Herbold (2011), and the present study.

My examination of the decision making process contributes to the literature in several ways. First, I observe information acquisition behavior in both real and hypothetical settings, whereas previous studies use only hypothetical payoffs. Second, I observe information acquisition within a commonly-used method of eliciting risk preferences that entails many similarly structured choices. Thus, I am able to examine the evolution of information acquisition as subjects progress through a standard method of eliciting risk preferences. Third, I avoid a framing effect that may have biased Johnson et al. (2008) toward finding information acquisition behavior consistent with integration models of decision making. In their study, each gamble presented to a subject in a choice was entirely contained in a single row or column. Thus, simple effort-minimizing row (or column) scans, from left to right, or right to left, would be confounded with an integrative search pattern in their study. In contrast, in this experiment, gambles were presented to subjects so that the probabilities are either directly above or directly below their associated payoffs, while alternatives are presented side-by-side.

Consistent with Johnson et al. (2008) and Glöckner and Herbold (2011), I find evidence to suggest that subjects acquire information in a manner consistent with the implicit calculation of expected utility. Specifically, individuals do not merely make choices "as if" they are integrating probabilities and outcomes, it appears that they actually are. Moreover, as they progress through a series of choices in a commonly used risk preference elicitation method, their information acquisition becomes progressively more consistent with integration models. Finally, on average, individuals appear to acquire information in real and hypothetical settings in similar ways. This insight is important to the ongoing debate about whether hypothetical choices can be used to elicit real preferences.

The Method

In the decision research literature, observing individual information acquisition behavior is known as "process tracing." Process tracing can be accomplished in a variety of ways, but the varying techniques have the common objective of providing a window (for the researcher) into the cognitive process that an individual uses to reach a decision. Early forms included manual techniques, such as the acquisition of information cards (Bettman and Jacoby, 1976; Payne, 1976). Other methods of process tracing include eye-tracking, verbal protocols, and computer-based programs.[3]

I use a software program called Mouselab to observe subjects' information acquisition behavior. In a way, Mouselab is a lower-tech, more expedient alternative to eye-tracking technology and provides analogous measures of attention, such as the

[3] Ford et al. (1989) reviews 45 experiments that use verbal protocols or "information board" techniques. See Cokely and Kelley (2009) for a recent example of a verbal protocol experiment. Verbal protocol experiments, also called "think aloud" protocols, have the subjects verbalize their decision-making process. Information board techniques were more common before the computer revolution in the 1980s and required subjects to go to a board and turn over cards with information about the alternatives being considered in the choice scenario.

frequency of fixations on an attribute, duration of fixations on an attribute, transitions between areas of interest (AOIs), and the first fixation.[4] Mouselab software has been used by psychologists as a lower-cost alternative to eye-tracking for more than two decades (Payne et al., 1988), and several recent studies by economists have used it to explore information acquisition and decision making behavior (Costa-Gomes et al., 2001; Gabaix et al., 2006).[5]

Mouselab works by allowing the researcher to "mask" selected information so that a subject must move the computer pointer over the cell with the information so that it will be revealed. Once the mouse pointer leaves the cell, the information is hidden again. By requiring a subject to unmask each cell and allowing them to view only one piece of information at a time, I am able to observe how individuals acquire information about risky choices in this stylized context.[6]

The Task

Data on subjects' information acquisition behavior was collected in the context of an experiment in which they were asked to choose between two lotteries. The

[4]"Fixations are moments during which the eye is relatively still, typically lasting around 200-500 milliseconds" (Rayner, 1998). A reaquisition occurs when the eyes are refocused on a piece of information after shifting their gaze somewhere else.

[5]Several studies have compared Mouselab with eye-tracking technology. Reisen et al. (2008) "found only minor differences between" eye-tracking technology and Mouselab in terms of their abilities to describe information search and strategies. However, in a comparison of process tracing data generated from eye-tracking and Mouselab, Lohse and Johnson (1996) find that "the method of recording information acquisition influenced the decision process." For example, subjects using eye-tracking equipment used less time, made more fixations, and made more reacquisitions than the subjects using Mouselab. Franco-Watkins and Johnson (2011) find similar results. This differences are unsurprising given the greater effort required to move a mouse relative to merely shifting one's gaze. However, there does not appear to be evidence to indicate that subjects alter the pattern of their acquisitions. In other words, a subject might make fewer acquisitions but does not acquire *different* information, or acquire the information in a different order, when using a mouse.

[6]Subjects were required only to "mouseover" a cell to reveal it and were not required to "click" on the cell. Requiring a click is a design feature that could have been employed but I believe that it would increase the differences between our design and an eye-tracking design without an obvious benefit.

lotteries presented to subjects were a modified version of the Multiple Price List (MPL) introduced by Holt and Laury (2002), which asks subjects to make ten choices between a *safe* lottery and a *risky* lottery. Both lotteries are characterized by uncertainty and have a high and low payoff, but the risky lottery has a greater variance in its payoffs. Table 1 illustrates the conventionally ordered MPL format. Subjects were provided only with the probability and payoff information in Columns (1) through (4).

Several modifications were made to the conventional MPL so that valid information acquisition behavior could be collected. First of all, the probabilities and the payoffs of the lotteries were masked, so subjects were required to unmask the attributes of the lotteries. Second, the order in which the choices were presented to subjects was randomized. This "shuffling" of the order of the MPL means that the *position* in which a subject completed a choice almost certainly differs from the choice order in the Holt and Laury MPL. Hence, for each subject a particular choice has its HL Decision Number (i.e., where that choice sits in the conventional MPL), and its Position Number (i.e., the order in which a subject actually completed the choice in the task).

The third modification is that small perturbations were made to the probabilities and payoffs to prevent subjects from easily recognizing the patterns in the MPL.[7] Fourth, the presentation of the lotteries was randomized along three dimensions to address framing effects: half of the subjects were presented with the probabilities in the top row, half saw the risky gamble on the left, and half saw the high payoff outcome on the left within each gamble. An example of how each pair of gambles was presented to subjects is shown in Figure 1.[8]

[7]For details of the MPL and the modifications should see Taylor (2011).

[8]The experiment was presented in color.

Search Patterns and Decision Making

To explore whether information is consistent with EU theory, I focus primarily on the *transitions* that subjects make between boxes in the process of acquiring information about the choices.[9] A brief comment on notation is in order before going further. I adopt the notation used byJohnson et al. (2008) and denote the payoffs in the following way, W_x^y, where W denotes payoffs, $x = safe, risky$ and $y = min, max$. Probabilities are labeled P_x^y, where P denotes probability. Hence, for example, W_{safe}^{min} denotes the minimum payoff in the safe option and P_{risky}^{max} denotes the probability of the maximum payoff in the risky option.

Transitions are classified into six types: within-gamble transitions, across-gamble transitions, same-cell transitions, within-outcome transitions, across-outcome transitions, and same-attribute transitions. *Within-gamble* transitions are transitions from one cell to another cell within the same gamble. This includes all transitions between cells with the same subscript (e.g., P_{safe}^{max} to P_{safe}^{min}). A transition from a cell in one gamble to a cell in the other gamble (from the safe gamble to the risky gamble or vice versa) is classified as an *across-gamble* transition. There are also rare occurrences when a subject moves the mouse out of a cell and then immediately back into the same cell—*same-cell* transitions. All possible transitions can be classified into one of these three transition types.

Within-gamble transitions can be delineated further into two distinct types: within-outcome transitions and across-outcome transitions. *Within-outcome* transitions are transitions between a probability and its corresponding payoff, or a payoff and its corresponding probability. The subscript and the superscript for transitions of

[9] Those interested in the details about what information was acquired, how long subjects spent on choices, and other measures of the breadth and depth of information acquisition can see Taylor (2011).

this type are the same— P^{max}_{safe} to W^{max}_{safe}, for instance. Transitions that occur within a gamble, but cannot be classified as within-outcome transitions (e.g., a transition from P^{max}_{safe} to W^{min}_{safe}), are referred to as *across-outcome* transitions for simplicity, although in a strict sense any transition that is not within-outcome could be classified as such.

The last type of transition occurs when a subject moves the mouse from an attribute in one gamble to the same attribute in the other gamble: *same-attribute* transitions. A transition from W^{max}_{safe} to W^{max}_{risky} is one potential same-attribute transition and this type of transition is the most consistent with a subject applying a heuristic that requires the comparison of attributes across gambles. For example, a subject using a "take-the-best" heuristic with the highest potential payoff criterion would be expected to transition from the cell with the high payoff in one gamble directly to the cell that contains the high payoff in the other gamble.

Interdimensional versus Intradimensional Nomenclature

Early studies that explored decision making via process-tracing data classified a transition by whether it is *interdimensional* or *intradimensional* (Payne, 1976; Payne and Braunstein, 1978; Rosen and Rosenkoetter, 1976). Interdimensional strategies are manifested in search patterns characterized by within-alternative transitions between the different "dimensions" of a particular alternative (i.e., within-outcome and within-gamble transitions) suggesting that a subject is integrating the pieces of information for that alternative. Intradimensional strategies are characterized by transitions across alternatives within the same attribute and are more consistent with the application of a heuristic because such transitions suggest that a subject is comparing alternatives along particular dimensions rather than by integrating information.[10]

[10]Arieli et al. (2011) adopt a different nomenclature. They refer to *holistic procedures* (interdimensional) and *component procedures* (intradimensional).

A within-outcome transition is the most narrowly defined type of interdimensional transition. In other words, this is the type of transition that is most consistent with a subject computing expected value. More broadly, however, all within-gamble transitions, both within-outcome and across-outcome, could be termed interdimensional transitions. In a strict sense, only same-attribute transitions are truly intradimensional and they are the least consistent with integration models. However, I classify any across-gamble transition as an intradimensional transition.

Occurrence and Adjacency

Two properties must hold for information acquisition behavior to be linked to decision making (Costa-Gomes et al., 2001). The first property is *occurrence*: information must be acquired before it can be used by a decision maker. The second property, *adjacency*, is that information that will be used together should be acquired in close proximity in a sequence of unmaskings (for example, a probability and a payoff in the computation of expected value). Adjacency assumes that it less costly to acquire information as needed than to memorize it.

In the current context, occurrence implies that if a subject is observed making decisions without acquiring information about the probabilities, then he could not be using that information to make decisions. This property is not controversial. Adjacency implies that when a subject transitions between a payoff and its associated probability, in either direction, then it is more likely that the product of the payoff and its associated probability, an expected value ingredient, is being used in the decision making process. Unless subjects are memorizing attributes so that the order of acquisition is irrelevant (i.e. if adjacency does not hold), search patterns characterized by intradimensional transitions suggest that subjects are applying heuristics.

If the occurrence and adjacency properties hold, then one straightforward way to test whether individuals are integrating information about the options presented to them, or applying a heuristic, is to check whether they tend to make predominantly within-outcome transitions or same-attribute transitions.

Heuristics

Individual decision making behavior can vary widely depending on the task and context, and there is no one dominant strategy applicable to every type of decision (Payne et al., 1993). The choices presented to subjects in this experiment are sufficiently complex that it is plausible that they could employ a heuristic when making their decisions. Gigerenzer and Gaissmaier (2011) stress that a rule or heuristic must be "ecologically rational," meaning that the heuristic is, or can be, adapted to the structure of the environment in which it is being used. Lexicographic rules and tallying, for example, can be adapted easily to choices between two gambles, such as those used in this experiment.

Lexicographic rules rank the *reasons* that an individual can use to decide between two alternatives, and decisions are then based on one reason only. In other words, in the two-alternative case, an individual using a lexicographic rule will proceed through a list of m different reasons and then make her choice based on the first reason for which one alternative has an advantage. In contrast, tallying rules do not impose an order on the reasons and an individual simply picks the option that is supported by the most reasons. Prior empirical evidence, as well as the results presented below, suggests that individuals do not use tallying rules when choosing between gambles (Brandstatter et al., 2006; Johnson et al., 2008). In particular, for the choices in this experiment, at least four same-attribute transitions are required if a subject is using a tallying rule and, as discussed below, subjects do not make anywhere near this many acquisitions. Thus, it is reasonable to focus on just the class of lexicographic heuristics.

A variety of one-reason lexicographic rules are readily adaptable to choices between two gambles such as those presented to subjects in this experiment. A risk-averse subject who chooses her preferred gamble based solely on the largest minimum payoff (maximin strategy) provides an example of a take-the-best heuristic being employed. Alternatively, she could choose the gamble with the highest potential payoff, highest probability, and so forth.

Reading Phase

In the heuristics literature, it is typically assumed that there is an initial reading "phase," in which subjects explore all the information so that they can identify the relevant pieces of information (Brandstatter et al., 2006). It is seems to be an empirical question, however, whether individuals actually delineate their searches into separate phases because expected utility theory does not require a reading phase. The most appropriate way to test whether subjects divide their searches into reading and choice phases is not entirely obvious. One reasonable check, though, is whether subjects acquire all the available information in as few acquisitions as possible.[11]

Johnson et al. (2008, p. 266) define the reading phase, in a straightforward way, as "all acquisitions made before all outcomes have been examined at least once." Thus, if subjects divided their searches into a reading phase and a choice phase, we might expect that the number of acquisitions that they make to acquire all eight pieces of

[11] Another possible check is the proportion of total acquisitions that occur before a subject acquires all the information. A small proportion could imply that a subject first scanned through all the information and then completed a more thorough evaluation in the choice phase. However, a large proportion does not rule out the possibility that subjects separated their searches into two phases. In fact, the median proportion of acquisitions made before a subject acquired all the information is seventy-five percent.

information would be very little in excess of eight. In fact, of the choices in which a subject viewed all the information, the median acquisition number at which all the information was viewed is ten.[12] This seems consistent with the idea of a reading phase because it implies that subjects did not do a lot of preliminary exploring before they eventually acquired all the available information. Since it is possible that subjects are dividing their searches into two phases, when appropriate, I examine information acquisition in the reading phase and choice phase separately.

Priority Heuristic

A more sophisticated heuristic is the *priority heuristic* , which Brandstatter et al. (2006) demonstrate has remarkable predictive ability and can account for decision anomalies such as the Allais paradox, the certainty effect, and the four-fold pattern of risk preferences.[13] The priority heuristic is a lexicographic heuristic that *prioritizes* the order of the reasons individuals use by incorporating previous empirical findings which suggest that payoffs (minimum payoffs in particular) matter more to subjects than probabilities, and psychological concepts, such as bounded rationality (Gigerenzer and Gaissmaier, 2011). Specifically, the priority heuristic hypothesizes that risk aversion will lead individuals to devote the most attention to minimum payoffs.

The priority heuristic is richer than the simple maximization of the minimum payoff, though. It also allows for the possibility that decisions are affected by other

[12] The sample for this statistic includes the 805 of 970 choices for which subjects viewed all eight pieces of information in the first task. Although the complementary nature of the probabilities requires that a subject unmask only one probability in each gamble, there are only 18 instances in which a subject acquired just one probability in each gamble and all four payoffs and did not also acquire the remaining probabilities. In other words, if a subject acquired the six pieces of information necessary to compute expected value, then they generally acquired all eight pieces of information.

[13] The four-fold pattern is: (a) risk aversion for gains if probabilities are high, (b) risk seeking for gains if probabilities are low, (c) risk aversion for losses if probabilities are low, and (d) risk seeking for losses if probabilities are high. Harbaugh et al. (2002b) and Harbaugh et al. (2002a) explore the robustness of the four-fold pattern.

potential payoffs in a choice. In particular, the priority heuristic formalizes the notion that the highest potential payoff in a choice will determine an individual's *aspiration level*, and that the aspiration level can affect individual preferences. The notion of an aspiration level is similar to Simon's theory of satisficing (Simon, 1983), which theorizes that bounded rationality will lead individuals to stop searching once an alternative meets a particular threshold. The priority heuristic posits that the aspiration level is 1/10 of the largest potential payoff and that a comparison of the aspiration level to the difference in the minimum payoffs is the first reason that individuals use to choose between alternatives.[14] In the present experiment, the size of the largest payoff ranges from seventy-three to eighty-one dollars, so the aspiration level would be either seven or eight dollars.

Although the priority heuristic has additional reasons (or decision rules), they are not relevant here because a subject using the priority heuristic in this experiment would stop at the first reason because all the choices are structured such that the difference in the minimum payoffs is always greater than the aspiration level.[15] Even in the case of one-reason choices, however, Johnson et al. (2008) highlight that the priority heuristic generates several testable hypotheses.

First, payoffs should receive the most attention. Specifically, the priority heuristic's hypothesis that individuals compare minimum payoffs directly implies that minimum payoffs should receive considerable attention and that there should be transitions between the minimum payoff in the safe option and the minimum payoff in the risky option. In other words, we should observe subjects directing their attention towards W^{min}_{safe} and W^{min}_{risky}, as well as making transitions between the two boxes.

[14]See Brandstatter et al. (2006) for an explanation of why 1/10 was chosen.

[15]Interested readers should see Brandstatter et al. (2006); Johnson et al. (2008) for further discussion of the stopping rules.

Second, probability-payoff transitions (within-outcome transitions) should be infrequent, since use of the priority heuristic implies that these transitions are unnecessary.

Third, during the *reading* phase, within-gamble transitions should be primarily between payoffs. This is because subjects are attempting to identify the smallest and largest payoffs so that the gambles can be compared using the reasons of the heuristic.

The Experiment

The following is a brief summary of the experiment. Please see Chapter II for full details of the experimental design.

Twenty-seven experimental sessions were conducted during March and April of 2011. A total of 98 people, recruited primarily from undergraduate chemistry, economics, and environmental studies courses at the University of Oregon, participated in the experiment. Subjects earned an average total payoff of \$52.68, with a maximum of \$86.00 and a minimum of \$6.90. Nearly all subjects finished within 40 minutes.

The experiment was completely computer-based and each subject's interaction with experimenters was minimal.[16] A typical session included about four (3.6) subjects but never more than six, and subjects were isolated by dividers.

All subjects completed two sets of ten choices. In addition to a \$5.00 show-up payment, subjects were paid based on the outcome of one randomly selected choice in one task (the real setting), but were not paid for the randomly selected choice in the other task (the hypothetical setting). The setting (i.e., real or hypothetical) under which a subject completed her first task was randomly assigned. Subjects were not aware that they would complete the task in both real and hypothetical conditions,

[16]The interaction between a subject and the experimenters was limited to the experimenters signing in the subject, reading a brief set of general instructions, verifying the subject's payment, and paying the subject.

but they were made aware of how their choices would affect their payoff (or not) immediately prior to the beginning of each task.

Once subjects completed both tasks and the payoffs were revealed, they completed an ability test that measures both numeracy and cognitive ability. Numeracy tests are designed to measure an individual's ability to understand and manipulate numeric and probabilistic information. The potential computational complexity of our tasks, along with the impact that numeracy could have on information acquisition behavior, made it prudent for us to elicit such a measure of numeracy. I adapted an eight-item test developed by Weller et al. (2011) that includes the three-item cognitive reflective test (CRT) introduced by Frederick (2005). Frederick designed the CRT to evaluate an individual's "System 2" cognitive processes — the ability to solve problems that require "effort, motivation, concentration, and the execution of learned rules" (2005, p. 26). In contrast, "System 1" processes are spontaneous and do not require significant effort.[17] Thus, I have separate measures for a subject's numeracy and his cognitive ability.[18]

After dropping one subject because she did not unmask a single attribute in seven of the ten real choices, our sample includes a total of 97 subjects. Forty-eight subjects completed the real choices in the first task; the remaining forty-nine completed the hypothetical choices first. Based on two-sided t-tests, if we partition the sample by whether a subject completed the real or hypothetical choices first, Table 2 shows that the samples are not statistically different on dimensions commonly found to be related to risk attitudes, such as gender and cognitive ability.

[17]See Frederick (2005) and Stanovich and West (2000) for discussions of System 1 and System 2 differences. Daniel Kahneman's *Thinking, Fast and Slow* highlights the differences in System 1 (Fast) and System 2 (Slow) cognitive processes (Kahneman, 2011).

[18]The complete test we use is included in Appendix A.

Results

Reading and Choice Phase Combined

In the 970 choices completed in the first task, the eight cells were acquired an average of 16.2 times per choice. This means that subjects made repeated visits to the same information. Correspondingly, subjects averaged 15.2 transitions in the first task. These measures decreased in the second task to 12.7 and 11.7, respectively. The distributions of acquisition frequencies and the transition frequencies are right-skewed. The median transition frequency was thirteen in the first task and ten in the second task.

More than eighty percent of all transitions that subjects made can be categorized as within-gamble transitions. Figure 9 shows the median transition frequency for within-gamble and across-gamble transitions in each choice. Same-cell transitions account for only about 1 percent of all transitions, making them sufficiently infrequent that they are ignored in the remainder of the paper.

The median within-gamble transition frequency in the first task was eleven, while the median across-gamble transition frequency was two. A Wilcoxon signed-rank test rejects the null hypothesis that the median transition frequencies are equal ($p < 0.001$). Such a large proportion of within-gamble transitions is consistent with an integrative decision making process. Moreover, the percentage of within-gamble transitions does not differ by setting (83.7 and 83.6 percent in the real and hypothetical settings, respectively), and differs only slightly by task (84.1 and 83.2 percent in the first and second task, respectively ($p=0.043$, for t-test of equal means)).

Across-gamble transitions account for approximately fifteen percent of total transitions, which means that the average subject transitioned across gambles only twice while making a choice in the first task. Additionally, there is not a statistically

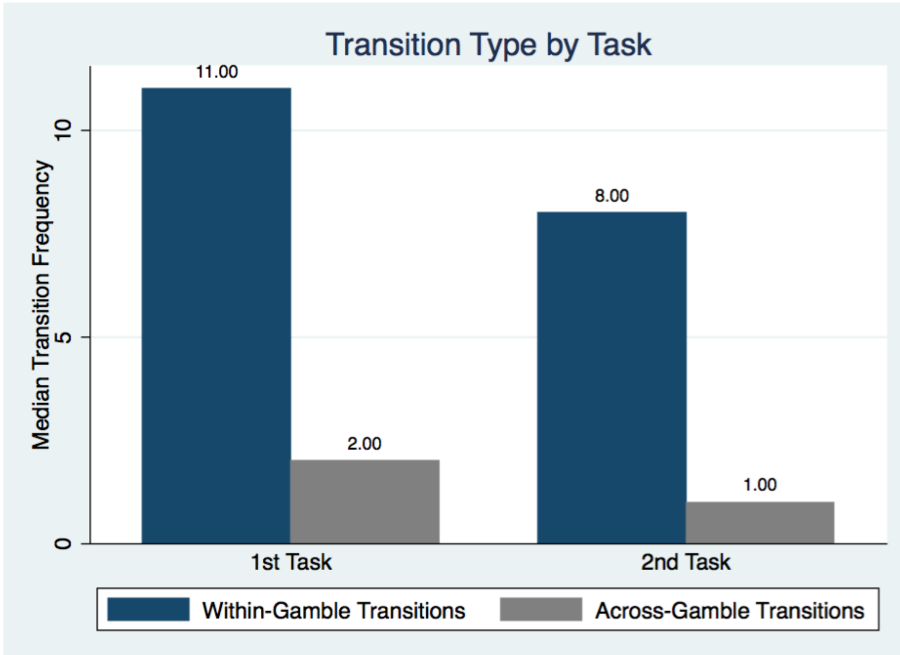


FIGURE 9. Median Transition Frequency by Type in Each Task.

significant difference in this percentage across settings in either task (p-value > 0.211). The median number of across-gamble transitions decreased in the second task, indicating that once subjects became familiar with the task they made very few intradimensional comparisons.

As discussed above, the use of a heuristic will generally require that a subject make same-attribute transitions and that within-outcome transitions will be infrequent. Figure 10 shows the median transition frequency of within-outcome, across-outcome, and same-attribute transitions in each task. Only about four percent of transitions are same-attribute transitions and the median is zero, so the third "bar" in each group in Figure 3 has height zero.

In contrast, the median within-outcome transition frequency is five. More than forty percent of the total transitions made by subjects in both tasks are within-outcome transitions (between a probability and its corresponding payoff or vice-versa) and the percentage of within-outcome transitions increased in the second task.

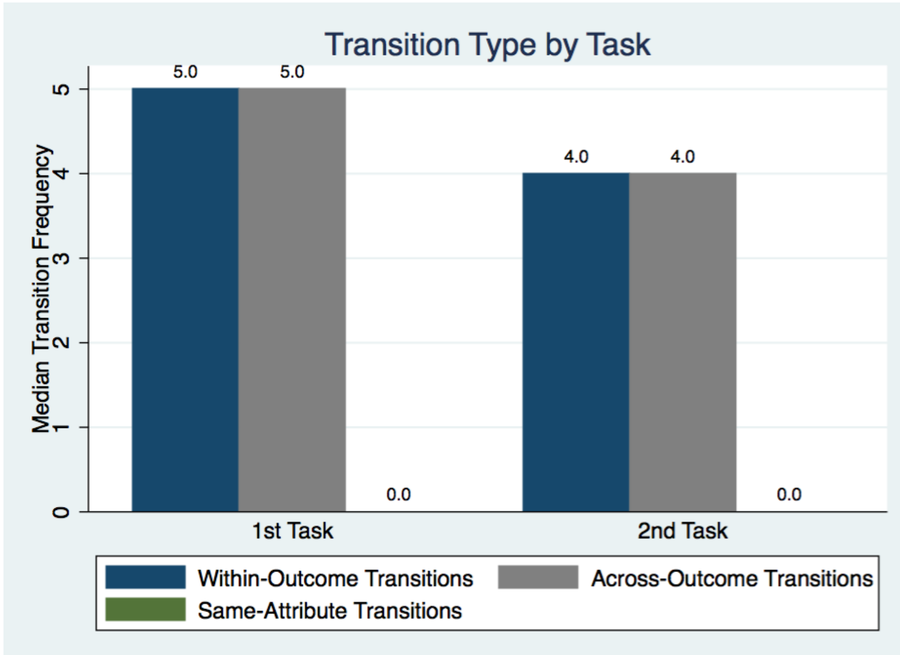


FIGURE 10. Median Transition Frequency by Type in Each Task.

It is helpful to put into context the percentage of within-outcome transitions. Consider a subject who unmasks every cell once and only once. The highest percentage that within-outcome transitions can account for in such a scenario is fifty-seven percent (four divided by seven transitions); the lowest is zero (a subject could unmask all the information without making a single within-gamble transition). In the case where a subject that minimizes the distance for each transition, Brandstatter et al. (2008) hypothesize that the percentage is zero for one-reason choices in the choice phase. Clearly, this percentage is much greater than zero when we examine the data without separating it into phases. In the next section, I show that it is greater than zero in the choice phase as well.

Reading Phase Transitions

Since the reading phase is an exploratory phase, there are few predictions about how subjects will acquire information in this phase even if they are using a particular decision heuristic because this is the phase in which subjects are presumed to be simply gathering information. However, as noted above, Johnson et al. (2008) assert that the priority heuristic would predict that subjects would make within-gamble transitions between payoffs so that the payoffs can be compared. Additionally, it is informative to examine whether information acquisition in the racing phase is significantly different from information acquisition for the choice as a whole.

In the reading phase, the median transition frequency in the first task was nine. The median within-gamble transition frequency was eight in the first task. Information acquisition behavior in the second task followed a similar pattern: the median transition frequency was eight and the median within-gamble transition frequency was seven. Thus, in both task, information acquisition patterns suggests that subjects gathered information about each gamble individually in the reading phase.

Figure 11 shows how information was acquired in the reading phase did not differ significantly from information acquisition in the choice as a whole. Within-gamble transitions were distributed evenly between within-outcome transitions and across-outcome transitions and same-attribute transitions were extremely rare, with the median being zero.

Did subjects appear to be comparing within-gamble payoffs in the reading phase as the priority heuristic predicts? The median transition frequency for within-gamble payoff transitions was two. Thus, these type of transitions were certainly not the primary type of transitions individuals made during the reading phase. However, given that subjects acquired all the information about the choices with relatively few transitions, the fact that only the median within-gamble payoff transitions is zero is not convincing evidence against the priority heuristic either because only two transitions of this type are required to identify the largest and smallest payoff in each

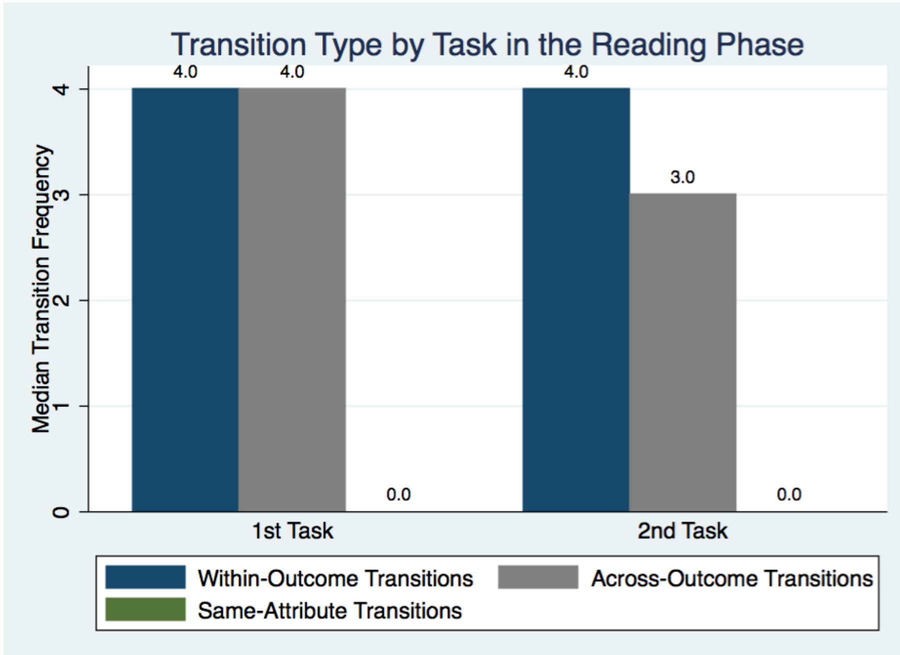


FIGURE 11. Median Transition Frequency by Type in Each Task in the Reading Phase.

gamble. It is worth noting that the median frequency for transitions of this type decreased to one in the second task.

Choice Phase Transitions

In the choice phase, the median transition frequency in the first task was five for subjects who acquired all the information and made at least one additional transition. It decreased slightly in the second task to four.[19]

Subjects' information acquisition behavior does not appear to differ significantly in the choice phase from the reading phase. Figure 12 shows the median transition frequency for within-outcome, across-outcome, and same-attribute transitions in the choice phase only. Again, the data appear to show an information acquisition process

[19]Transitions were non-zero for 655 of the 805 choices for which subjects acquired all eight pieces of information in the first task. In the second task, transitions were non-zero for 492 of the 656 choices for which all the information was acquired.

consistent with integration models rather than a heuristic. The median same-attribute transition frequency was zero and the median within-gamble transition frequency was four (two within-outcome and two across-outcome). Thus, even when the decision process is divided into two phases, the choice phase search process involved primarily within-gamble transitions.

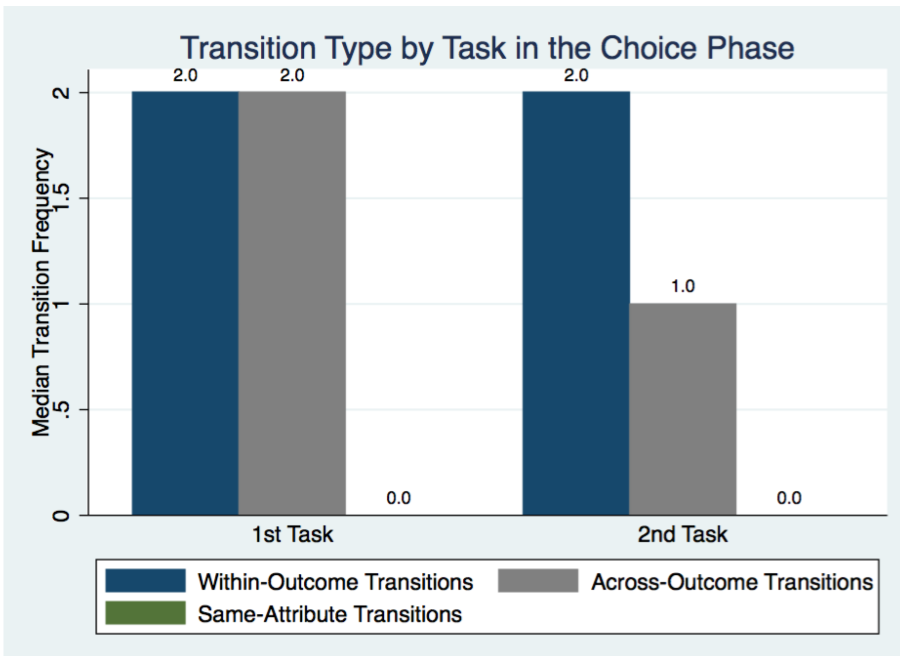


FIGURE 12. Median Transition Frequency by Type in Each Task in the Choice Phase.

Given that the median frequency of same-attribute transitions is zero, the median frequency of transitions between the minimum payoffs is necessarily zero. In fact, of the 655 choices in which a subject made at least one transition in the choice phase there are only thirty instances of a transition across gambles between minimum payoffs in the first task. In the second task, there were only eighteen transitions of this type.

If I focus on just the first choice that subjects made, because that is the one to which they devote the most time and make the most acquisitions, the results are not significantly different. The median transition frequency between minimum payoffs is zero, and there are only two instances of a subject making such a transition.

Regions of Good and Bad Performance

Thus far, the evidence presented is rather clear that subjects were not using a intradimensional heuristic for these choices. However, are these the type of choices for which people would resort to a heuristic? In response to criticism that there are choice sets for which the priority heuristic cannot predict well, Brandstatter et al. (2008) point that the same can be said for every model of risky choice. This raises the question of whether the choices presented to the subjects in this experiment are well-suited for the use of a heuristic. For example, Brandstatter et al. (2008) show that the priority heuristic more accurately predicts subjects' choices than expected utility theory and cumulative prospect theory when the ratio between the gambles' expected values is less than or equal to two. The situation is reversed, however, when the ratio becomes large and the choices are "easy."[20]

The ratio of expected values in this experiment range from 1.1 to 3.45 (gambles with identical expected values have a ratio of one), but the ratio is less than two for eight of the nine choices that contain uncertainty. If we focus on just the most "difficult" choices, namely, the two choices for which the ratio of expected values is approximately 1.1, HL Decision numbers 4 and 5, the results are unchanged.[21] Figure 13 shows that the median same-attribute transition frequency was zero for these decisions; the median within-gamble transition frequency was four.

[20]Focusing on the ratio of expected values ignores absolute differences between options, which could be extremely salient when considering gambles with very large stakes. However, given that the stakes are not in the hundreds of thousands of dollars in this experiment, it seems reasonable to use the ratio of expected values here.

[21]Table 1 shows that these two choices also have the smallest difference in expected value.

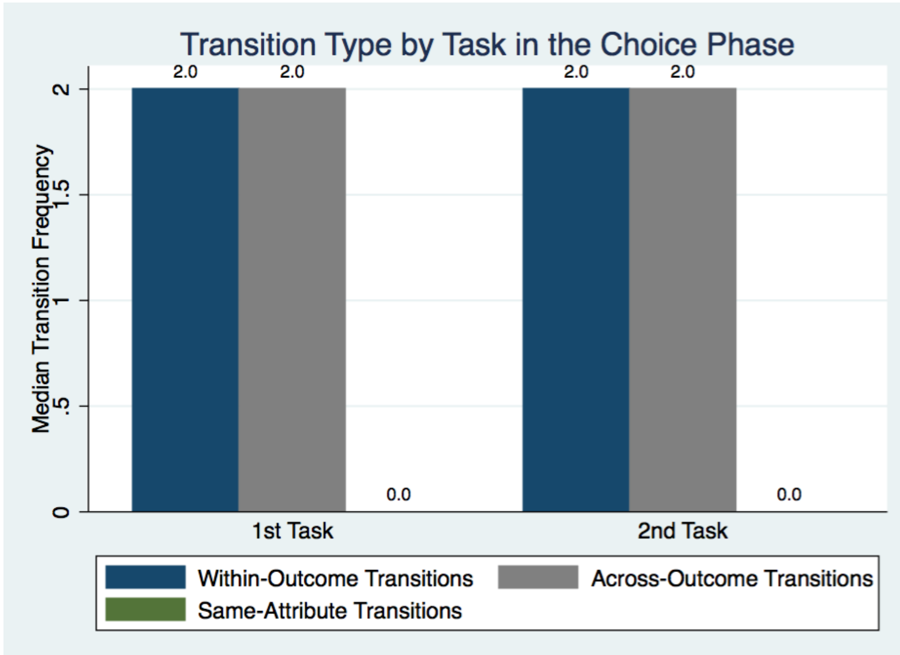


FIGURE 13. Median Transition Frequency by Type in Each Task in the Choice Phase For Choices with a Low Ratio of Expected Value (HL Decision Numbers 4 and 5).

Another possible explanation for why subjects do not appear to be using a intradimensional heuristic is because the probabilities do not differ substantially across the gambles. In cases where all the probabilities are the same, Brandstatter et al. (2008) propose that individuals may use the *toting-up heuristic*. This heuristic is employed when a subject encounters a choice for which all the probabilities are the same, so they "add up the outcomes of each gamble and select the gamble with the higher sum" (Brandstatter et al. (2008), p. 284).

The evidence here does not support the hypothesis that subjects were using such a toting-up heuristic. Even if we allow for any transition from one probability to any other probability to be considered as evidence in support of such a rule, the median transition frequency of this type is equal to one during the choice phase. Three transitions are necessary to sum up the payoffs, however, and there were no occasions in which a subject did a simple row search in the choice phase, let alone a row search of the payoffs.

Overall, the evidence presented here is generally more supportive of the hypothesis that individuals are integrating information in a manner consistent with expected utility rather than a heuristic. In fact, same-attribute transitions are relatively rare and within-gamble transitions make up an overwhelming majority of transitions between cells. Furthermore, within-outcome transitions, the type of transition most supportive of the hypothesis that a subject is computing expected utility, account for more than forty percent of total transitions. Across-outcome transitions, which are also within-gamble transitions, account for at least another forty percent of total transitions. Meanwhile, same-attribute transitions—the transition type most consistent with an individual applying a heuristic—account for less than four percent of total transitions.[22]

Evolution of Transition Behavior

Tracking subject information acquisition using the Holt and Laury MPL allows us to observe how subjects' information acquisition behavior evolves as they make many similar choices. As subjects progressed through choices in the first task, they tended to make fewer total transitions for each choice. This suggests that as subjects learned about the arrangement of the information in the choice scenarios they optimized their search strategies. Figure 14 shows the decline in median transition frequency by position number of the choice. The median transition frequency on the first choice is

[22]It is possible that the spatial separation of the gambles may have biased subjects toward using within-gamble transitions to acquire information. In other words, because it is more "work" to make a same-attribute transition relative to a within-gamble transition subjects could, conceivably, make fewer same-attribute transitions. However, it is not obvious how the information can be presented to subjects in this type of choice context in a way that will avoid this bias but does not also increase the difficulty of the choices with an unintuitive arrangement of the information.

fourteen. The median for the last choice is ten. Time spent on the choices follows a similar pattern as subjects progressed through the set of choices in the first task.

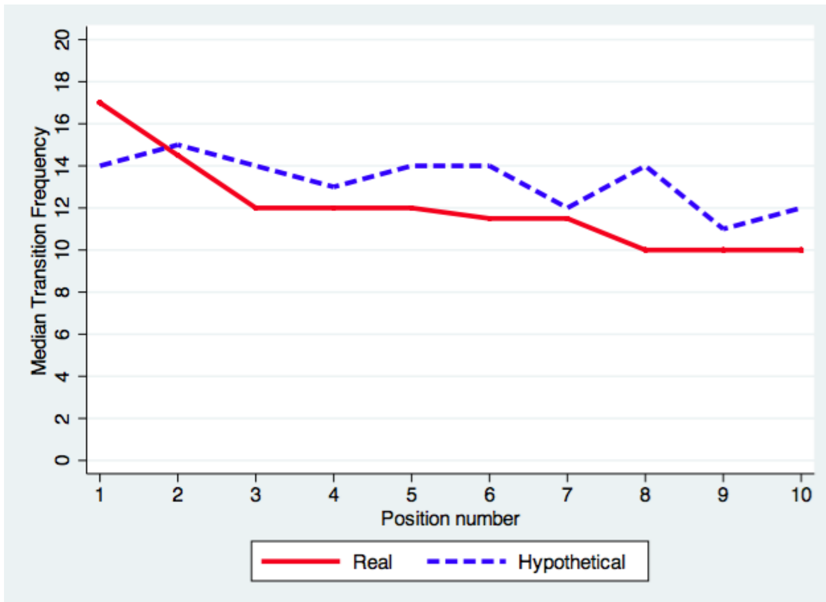


FIGURE 14. Median total transition frequency in real and hypothetical setting in the first task. Position number indicates the order in which the choice was made.

The percentage of transitions that can be classified as within-gamble or across-gamble are plotted in Figure 15. The figure is consistent with the results in Figure 2, and indicates that, although subjects made fewer transitions as they progressed through the task, they primarily explored each gamble separately throughout the task and did not dramatically change the percentage of within-gamble or across-gamble transitions as they progressed through the first task (Recall that these two types of transitions effectively account for all transitions.).

Interestingly, when within-gamble transitions are separated into within-outcome and across-outcome transitions, we see that the type of transition most consistent with integration models, within-outcome transitions, increases on average as a percentage of total transitions. The graph on the left of Figure 16 shows that the percentage

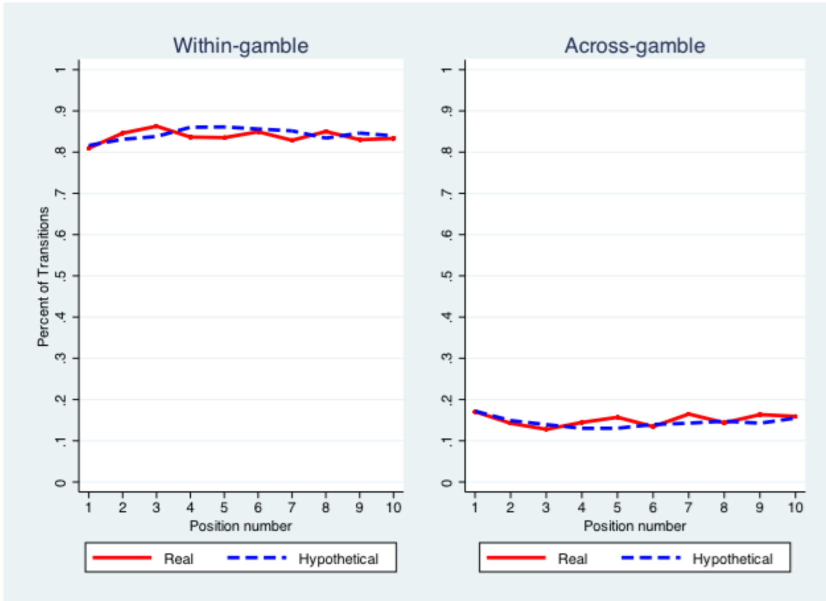


FIGURE 15. Percent of transitions accounted for by the specified type of transition type in real and hypothetical setting. Position number indicates the order in which the choice was made.

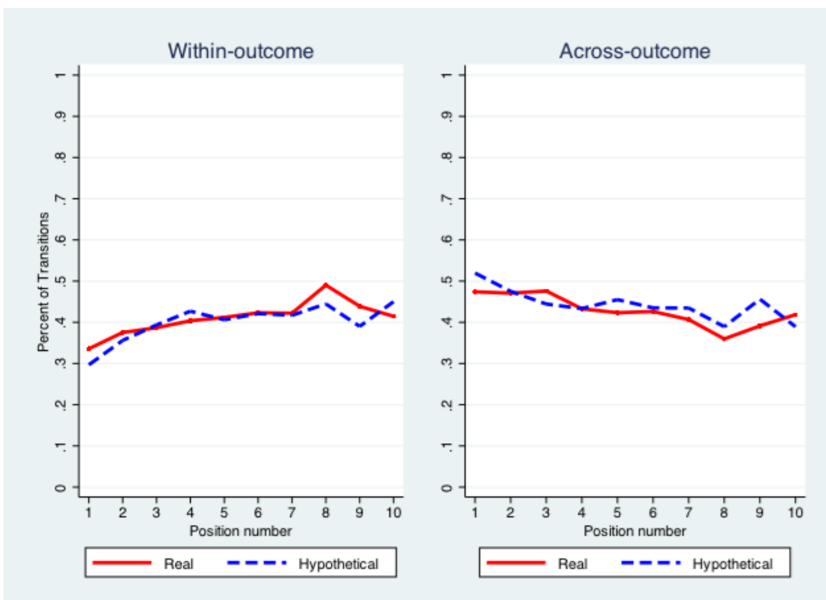


FIGURE 16. Percent of transitions accounted for by the specified type of transition type in real and hypothetical setting in the first task. Position number indicates the order in which the choice was made.

of within-outcome transitions increased between the first and tenth choices subjects completed. The percentage of within-gamble transitions is relatively constant, so the within-outcome and across-outcome percentages should mirror one another fairly closely, and they do: the percentage of across-outcome transitions decreased as subjects progressed through the choices in the first task.

To provide a more complete picture of subject search behavior, the evolution of median transition frequencies for within-outcome and across-outcome transitions is shown in Figure 17. On the left, the median frequency of within-outcome transitions appears relatively flat in the hypothetical setting and slightly decreasing in the real setting. In the graph on the right, the median frequency of across-outcome transitions clearly decreases in both settings. In fact, the graph on the right shows that the primary source of the decrease in total transitions shown in Figure 14 is subjects' make fewer across-outcome transitions as they progress through the task.

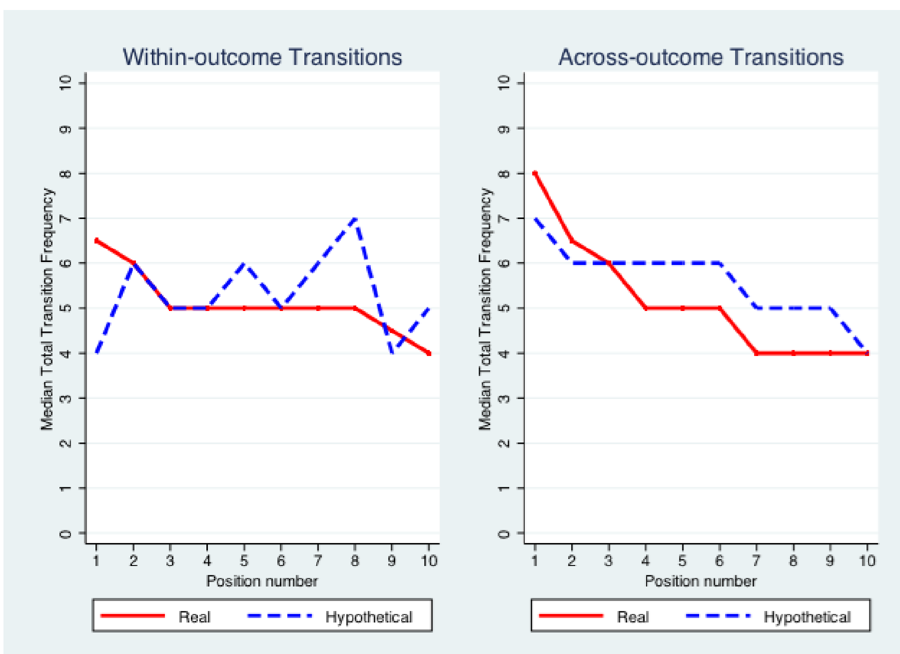


FIGURE 17. Median transition frequency accounted for by the specified type of transition type in real and hypothetical setting. Position number indicates the order in which the choice was made.

One way to interpret this evolution of search behavior is that subjects optimized or streamlined their search behavior by reducing within-gamble exploration. In other words, they minimized across-outcome transitions as they became more familiar with these decisions, perhaps because such transitions do not improve the quality of the decision if they are using an expected utility criterion.

The second task shows that search strategies do not change significantly as subjects progressed through the choices in the second task. In other words, in the second task, information acquisition behavior looks very similar whether a subject was completing the first choice, the second choice, and so forth. This suggests that the optimizing occurs during the first task and any gains in decision-making efficiency carry over into the second task, which is reasonable.

Hypothetical versus Real

Figures 14 through 17 in the previous section suggest no significant differences in how subjects acquired information across settings. In the brief analysis to follow, I focus on percentages because there are information acquisition differences across real and hypothetical settings in the second task in terms of the absolute number of transitions. Specifically, the median transition frequency is eleven in the real setting and nine in the hypothetical setting in the second task (Wilcoxon p-value = 0.076).

14. Within-gamble transitions account for well over eighty percent of total transitions in both settings, and in both tasks. The third column shows that a test of equal means fails to reject the null hypothesis that these percentages are equal for both the first task and the second task, even though there are differences in the absolute numbers of transitions across settings in the second task.

TABLE 14. Within-Gamble Transitions

Percentage of total transitions that occur within the same gamble. P-values shown for t-test of equal means.

	Real	Hypothetical	t-test
First Task	0.838	0.843	0.317
	Hypothetical	Real	
Second Task	0.833	0.831	0.659

The results are similar for the percentage of within-outcome transitions. Table 15 shows that at least forty percent of transitions can be classified as within-outcome transitions in all four treatments. However, a test of equal means rejects the null hypothesis that the percentages are equal in the second task at the nine percent level. Despite this result, the practical importance of a two percentage point difference in this context is unclear.

TABLE 15. Within-Outcome Transitions

Percentage of total transitions that occur between the probability and payoff of the same potential outcome. P-values shown for t-test of equal means.

	Real	Hypothetical	t-test
First Task	0.410	0.400	0.366
	Hypothetical	Real	
Second Task	0.458	0.437	0.088

The results are nearly identical in the choice phase for within-gamble and within-outcome transitions. The mean percentage of within-gamble transitions in the choice phase is at least 76 percent in all four treatments (the median percentages are well over eighty percent), and tests of equal means across settings fail to reject the null of hypothesis in both tasks ($p > 0.161$). Within-outcomes transitions account for more than forty percent of total transitions in each treatment in the choice phase as well (i.e., for all choices after all eight cells have been viewed at least once). Again, the null hypothesis of equal means across settings is not rejected ($p > 0.491$).

Conclusion

Individual decision makers often violate the maintained hypotheses of expected utility theory (EUT) in systematic ways when making choices involving uncertainty. In addition to the systematic violations, some argue that the complexity of many decisions makes it too costly, or even impossible, for humans to compute expected values. The response to this criticism has been that individuals may not compute expected value, but they behave "as if" they do.

Despite the importance of questions about how individuals process information while making decisions under uncertainty, there have been few attempts to actually observe whether individuals acquire information in a pattern consistent with expected utility theory. Three recent studies that have examined choice behavior with this question in mind and have found conflicting evidence (Johnson et al., 2008; Arieli et al., 2011; Glöckner and Herbold, 2011). In contrast to these studies, which use only hypothetical choices, I examine information acquisition using gambles with both real and hypothetical payoffs. I also observe information acquisition behavior in the context of a commonly used risk preference elicitation method, the Holt and Laury multiple-price list. Finally, I avoid a framing effect that may have biased the results of Johnson et al. (2008) toward finding information acquisition behavior consistent with expected utility theory.

The results in this study suggest that subjects acquire information consistent with expected utility theory (or with some other integration model, such as prospect theory). Moreover, they tend to acquire information in sequences consistent with expected utility in both the reading and the choice phase of the search process, and they do not acquire information in a manner consistent with the application of a heuristic. These results are consistent with Johnson et al. (2008) and Arieli et al. (2011), which found a similarly large percentage of information acquisition to be consistent with the calculation of expected utility.

The evolution of search behavior as subjects progress through their choices further supports the hypothesis that individuals integrate corresponding probabilities and payoffs when making choices in this context. As subjects progressed through the experiment, they slightly modified their search strategies in a manner consistent with someone computing expected utility. In particular, the percentage of transitions that are most consistent with an integrative search increased. Additionally, subjects reduced the total number of transitions (which means less overall searching). This behavior implies that subjects optimized their searches as they progressed through the set of choices, even though there was no time constraint.

The use of both real and hypothetical payoffs enabled a comparison of search behavior in both real and hypothetical settings, and subjects do not appear to acquire information differently in real and hypothetical settings while completing the first task. This result is significant because it suggests that, on average, individuals do not use one type of decision making process when making hypothetical choices and a fundamentally different type of decision making process when making real choices. This result is consistent with the recent findings of Kang et al. (2011) that common areas of the brain are activated when individuals make real and hypothetical purchase decisions in a consumer goods context.

Finally, it is worth reiterating that this study observed information acquisition behavior in the context of a standard risk preference elicitation method that assumes individuals are making decisions in accordance with expected utility. The choices that subjects encountered in this study are designed in such a way that subjects can compute expected value, but are sufficiently difficult in terms of the computational effort required that subjects could choose to adopt a decision making heuristic instead. However, the evidence presented here strongly suggests that subjects do not only make choices "as if" they are computing expected utilities, it appears that they are *actually* computing expected utilities.

CHAPTER IV

BIAS AND BRAINS: ARE INDIVIDUALS WITH HIGHER COGNITIVE ABILITY TREATING HYPOTHETICAL CHOICES AS PUZZLES?

Reliable estimates of risk preferences are frequently required to make appropriate evaluations of the welfare implications of many policy changes. It is often necessary, however, for risk preferences to be measured using hypothetical choices. For example, many environmental policy changes effect the allocation of non-market goods, or goods that will actually be provided at some time in the far distant future (e.g., reducing the magnitude of the increase average temperatures, emissions reductions).

There is some evidence, however, to suggest that individuals may respond differently to hypothetical choices involving risk than they do in similar real-choice contexts. In an influential experiment designed to evaluate objections to laboratory-based estimates of risk aversion, Holt and Laury (2002) demonstrate that, on average, individuals tend to make choices that imply that they are more tolerant of risk in a hypothetical setting than in a real setting. The Holt and Laury study also finds that this "hypothetical bias" worsens as the size of the stakes increases. Several other studies, including Holt and Laury (2005), also find evidence of hypothetical bias in risk preferences (Battalio et al., 1990; Harrison, 2006).

However, some studies find no significant difference in average risk attitudes across hypothetical and real contexts (see for example Beattie and Loomes (1997); Camerer (1995); Camerer and Hogarth (1999), and Kuhberger et al. (2002)). For instance, the results presented in Chapter II provide no indication of hypothetical bias in average risk preferences for the sample used in this study. Furthermore, Chapters II and III taken together suggest that there is no evidence that individuals acquire information differently in hypothetical settings as opposed to real settings.

In this chapter, I estimate individual risk preferences using richer models that incorporate individual controls for cognitive ability and information acquisition. More importantly, I allow for the interaction between these individual characteristics and hypothetical settings, which enables the exploration of how individual characteristics are related to risk preferences in different ways across real and hypothetical setting.

This is the first study to compare whether the relationship between cognitive ability and risk aversion differs across real and hypothetical choice settings. Several studies document a tendency for higher cognitive ability to be associated with lower risk aversion (Frederick, 2005; Benjamin et al., 2006; Oechssler et al., 2009; Dohmen et al., 2010; Campitelli and Labollita, 2010), but these studies use either hypothetical choices or small expected payoffs.[1] I use relative large payoffs—ranging from \$1.90 to \$81.00—and every subject makes choices that have real payoffs, as well as choices that are hypothetical. Moreover, both options from which a subject chooses are characterized by risk, in contrast to previous studies that ask subjects to choose between a risky option and a sure option.[2]

Significantly, allowing for risk preferences to vary by cognitive ability across settings reveals that cognitive ability is inversely related to risk aversion when the choices are hypothetical, but it is unrelated to risk aversion when the choices are real. This result implies that apparent differences in average risk preferences across settings in previous studies could have been the result of a sample with higher cognitive ability. For instance, the most influential study of risk preferences across settings

[1]Dohmen et al. (2010) do use a relatively large potential payoff in their study. They ask subjects to choose between a sure option and a gamble, for which the potential payoffs are 0 euros and 300 euros and each potential outcome had a fifty percent probability of realization. However, only one in seven subjects was actually paid, thereby reducing the expected payoffs substantially. Oechssler et al. (2009) use real payoffs with a 75 percent probability of 20 euros, but only one in one hundred subjects was actually paid.

[2]Benjamin et al. (2006) is the exception here. They ask subjects to choose between two gambles in one trial, but they use small stakes (i.e., no more than \$1.60).

used a sample in which half of the subjects were undergraduates, one-third were MBA students, and 17 percent were faculty members (Holt and Laury, 2002). A general population sample might have produced different results.

This interaction between cognitive ability and setting signifies that the issue of hypothetical bias is more nuanced than simply comparing mean behavior across settings. It also provides a potential explanation for hypothetical bias in risk preferences. Such behavior is consistent with the hypothesis put forth by Schkade and Payne (1994) that subjects approach hypothetical choices presented to them in a contingent valuation survey as "puzzles" that they must solve. If subjects are, indeed, "constructing" solutions to the hypothetical choices about risk preferences, then differences in cognitive ability could easily generate systematic differences in their responses to questions about risk attitudes. In particular, it appears as if those with a greater ability to "think slow," in the language of Kahneman (2011), attempt to "solve" the hypothetical choices and respond in accordance with risk neutrality.

Summary of Experiment

This section briefly summarizes some key features of the experiment and discusses the measurement of cognitive ability. A complete exposition of the details of the experiment can be found in Chapter 2.

Subject risk preferences were measured using the Holt and Laury Multiple Price List (MPL) format (see Holt and Laury, 2002). The MPL presents subjects with ten choices. Each choice has two gambles and subjects are asked to indicate which gamble they prefer for each two-alternative choice. Given a subject's choices over the entire choice set, inferences can be made about her risk preferences.

In this experiment, all subjects completed the MPL in both real and hypothetical settings. Half of the subjects were randomly selected to receive the real-payoff treatment in the first task, which meant that one choice from the set of ten included in the MPL would be randomly selected for payment once the subject had completed all ten choices. The other half completed the hypothetical choices in the first task. This means they made similar choices but they did not receive a payoff for any of their choices. In the second task, each subject completed the choices under whichever treatment they had not yet completed.

Once subjects completed the choice tasks and their payoffs were revealed, they were asked to complete a test that measured both numeracy and cognitive ability. Numeracy tests are designed to measure an individual's ability to understand and manipulate numeric and probabilistic information. The computational complexity of our tasks, along with the impact that numeracy could have on information acquisition behavior, made it prudent for us to elicit such a measure of numeracy. I adapted an eight-item test developed by Weller et al. (2011) that includes two of the three items in the cognitive reflective test (CRT) introduced by Frederick (2005). The third item from Frederick's CRT is included in the test as well, so I have two separate measures of ability: a numeracy score that measures a subject's numeracy, and a CRT score that measures a subject's cognitive ability.[3]

The CRT was designed to evaluate an individual's "System 2" cognitive processes — the ability to solve problems that require "effort, motivation, concentration, and the execution of learned rules" (Frederick (2005), pp. 26). In contrast, "System 1" processes are spontaneous and do not require significant effort.[4] Kahneman (2011) highlights that, although some problems can easily be handled with System 1 ("Fast") processes, more difficult problems require the use of System 2 ("Slow") processes. Performance on the CRT, along with other measures of cognitive ability, has been

[3]The complete test is included in Appendix A.

[4]See Stanovich and West (2000) for a discussion of System 1 and System 2.

shown to be predictive of risk preferences in previous studies (Frederick (2005); Benjamin et al. (2006); Oechssler et al. (2009); Dohmen et al. (2010), and Campitelli and Labollita (2010)). However, these studies use hypothetical choices or stakes with relatively small expected values. In contrast, I use both relatively large stakes and I collect choice information from both hypothetical and real settings for each subject.

The final section of the experiment was a debriefing questionnaire that inquired whether subjects had been distracted during the experiment, whether they were liquidity constrained, the income level of their household, their education level and educational aspirations, the extent of their math and probability training, their academic major, and their gender. Table 2 provides selected summary statistics based on subjects' responses to this questionnaire.[5]

Twenty-seven experimental sessions were conducted during March and April of 2011. A total of 98 people, recruited primarily from undergraduate chemistry, economics, and environmental studies courses at the University of Oregon, participated in the experiment. Subjects earned an average total payoff of \$52.68, with a maximum of \$86.00 and a minimum of \$6.90. Nearly all subjects finished within 40 minutes.

After dropping one subject because she did not unmask a single attribute in seven of the ten real choices, the sample includes a total of 97 subjects. Forty-eight subjects completed the real choices in the first task; the remaining forty-nine completed the hypothetical choices first. Based on two-sided t-tests, if the sample is partitioned by whether a subject completed the real or hypothetical choices first, the samples are not statistically different on dimensions commonly found to be related to risk attitudes.[6]

---

[5]The complete set of summary statistics is included in Appendix A. Where appropriate, some response categories are combined into a single category.

[6]See Table A.1 in Appendix A.

Approximately forty percent of the sample is female. Subjects answered 4.49 questions correctly, on average, on the six-item numeracy test. The average number of correct answers on the three-item CRT is 1.27, which is comparable to the 1.24 average score reported in Frederick (2005). The two scores have of correlation coefficient of 0.397. Males tended to score higher on both the numeracy test and the CRT, but not to a statistically different degree (p-value > 0.6044). Only five subjects in the entire sample indicated that they were "distracted at all during this experiment." More than half of the subjects have completed a course in which at least two weeks were spent on probability and statistics. Finally, unlike many experiments that use primarily economics and psychology majors, our sample is more than one-fifth science majors, nearly one-fifth environmental studies/science majors, and another one-fifth business majors.

Risk Aversion and Cognitive Ability

In Chapter II, I show that I do not find differences in implied risk preferences between real and hypothetical settings. Similar to previous studies that have used the MPL to estimate risk preferences, however, I do find that subjects are risk averse on average.

As noted in the introduction, there are several studies that have examined the relationship between cognitive ability and risk aversion. Those studies have generally found that greater cognitive ability is associated with less risk aversion. However, all of these studies use either hypothetical payoffs, relatively small payoffs, or a random lottery mechanism, which reduces the expected payoff and introduces the potential that a subject may not be paid for any of his choices.

Of particular interest here is Frederick (2005), which classified individuals as either low-CRT or high-CRT depending on the number of correct responses on the

CRT. Individuals with zero or one correct answers are classified as low-CRT, while those who answered correctly for two or three questions are classified as high-CRT. Using hypothetical choices, Frederick (2005) finds that high-CRT individuals tend to indicate that they are more tolerant of risk.

If I limit the sample to the hypothetical choices and compare the mean number of safe choices of low-CRT individuals to the mean of high-CRT individuals, I find a similar result. In particular, high-CRT subjects made 5.5 total safe choices, while low-CRT subjects made 6.6 safe choices. The difference in means between the two groups is statistically significant (p-value=0.0484).

Figure 18 shows the average number of safe choices made by low-CRT and high-CRT individuals in both the real setting and the hypothetical setting in the first task. When making choices with real payoffs, it appears that individuals with greater cognitive ability make slightly more safe choices than those with lower cognitive ability. Yet, when the choices are made in a hypothetical setting, those with greater cognitive ability make statistically fewer safe choices than they do in the real setting.

To explore how individual characteristics potentially interact with the hypothetical setting, I obtain estimates of the risk preference parameter employing maximum likelihood methods.[7] Utility is assumed to be described by the standard constant relative risk aversion (CRRA) utility function, $u(w) = \frac{w^{(1-r)}}{(1-r)}$ where w is the gamble payoff. Under expected utility theory, subjects are assumed to make an expected utility calculation of the form:

$$EU_g = p_h \times U(w_h|r) + p_l \times U(w_l|r) \tag{4.1}$$

[7] Harrison and Rutström (2008) provides an invaluable overview of maximum likelihood methods to estimate risk preferences as well as guidelines for programming the models.

FIGURE 18. Total Safe Choices in the First Task by Cognitive Ability Classification (CRT Low=0 or 1 correct answers, CRT High=2 or 3 correct answers)

for each gamble g, where the probability of the high outcome h is denoted p_h and the probability of the low outcome is p_l.

If the safe gamble is denoted as EU_S and the risky gamble as EU_R, a latent index can be specified, based on subject preferences, that allows for deviations from the deterministic choice specified by the expected utility calculation:

$$\nabla EU = \frac{EU_S - EU_R}{\mu} \tag{4.2}$$

Under this specification, commonly referred to as the Fechner error specification, the parameter μ can be interpreted as noise, and as this noise parameter increases the latent index becomes less sensitive to differences in the expected utilities of the gambles and the choices are characterized by more randomness (Hey and Orme, 1994).[8] The index in (4.2) forms the basis of a conditional likelihood function that is linked to choices using either a normal or logistic cumulative distribution function. The log-likelihood function can be denoted $\mathcal{L} = \sum_i l(r, \mu | C_i, X_i)$, where C_i denotes subjects' choices and X_i denotes individual characteristics. This log-likelihood is maximized with respect to unknown parameters r and μ.

Table 16 presents the resulting estimates of the coefficient of relative risk aversion, r, and the noise parameter, μ, where each parameter is allowed to be a function of the real or hypothetical setting, gender, the CRT score and the numeracy score separately, and whether the subject opened sufficient cells to compute expected value for all ten choices.[9] Specification (1) in Table 16 includes an indicator for the hypothetical setting and controls for individual characteristics. The results for these simple linear-in-variables indexes for the systematically varying parameters r and μ are shown in the first three columns of the table. The coefficient on the indicator variable for hypothetical setting is small and statistically insignificant, suggesting that there is no simple additively separable form of hypothetical bias in either of these two parameters.

Consistent with the reviews by Eckel and Grossman (2008) and Croson and Gneezy (2009), the coefficient on the female indicator variable indicates that females are more risk averse at a statistically significant level. The insignificant coefficient on the hypothetical setting indicator variable in the expression for μ in Specification (1) in Table 3 also suggests that there is not additional noise in the hypothetical setting relative to the real setting. The breadth of subjects' acquired information is,

[8] I also explored a Luce error specification, $\nabla EU = \frac{EU_S^{\frac{1}{\mu}}}{EU_S^{\frac{1}{\mu}} + EU_R^{\frac{1}{\mu}}}$, but the preferred specification fails to converge. However, for specifications that did converge, the results were not substantively different from the results of comparable specifications using a Fechner error structure. The results reported here are robust to the choice of either normal or logistic cumulative distribution function.

[9] Attempts to include the indicator variable for the completion of a statistics or probability course resulted in convergence problems, so this regressor is not included in our model.

unsurprisingly, related to the level of noise: subjects who took the opportunity to acquire EV-sufficient information for all ten choices made less noisy choices.

TABLE 16. Estimates of risk parameter "r" and noise parameter "μ." CRRA utility function, Fechner error specification. First task only.

	Specification (1)			Specification (2)		
	No Interaction Terms			With Interaction Terms		
VARIABLES	Estimate	Std. Err.†	p-value	Estimate	Std. Err.†	p-value
r						
1(Hypot)	0.021	0.086	0.807	-0.088	0.399	0.825
1(Female)	0.197**	0.088	0.025	0.296***	0.108	0.006
1(Female)×1(Hypot)				-0.039	0.156	0.804
CRT Score	0.003	0.046	0.952	0.089	0.058	0.122
CRT×1(Hypot)				-0.192**	0.089	0.031
Numeracy Score	-0.073	0.046	0.111	-0.113	0.071	0.112
Numeracy×1(Hypot)				0.093	0.080	0.249
1(EV-sufficient All Choices)	0.098	0.101	0.335	0.186	0.131	0.156
1(EV-sufficient All)×1(Hypot)				-0.138	0.190	0.469
Constant	0.816***	0.208	0.000	0.813**	0.321	0.011
μ						
1(Hypot)	0.034	0.559	0.952	0.739	2.199	0.737
1(Female)	-0.532	0.450	0.236	-0.941**	0.398	0.018
1(Female)×1(Hypot)				-0.408	0.749	0.586
CRT Score	0.056	0.232	0.810	-0.475*	0.260	0.068
CRT×1(Hypot)				1.448***	0.501	0.004
Numeracy Score	0.126	0.214	0.557	0.257	0.207	0.215
Numeracy Score×1(Hypot)				-0.628**	0.308	0.042
1(EV-sufficient All Choices)	-1.770*	0.917	0.054	-2.378*	1.396	0.088
1(EV-sufficient All)×1(Hypot)				0.973	1.731	0.574
Constant	2.744**	1.173	0.019	3.360***	1.237	0.007
Observations	970			970		
LogL	-374.4			-367.0		

Notes: † Standard error estimates clustered on subject. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In Specification (2) in Table 16, by including interaction terms between the hypothetical setting indicator and each of the individual characteristics, I test whether there are differences across settings in the incremental effects on r and μ of each of our individual controls. Here, again, we find that females are more risk averse at a statistically significant level.

The most important finding from this experiment, however, is that the coefficient on the interaction between the CRT score and the hypothetical setting indicator is statistically significant. It suggests that subjects with higher cognitive ability are less risk averse in the hypothetical setting than those who have lower cognitive ability. A higher CRT score is apparently not related to risk aversion in the real setting, however. Figure 19 plots the predicted coefficient of relative risk aversion against CRT score in the real and the hypothetical settings.

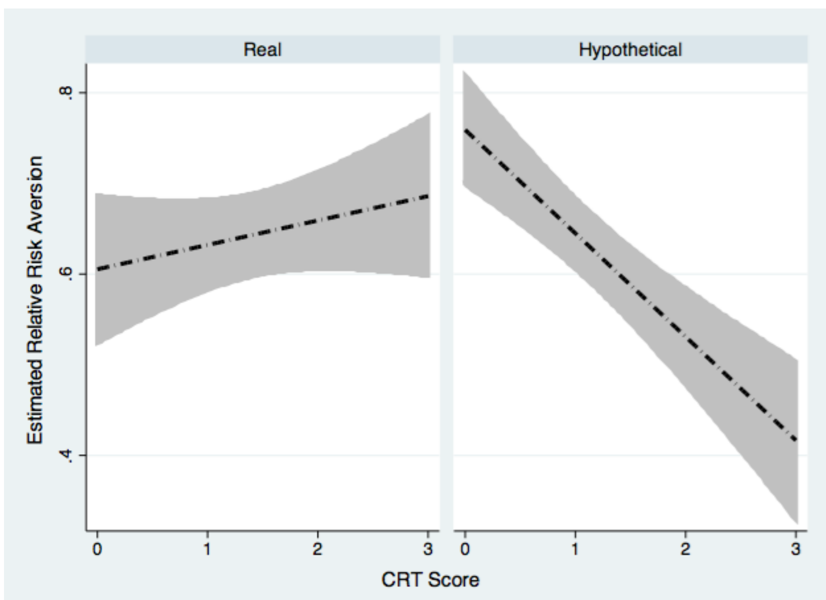


FIGURE 19. Real versus Hypothetical: Estimated Relative Risk Aversion with CRRA Utility and Fechner Noise by CRT Score, First Task

The difference in the slopes between the two graphs in Figure 19 is striking. Whereas the estimated coefficients of relative risk aversion are nearly identical for subjects with low and high cognitive ability in the real setting, there is a dramatic difference in estimated risk aversion between subjects with low and high cognitive ability in the hypothetical setting.

Specification (2) in Table 16 also suggests that the estimated noise parameter is related to cognitive ability in an interesting way. Higher cognitive ability is associated with less noise in the real choices, but with more noise in the hypothetical choices. One possible explanation for the differential in noise for higher cognitive ability across settings may be that subjects are more likely to focus and to try to make "accurate" choices when the stakes are real. This is only speculation, of course, and this issue is worthy of further study.

Discussion of Risk Aversion Results

Given the additional effort required for the choices in this experiment relative to the choices in an experiment that used the conventional MPL, I believed my design would be more likely to detect hypothetical bias. There is no evidence of a hypothetical bias in average risk preferences. However, I do find that cognitive ability, as measured by the CRT test, is related to risk aversion for hypothetical choices but not for real choices.

The main implication of this finding is that apparent hypothetical bias in average risk attitudes in a between-subjects comparison may be the result of omitted variable bias. In particular, a study that used a sample of high cognitive ability individuals would tend to find hypothetical bias in average risk preferences.

Why is the relationship between cognitive ability and risk aversion limited to hypothetical choices? One possible explanation is that some individuals treat hypothetical choices as "puzzles." In a verbal-protocol process-tracing study of how subjects arrived at values to hypothetical choices eliciting willingness to pay for an environmental good, Schkade and Payne (1994) conclude that individuals "constructed" responses to the hypothetical choices presented to them in contingent valuation (CV) studies. As Green et al. (1998) put it, "subjects often treat CV questions as puzzles to which they must 'construct' a solution."[10] Similarly, Chilton and Hutchinson (2003) find evidence that *some* subjects appear to "construct" their responses to questions that elicit their hypothetical willingness to pay.

Although the studies mentioned above explore willingness to pay and not risk preferences, if subjects view hypothetical choices in general as puzzles to be solved, then it is entirely possible that "solutions" will differ by cognitive ability. The mechanism could work in at least two ways. One possibility is that *all* subjects view the hypothetical choices as puzzles, but only high cognitive ability individuals are able or inclined to "solve" the puzzle. Those with higher cognitive ability will determine that "the solution" is risk neutrality and will make their choices accordingly; those with low cognitive ability cannot solve the puzzle so their choices reflect their actual preferences.

It is also possible that the hypothetical contexts triggers a "problem-solving switch" in high cognitive ability individuals *only*. Toplak et al. (2011) suggests that those who score higher on the CRT have higher cognitive ability, and also differ in their *thinking disposition*. The study finds that high-CRT subjects also score higher on measures of open-minded thinking (more open-minded), consideration of future consequences (more considerate), and superstitious thinking (less superstitious). Thus, it is also possible that only high-cognitive-ability individuals even recognize that there is a puzzle to be solved.

Whether some individuals, or all individuals, treat hypothetical choices about risk preferences as puzzles deserves additional study. Whatever the particular mechanism

[10] In the study, subjects were asked about their willingness to pay for nets to cover oil ponds to keep migratory birds from drowning in the aftermath of the *Exxon Valdez* oil spill. The authors used a verbal protocol method to track how subjects arrived at their responses.

at work, the interaction between cognitive ability and setting indicates that it is important for future studies comparing risk preferences across hypothetical and real settings to control for cognitive ability. Moreover, if cognitive ability is the source of hypothetical bias in risk preferences, it is also possible that it can be measured routinely and used to calibrate responses to hypothetical choices about risk.

Discussion and Conclusion

I conducted an experiment to test whether differences in estimated risk preferences across real and hypothetical settings are associated with differences in either information acquisition behavior or cognitive ability. Although the potential payoffs in this experiment were relatively large and the research design increased the amount of subject effort required in the tasks, there is no evidence of "hypothetical bias" in average risk preferences in this sample.

Notably, however, I find that risk aversion is inversely related to cognitive ability when the choices are hypothetical, but is unrelated to cognitive ability when the choices are real. This relationship hints at a potential explanation for hypothetical bias in average risk preferences detected in some previous studies (and represents a possible source for the conflicting prior evidence concerning this tendency). Specifically, if a study uses a sample of subjects with higher average cognitive ability, then a between-subject comparison would be more likely to indicate that there is hypothetical bias in average behavior.

This new evidence that risk preferences are related to cognitive ability in hypothetical settings is consistent with the hypothesis that some, or all, subjects treat hypothetical choices as puzzles to which they must construct a solution. If all subjects are in fact treating hypothetical choices as puzzles, it is plausible that the solutions will differ by cognitive ability because subjects with higher cognitive ability can “solve” the puzzle. It is also possible that only the high-cognitive-ability subjects treat hypothetical choices as puzzles because they recognize that there is a puzzle to be solved.

The interaction between cognitive ability and real versus hypothetical settings poses both a challenge and an opportunity for the use of hypothetical choices to study risk preferences. It poses a challenge because it is consistent with the hypothesis that at least some subjects treat hypothetical choices as puzzles to be solved, which confounds the measurement of risk preferences. It also presents an opportunity because if the source of hypothetical bias is identified then it is possible to control for it and responses can be appropriately calibrated.

Hypothetical scenarios are extremely valuable because they allow economists to explore questions about risk preferences that would be prohibitively costly or impractical with real consequences. In some cases, the only way to proceed with research is to use hypothetical choice scenarios. Hypothetical bias, however, remains an issue. The evidence in this study provides a possible explanation for hypothetical bias in risk preferences. This issue deserves further study because if the actual source of bias can be identified, then perhaps it can be better mitigated.

APPENDIX A

SUPPLEMENTARY MATERIAL

Script of Instructions Read Aloud

Thank you for participating in this experiment. Your participation will allow us to explore questions economists have about individual decision making.

In this experiment, you will be asked to make some choices. There are no right or wrong choices. Additionally, although there are other people simultaneously completing the experiment, you are not competing against anyone.

After you have made all of your choices, your payoff will be displayed on the computer. Please stop at this page and raise your hand to let us know that you are at your "payoff page" and we will verify your payoff and prepare your payment.

Note that there are some additional questions after the payoff page that we ask that you also complete.

You will be paid in cash today before you leave. You will also be paid a $5 show-up fee. If you decide to leave early, you may do so. If you do leave early, you will be allowed to keep the $5 show-up payment, but you will forfeit any experimental earnings.

We understand that you may have questions, but it is important that we maintain the integrity of the experiment by minimizing talking and other disruptions. If you have a question before you begin making your choices, please write it down on the card next to your computer and raise your hand. We will then check to see if it is a question that we can answer.

Finally, we ask that you do not discuss this experiment with anyone who may also participate in this experiment so that we can maintain the integrity of our results. Thank you.

Numeracy and Cognitive Ability Test

Please answer the following questions to the best of your ability.

1. Imagine that we roll a fair, six-sided die 1,000 times. (That would mean that we roll one die from a pair of dice.) Out of 1,000 rolls, how many times do you think the die would come up as an even number?

2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket from BIG BUCKS?

3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?

4. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000 people:

5. If the chance of getting a disease is 20 out of 100, this would be the same as having a ______% chance of getting the disease.

6. Suppose you have a close friend who has a lump in her breast and must have a mammogram. Of 100 women like her, 10 of them actually have a malignant tumor and 90 of them do not. Of the 10 women who actually have a tumor, the mammogram indicates correctly that 9 of them have a tumor and indicates incorrectly that 1 of them does not have a tumor. Of the 90 women who do not have a tumor, the mammogram indicates correctly that 81 of them do not have a tumor and indicates incorrectly that 9 of them do have a tumor. The table below summarizes all of this information. Imagine that your friend tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?

	Tested positive	Tested negative	Totals
Actually has a tumor	9	1	10
Does not have a tumor	9	81	90
Totals	18	82	100

FIGURE A.1. Table provided for subjects

7. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? [CRT question]

8. In a lake, there is a patch of lilypads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? [CRT question]

9. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? [CRT question]

TABLE A.1. Full Summary Statistics

Comparison of summary statistics across treatments. "Real First" indicates subjects completed the real choices in the first task; "Hyp. First" indicates subjects completed the hypothetical choices in the first task

Variables	Real first	Hyp. First	p-value
N	48	49	–
Female	.46	.35	.268
Numeracy Score	4.56	4.43	.545
CRT Score	1.42	1.12	.138
1(Distracted)	.04	.06	.667
Stats/Prob. Course	.58	.55	.751
Major			
Economics	.10	.10	.973
Environmental Science	.21	.16	.573
Science	.23	.22	.957
Business	.21	.18	.763
Other	.25	.33	.411
Educational attainment			
1(H.S. or Some college)	.85	.84	.815
1(2-yr or bachelor's)	.15	.16	.815
1(Years college $\leq$ 1)	.44	.24	.046
1(Years college $>$ 1)	.56	.75	.046
Educational aspirations			
1(2-yr, BA, or done)	.44	.37	.486
1(MA or professional)	.40	.43	.746
1(Doctoral)	.17	.20	.640
Status			
1(Student only)	.69	.74	.613
1(Student & PT emp.)	.21	.12	.259
1(Status: other)	.10	.14	.568
Other Demographics			
Dependent	.73	.84	.202
Age	22.1	21.5	.409
Height: Males (inches)	69.8	71.1	.135
Height: Females (inches)	64.8	66.7	.104
Income category			
1(<$10,000)	.21	.27	.515
1($10,000-$74,999)	.27	.22	.601
1(>$75,000)	.40	.39	.936
1(Declined to answer)	.13	.12	.970

Notes: One subject was dropped dropped from the analysis because she did not unmask a single attribute in seven of her ten choices in the real task.

APPENDIX B

EXAMPLE OF EXPERIMENT

This appendix includes screenshots of the experiment.

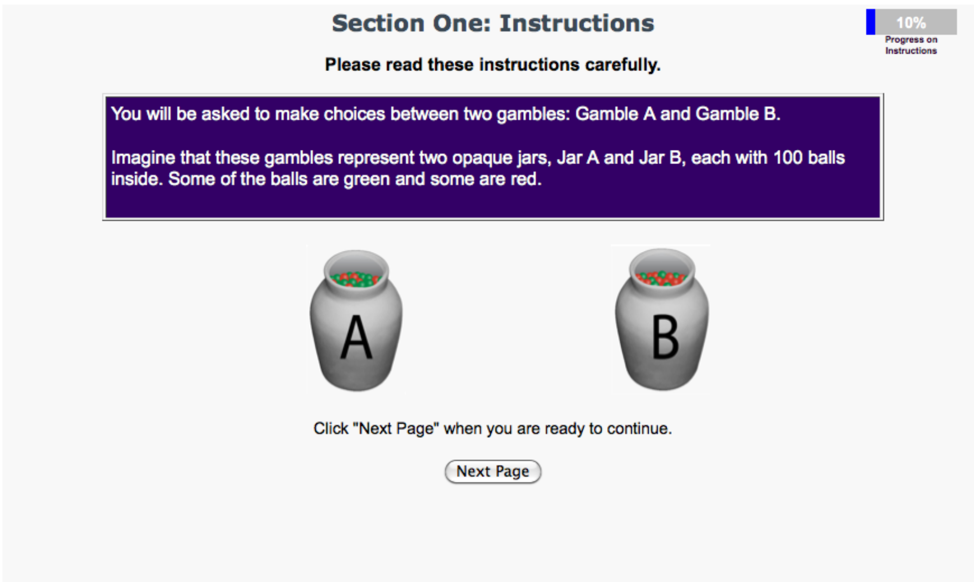


FIGURE B.1. Experiment, Introduction.

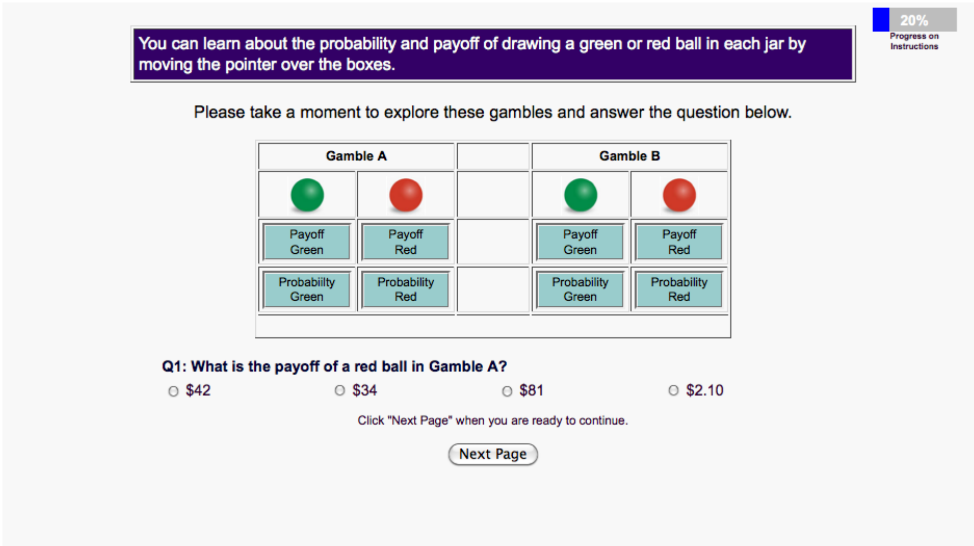


FIGURE B.2. Experiment Tutorial, page 1 of 4. The first of four pages with instructions about how the gambles will be presented. This page also includes a question to test a subject's understanding. Subjects can not continue unless they answer the question correctly.

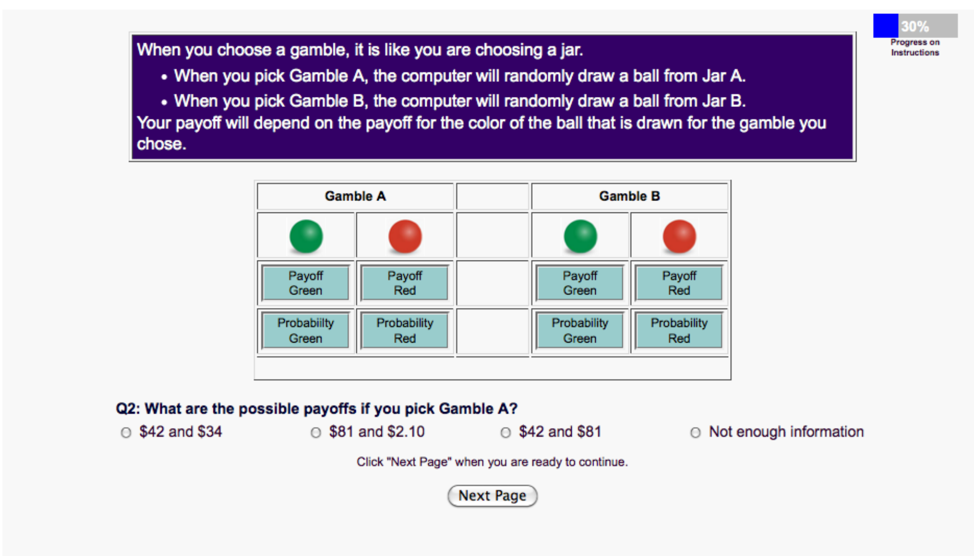


FIGURE B.3. Experiment Tutorial, page 2 of 4. The second of four pages with instructions about how the gambles will be presented. This page also includes a question to test a subject's understanding. Subjects can not continue unless they answer the question correctly.

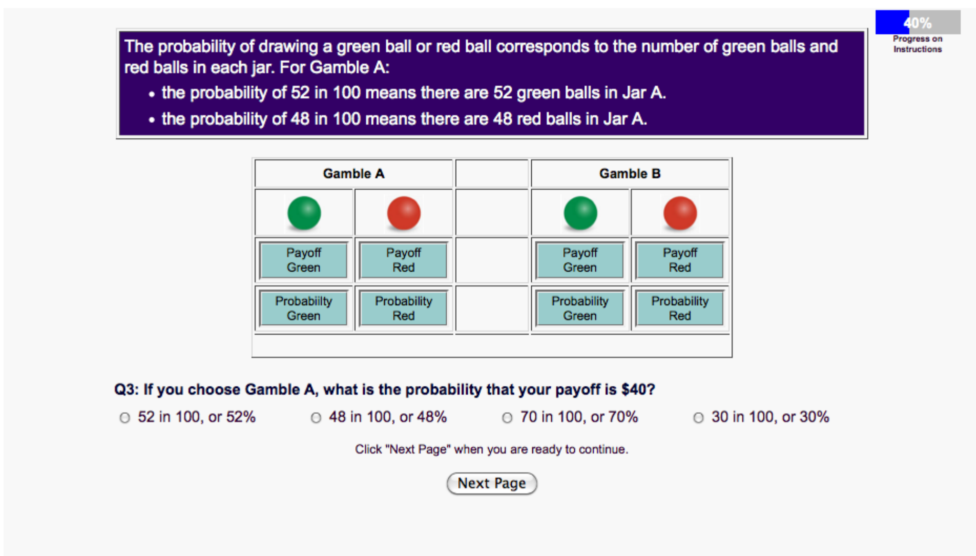


FIGURE B.4. Experiment Tutorial, page 3 of 4. The third of four pages with instructions about how the gambles will be presented. This page also includes a question to test a subject's understanding. Subjects can not continue unless they answer the question correctly.

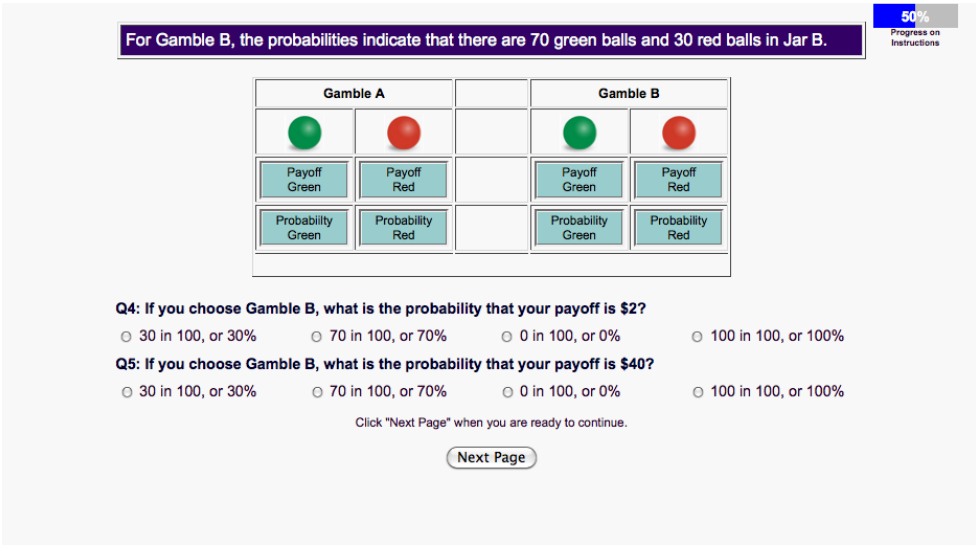


FIGURE B.5. Experiment, 4 of 4 pages. The fourth of four pages with instructions about how the gambles will be presented. This page includes two questions to test a subject's understanding. Subjects cannot continue until they answer all questions correctly.

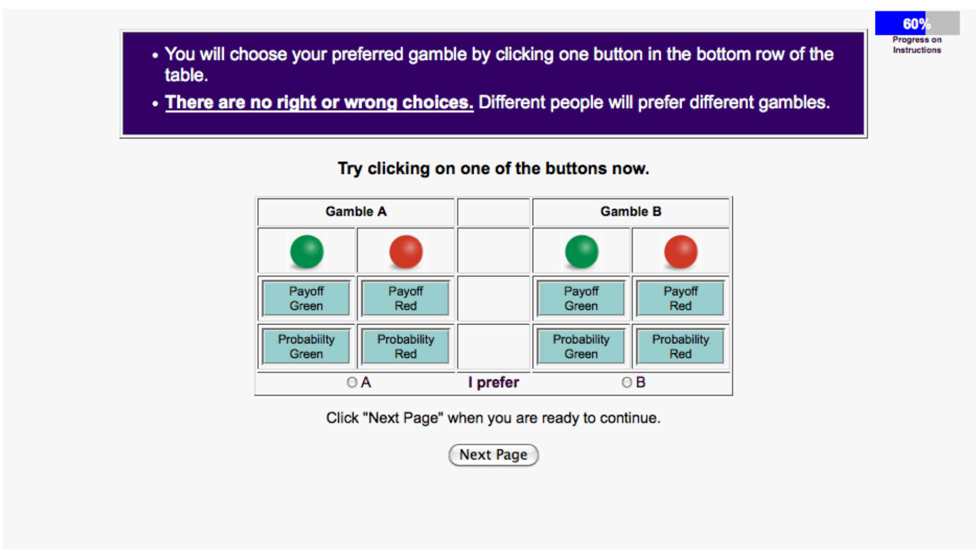


FIGURE B.6. How to Indicate Preferred Choice. This page instructs subjects on how to indicate their preference between the pair of gambles.

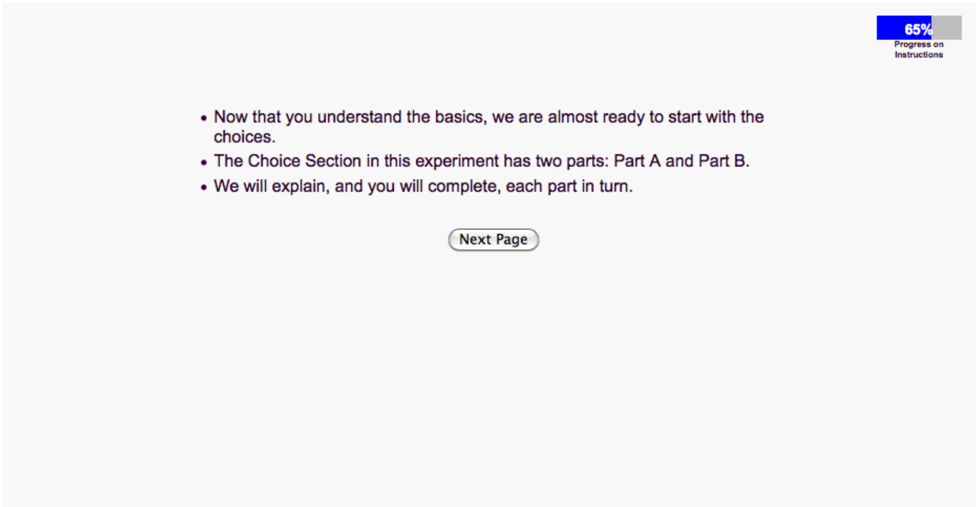


FIGURE B.7. Experiment, Page 7. This page explains that there will be two sections in which they will make choices.

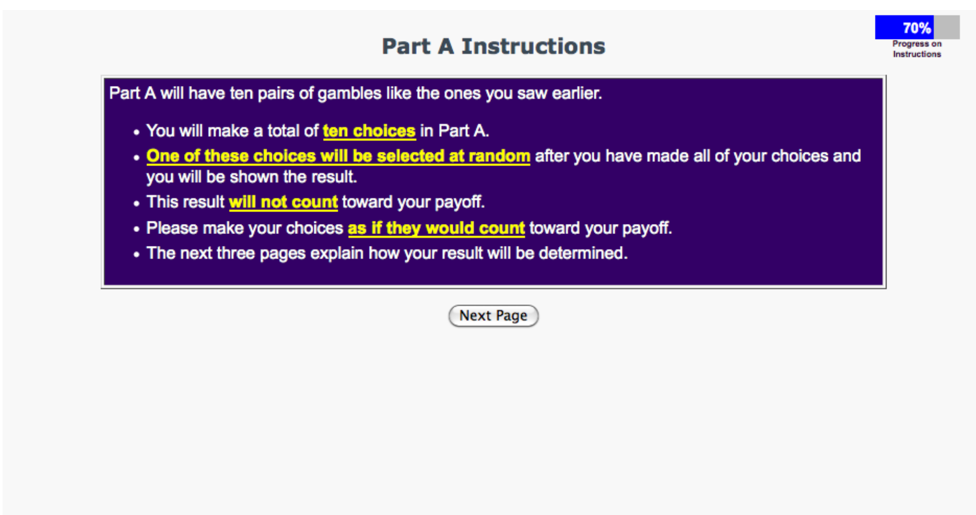


FIGURE B.8. Example of Part A Instructions. Instructions for the choices in Part A. These instructions differ by whether the subject was selected to complete real or hypothetical choices first. Subjects selected to complete the real choices first saw the instructions shown in Figure B.15.

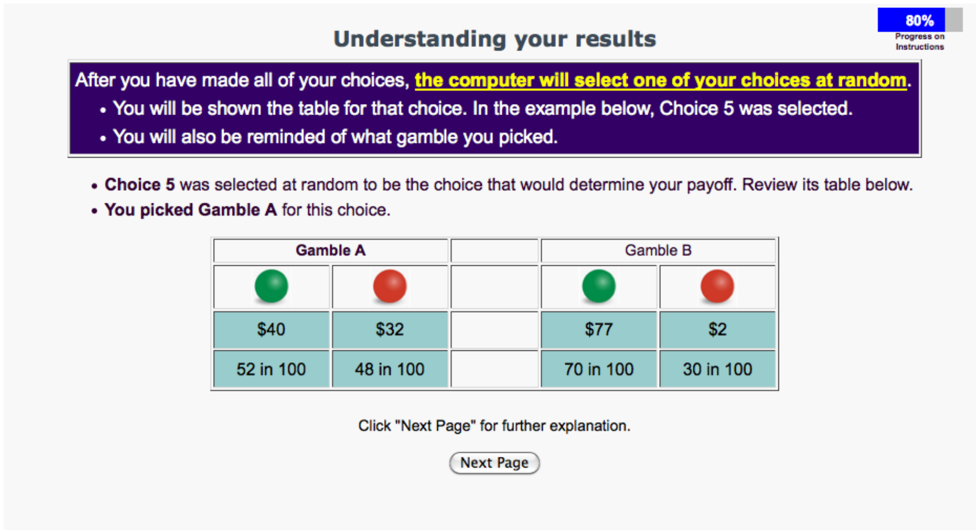


FIGURE B.9. Explanation of Results, page 1 of 3. The first of three pages explaining how subjects' payoffs will be determined. These differ depending on whether subjects completed real or hypothetical choices first.

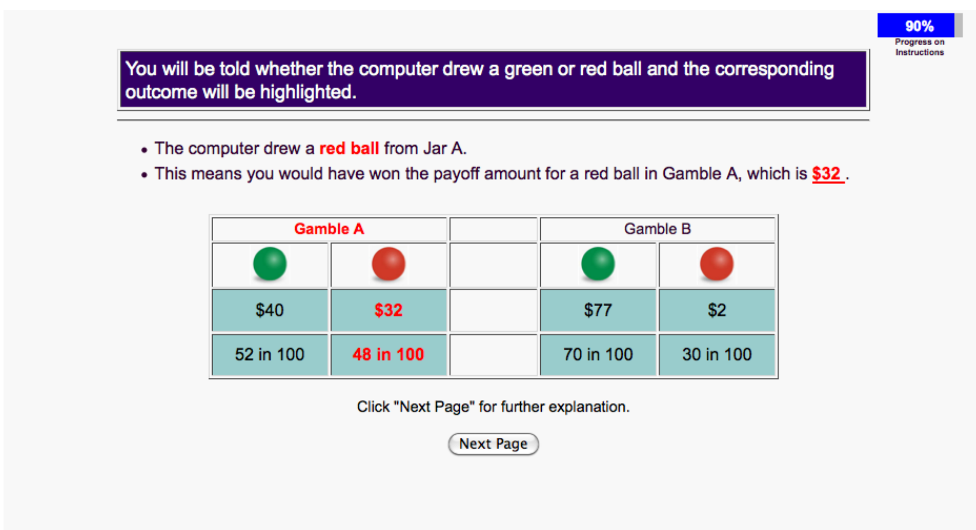


FIGURE B.10. Explanation of Results, page 2 of 3. The second of three pages explaining how subjects' payoffs will be determined. These differ depending on whether subjects completed real or hypothetical choices first.

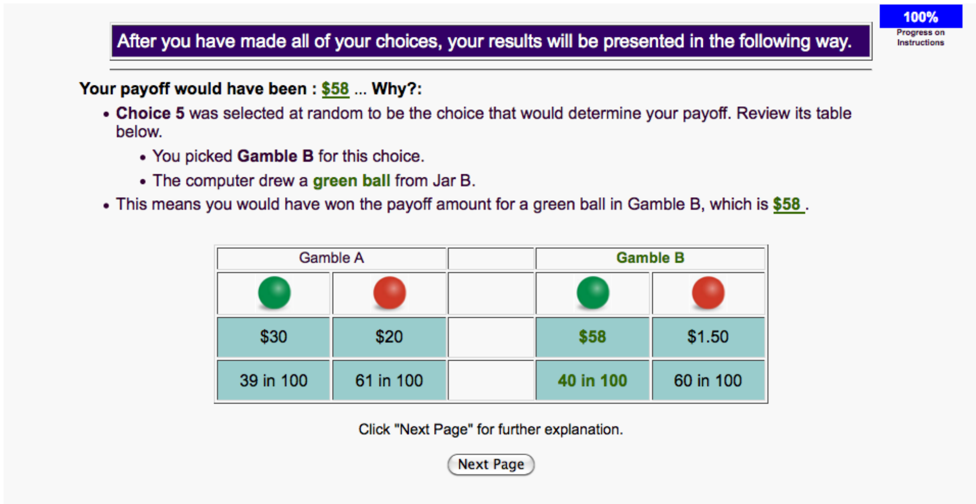


FIGURE B.11. Explanation of Results, page 3 of 3. The third of three pages explaining how subjects' payoffs will be determined. These differ depending on whether subjects completed real or hypothetical choices first.

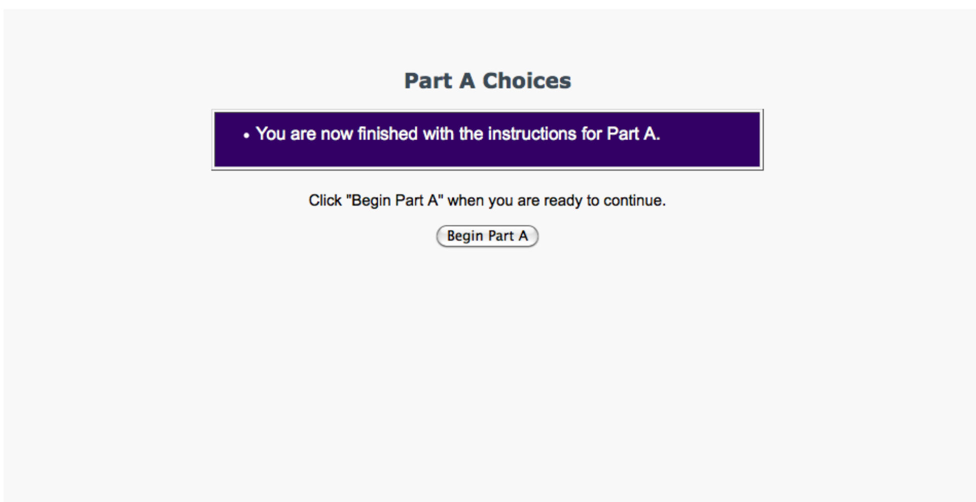


FIGURE B.12. Begin Part A Choices.

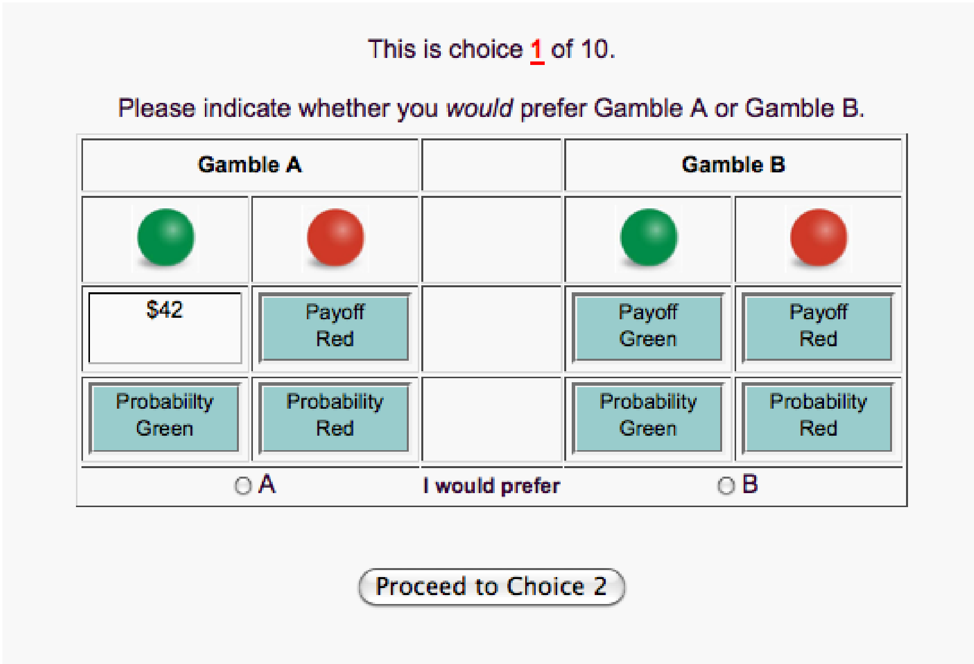


FIGURE B.13. Example of Hypothetical Choice

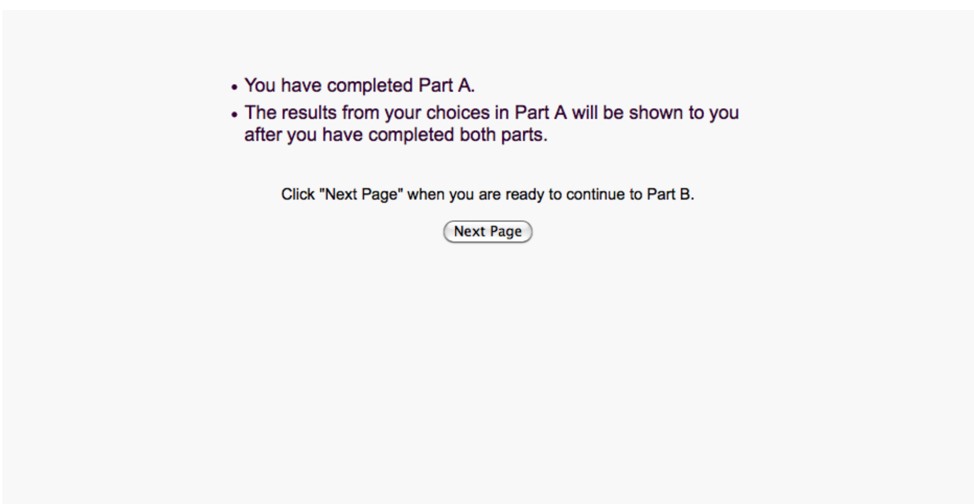


FIGURE B.14. Transition Between Choice Sections

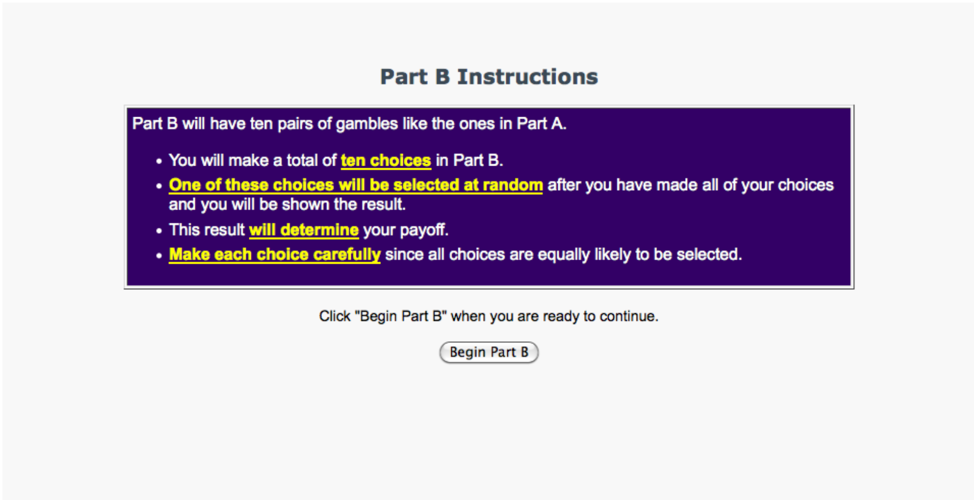


FIGURE B.15. Example of Part B Instructions. Instructions for the choices in Part B. These differ depending on whether subjects completed real or hypothetical choices first. Subjects selected to complete the real choices first saw these instructions before making their choices in Part A.

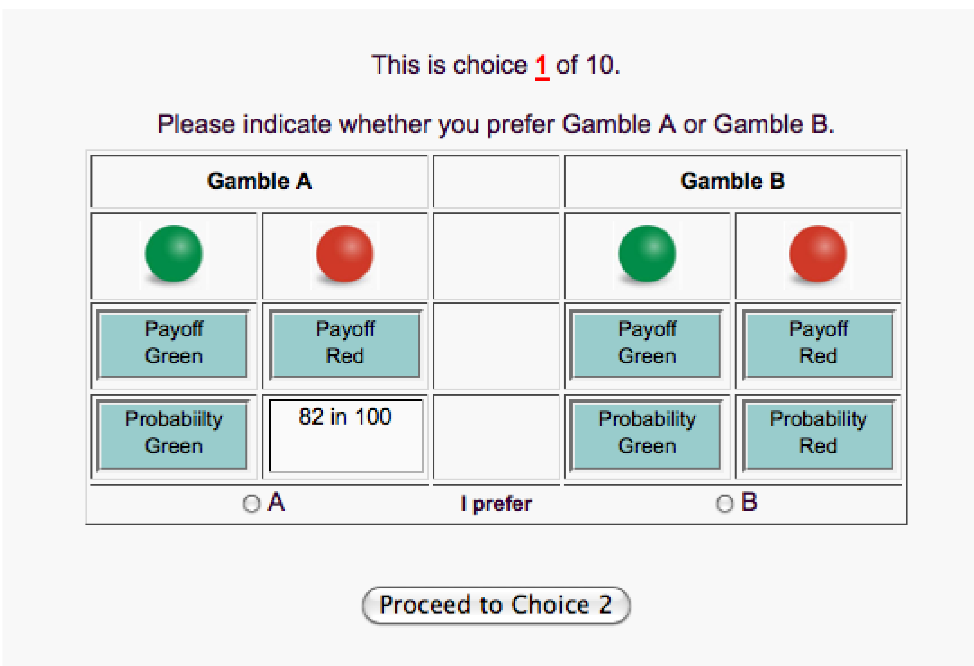


FIGURE B.16. Example of Real Choice

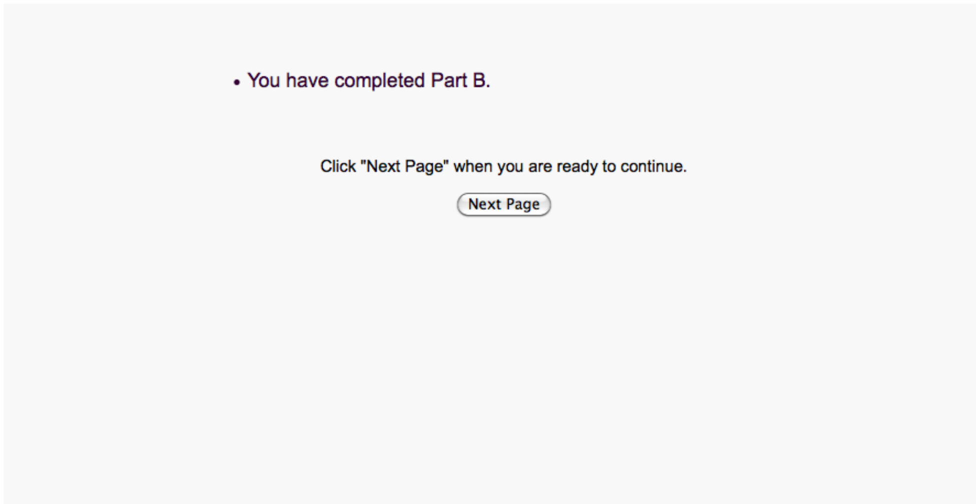


FIGURE B.17. Completion of Part B.

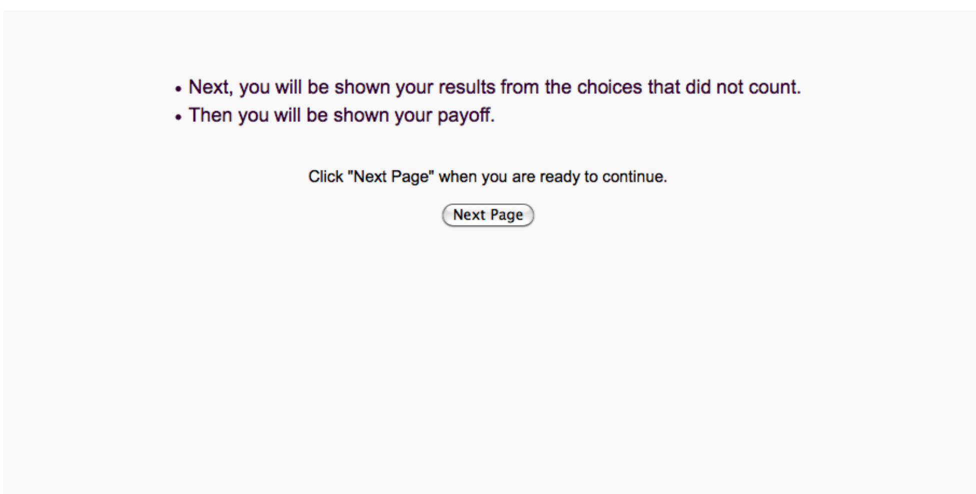


FIGURE B.18. Transition to Results.

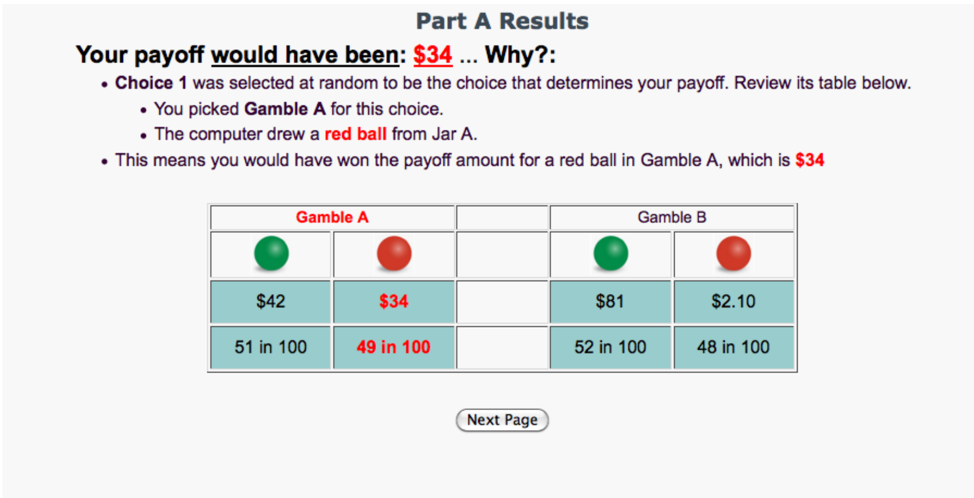


FIGURE B.19. Example of Hypothetical Task Results

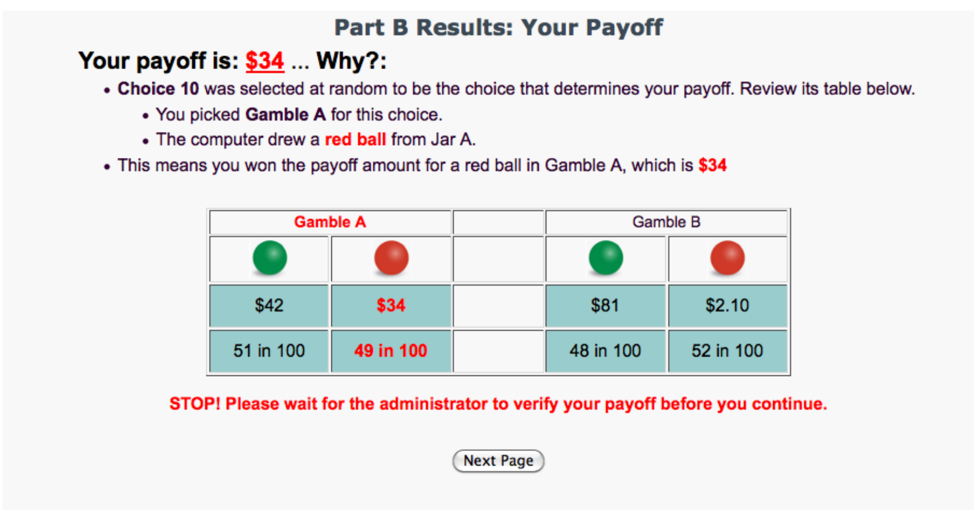
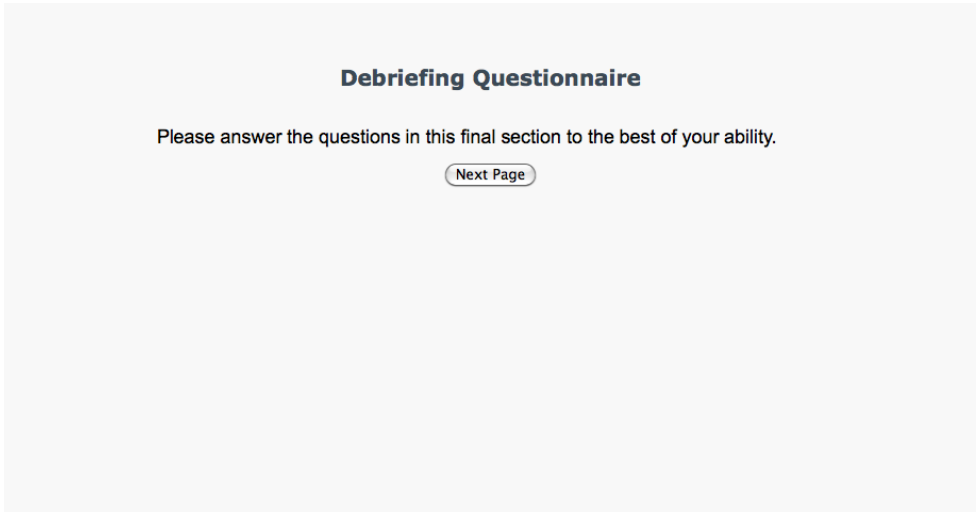


FIGURE B.20. Example of Real Task Results

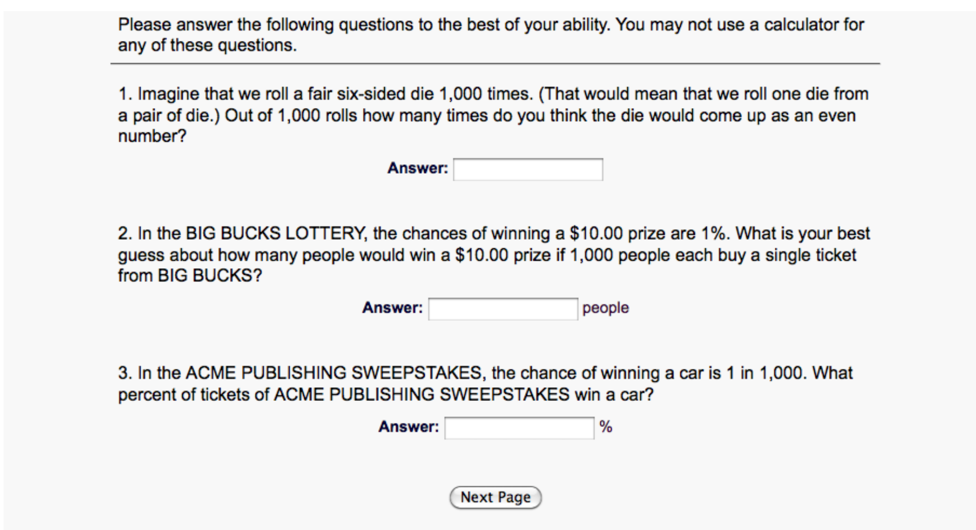


Debriefing Questionnaire

Please answer the questions in this final section to the best of your ability.

Next Page

FIGURE B.21. Numeracy Test Instructions.



Please answer the following questions to the best of your ability. You may not use a calculator for any of these questions.

1. Imagine that we roll a fair six-sided die 1,000 times. (That would mean that we roll one die from a pair of die.) Out of 1,000 rolls how many times do you think the die would come up as an even number?

Answer:

2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket from BIG BUCKS?

Answer: people

3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?

Answer: %

Next Page

FIGURE B.22. Numeracy Test, page 1 of 4. Questions 1 through 3 of the numeracy test.

Please answer the following questions to the best of your ability. You may not use a calculator for any of these questions.

4. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000 people:

Answer: ________ people

5. If the chance of getting a disease is 20 out of 100, this would be the same as having a _____% chance of getting the disease.

Answer: ________ % chance

Next Page

FIGURE B.23. Numeracy Test, page 2 of 4. Questions 4 through 5 of the numeracy test.

Please answer the following questions to the best of your ability. You may not use a calculator for any of these questions.

6. Suppose you have a close friend who has a lump in her breast and must have a mammogram. Of 100 women like her, 10 of them actually have a malignant tumor and 90 of them do not. Of the 10 women who actually have a tumor, the mammogram indicates correctly that 9 of them have a tumor and indicates incorrectly that 1 of them does not have a tumor. Of the 90 women who do not have a tumor, the mammogram indicates correctly that 81 of them do not have a tumor and indicates incorrectly that 9 of them do have a tumor. The table below summarizes all of this information. Imagine that your friend tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?

	Tested positive	Tested negative	Totals
Actually has a tumor	9	1	10
Does not have a tumor	9	81	90
Totals	18	82	100

Answer: ________ out of ________

Next Page

FIGURE B.24. Numeracy Test, page 3 of 4. Question 6 of the numeracy test.

Please answer the following questions to the best of your ability. You may not use a calculator for any of these questions.

7. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

Answer: ______ cents

8. In a lake, there is a patch of lilypads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Answer: ______ days

9. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

Answer: ______ minutes

Next Page

FIGURE B.25. Cognitive Reflective Test (CRT), page 4 of 4. Questions 7 through 9 of the numeracy test.

About You

1. Were you distracted at all during this experiment?

○ Yes ○ No

2. What is the highest level of education you have completed?

○ High school or less ○ Some college, but no degree ○ 2-year degree ○ Bachelor's degree ○ Master's degree ○ Professional degree ○ Doctoral degree

3. If your education is not complete, what is your goal?

○ High school ○ 2-year degree ○ Bachelor's degree ○ Master's degree ○ Professional degree ○ Doctoral degree ○ I'm done

4. If you have attended any college, how many years of college have you completed?

○ Less than 1 year ○ 1 year ○ 2 years ○ 3 years ○ 4 years ○ More than 4 years ○ Not applicable

5. If you have attended any college, what is (was) your major field of study?

______

6. If you have attended any college, how many college-level math courses have you completed?

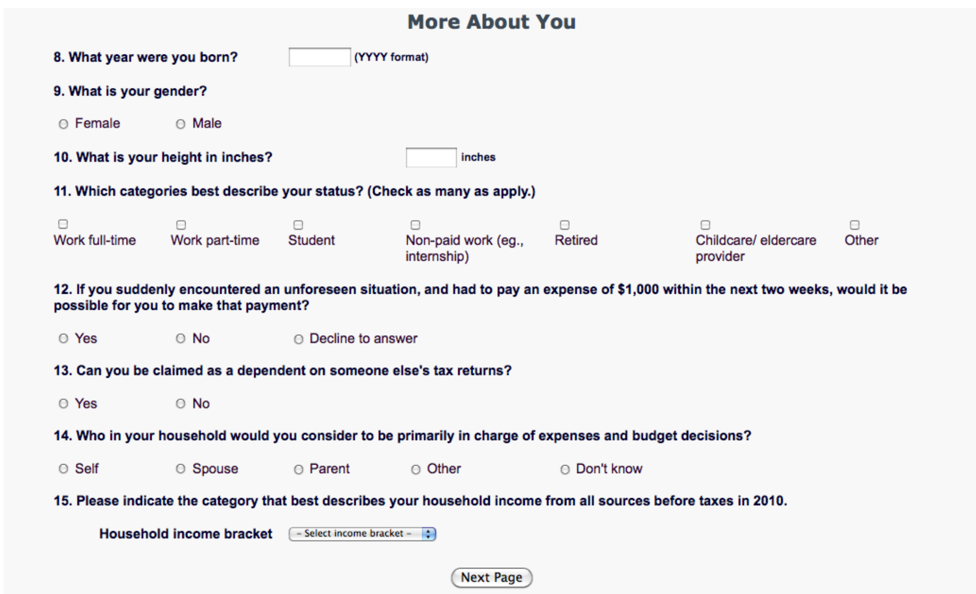
○ One ○ Two ○ Three ○ Four ○ Five ○ More than five ○ Not applicable

7. Have you completed a college-level course in which at least two lectures were spent on probability or statistics?

○ Yes ○ No

Next Page

FIGURE B.26. Debriefing Questionnaire, page 1 of 2. First page of debriefing questionnaire.



More About You

8. What year were you born? ☐ (YYYY format)

9. What is your gender?

○ Female ○ Male

10. What is your height in inches? ☐ inches

11. Which categories best describe your status? (Check as many as apply.)

☐ Work full-time ☐ Work part-time ☐ Student ☐ Non-paid work (eg., internship) ☐ Retired ☐ Childcare/ eldercare provider ☐ Other

12. If you suddenly encountered an unforeseen situation, and had to pay an expense of $1,000 within the next two weeks, would it be possible for you to make that payment?

○ Yes ○ No ○ Decline to answer

13. Can you be claimed as a dependent on someone else's tax returns?

○ Yes ○ No

14. Who in your household would you consider to be primarily in charge of expenses and budget decisions?

○ Self ○ Spouse ○ Parent ○ Other ○ Don't know

15. Please indicate the category that best describes your household income from all sources before taxes in 2010.

Household income bracket – Select income bracket –

Next Page

FIGURE B.27. Debriefing Questionnaire, page 2 of 2. Second page of debriefing questionnaire.

REFERENCES CITED

Allais, Maurice, "Le comportement de l'homme rationel devant le risque: Critique des postulats et axioms de l'ecole americaine [Rational man's behavior in face of risk: Critique of the American School's postulates and axioms]," *Econometrica*, 1953, *21*, 503–546.

——, "Allais Paradox," in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, 2 ed., Palgrave MacMillan, 2008.

Arieli, Amos, Yaniv Ben-Ami, and Ariel Rubinstein, "Tracking Decision Makers under Uncertainty," *American Economic Journal: Microeconomics*, 2011, *3*, 68–76.

Battalio, Raymond C., John H. Kagel, and Komain Jiranyakul, "Testing Between Alternative Models of Choice Under Uncertainty: Some Initial Results," *Journal of Risk and Uncertainty*, 1990, *3*, 25–50.

Beattie, J. and G. Loomes, "The Impact of Incentives Upon Risky Choice Experiments," *Journal of Risk and Uncertainty*, 1997, *14* (2), 155–168.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro, "Who is "Behavioral"? Cognitive Ability and Anomalous Preferences," *Available at SSRN: http://ssrn.com/abstract=675264*, May 2006.

Bettman, James R. and Jacob Jacoby, "Patterns of Processing in Consumer Information Acquisition," in Beverlee B. Anderson, ed., *Advances in Consumer Research*, Vol. 3, Cincinnati, Ohio: Association for Consumer Research, 1976, pp. 315–320.

Brandstatter, Eduard, Gerd Gigerenzer, and Ralph Hertwig, "The Priority Heuristic: Making Choices Without Trade-Offs," *Psychological Review*, 2006, *113* (2), 409–432.

——, ——, and ——, "Risky Choice with Heuristics: Reply to Birnbaum (2008), Johnson, Schulte-Mecklenbeck, and Willemsen (2008), and Rieger and Wang (2008)," *Psychological Review*, 2008, *115* (1), 281–290.

Camerer, C. F. and R. M. Hogarth, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty*, 1999, *19* (1-3), 7–42.

Camerer, Colin, "Individual Decision Making," *Handbook of Experimental Economics*, 1995, pp. 587–616.

Campitelli, G. and M. Labollita, "Correlations of Cognitive Reflection with Judgments and Choices," *Judgment and Decision Making*, 2010, *5* (3), 182–191.

Carson, Richard T. and Theodore Groves, "Incentive and Informational Properties of Preference Questions," *Environmental and Resource Economics*, 2007, *37*, 181–210.

Chilton, S.M. and W.G. Hutchinson, "A Qualitative Examination of How Respondents in a Contingent Valuation Study Rationalize Their WTP Responses to an Increase in the Quantity of the Environmental Good," *Journal of Economic Psychology*, 2003, *24*, 65–75.

Cokely, Edward T. and Colleen M. Kelley, "Cognitive Abilities and Superior Decision Making Under Risk: A Protocol Analysis and Process Model Evaluation," *Judgment and Decision Making*, February 2009, *4* (1), 20–33.

Costa-Gomes, M, VP Crawford, and B Broseta, "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 2001, pp. 1193–1235.

Croson, Rachel and Uri Gneezy, "Gender Differences in Preferences," *Journal of Economic Literature*, 2009, *47* (2), 448–474.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde, "Are Risk Aversion and Impatience Related to Cognitive Ability?," *American Economic Review*, 2010, *100* (3), 1238–1260.

Eckel, Catherine C. and Philip J. Grossman, "Men, Women, and Risk Aversion: Experimental Evidence," *Handbook of Experimental Economics: Results 1, Volume 1*, 2008, pp. 1061–1073.

Falk, Armin and James J. Heckman, "Lab Experiments Are a Major Source of Knowledge in the Social Sciences," *Science*, 2009, *326*, 535–538.

Ford, J. K., N. Schmitt, S. L. Schechtman, B. M. Hults, and M. L. Doherty, "Process Tracing Methods - Contibutions, Problems, and Neglected Research Questions," *Organizational Behavior and Human Decision Processes*, 1989, *43* (1), 75–117.

Franco-Watkins, Ana M. and Joseph G. Johnson, "Decision Moving Window: Using Interactive Eye Tracking to Examine Decision Processes," *Behavior Research Methods*, 2011, *43*, 853–863.

Frederick, S, "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 2005, pp. 25–42.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg, "Costly Information Acquisition: Experimental Analysis of a Bounded Rational Model," *American Economic Review*, September 2006, *96* (4), 1043–1068.

Gigerenzer, Gerd and Wolfgang Gaissmaier, "Heuristic Decision Making," *Annual Review of Psychology*, 2011, *62*, 451–82.

___ , Peter M. Todd, and the ABC Research Group, *Simple Heuristics That Make Us Smart*, Oxford University Press, 1999.

Glöckner, Andreas and Ann-Katrin Herbold, "An Eye-tracking Study on Information Processing in Risky Decisions: Evidence for Compensatory Strategies Based on Automatic Processes," *Journal of Behavioral Decision Making*, 2011, *24*, 71–98.

Green, Donald, Karen E. Jacowitz, Daniel Kahneman, and Daniel McFadden, "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods," *Resource and Energy Economics*, 1998, *20*, 85–116.

Harbaugh, William T., Kate Krause, and Lise Vesterlund, "Prospect Theory in Choice and Pricing Tasks," *University of Oregon Working Paper*, 2002.

___ , ___ , and ___ , "Risk Attitudes of Children and Adults: Choices Over Small and Large Probability Gains and Losses," *Experimental Economics*, 2002, *5*, 53–84.

Harrison, Glenn W., *Hypothetical Bias Over Uncertain Outcomes*, Northampton, MA: Edward Elgar, 2006.

___ and Elisabet Rutström, *Risk Aversion in the Laboratory*, Vol. 12, JAI Press,

___ , Morten I. Lau, and E. Elisabet Rutström, "Estimating Risk Attitudes in Denmark: A Field Experiment," *Scandinavian Journal of Economics*, 2007, *109* (2), 341–368.

Harrison, GW, E Johnson, MM McInnes, and EE Rutström, "Risk aversion and incentive effects: Comment," *American Economic Review*, 2005, pp. 897–901.

Hey, John D. and Chris Orme, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 1994, *62* (6), 1291–1326.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *American Economic Review*, 2002, *92* (5), 1644–55.

___ and ___ , "Risk Aversion and Incentive Effects: New Data without Order Effects," *American Economic Review*, 2005, *95* (3), 902–912.

Johnson, EJ, M Schulte-Mecklenbeck, and MC Willemsen, "Process Models Deserve Process Data: Comment on Brandstatter, Gigerenzer, and Hertwig (2006)," *Psychological Review*, 2008, pp. 263–272.

Kahneman, D and A Tversky, "Prospect Theory - Analysis of Decision Under Risk," *Econometrica*, 1979, pp. 263–291.

Kahneman, Daniel, *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011.

___ **and Amos Tversky**, *Judgment Under Uncertainty: Heuristics and Bias*, New York: Cambridge University Press, 1982.

Kang, Min Jeong, Antonio Rangel, Mickael Camus, and Colin F. Camerer, "Hypothetical and Real Choice Differentially Activate Common Valuation Areas," *The Journal of Neuroscience*, 2011, *31* (2), 461–468.

Kuhberger, A, M Schulte-Mecklenbeck, and J Perner, "Framing decisions: Hypothetical and real," *Organizational Behavior and Human Decision Processes*, 2002, pp. 1162–1175.

Lévy-Garboua, Louis, Hela Maafi, David Masclet, and Antoine Terracol, "Risk Aversion and Framing Effects," *Experimental Economics*, 2011, *Online First.*

Lohse, G. L. and E. J. Johnson, "A Comparison of Two Process Tracing Methods for Choice Tasks," *Organizational Behavior and Human Decision Processes*, 1996, *68* (1), 28–43.

Murphy, James J. and Thomas H. Stevens, "Contingent Valuation, Hypothetical Bias, and Experimental Economics," *Agricultural and Resource Economics Review*, 2004, *33* (2), 182–192.

Newell, A. and H.A. Simon, *Human Problem Solving*, Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Oechssler, J, A Roider, and PW Schmitz, "Cognitive abilities and behavioral biases," *Journal of Economic Behavior & Organization*, 2009, pp. 147–152.

Payne, J. W., "Task Complexity and Contingent Processing in Decision-Making - Information Search and Protocol Analysis," *Organizational Behavior and Human Performance*, 1976, *16* (2), 366–387.

___ **and M. L. Braunstein**, "Risky Choice - Examination of Information Acquisition Behavior," *Memory & Cognition*, 1978, *6* (5), 554–561.

——— **, J. R. Bettman, and E. J. Johnson**, “Adaptive Strategy Selection in Decision-Making,” *Journal of Experimental Psychology-Learning Memory and Cognition*, 1988, *14* (3), 534–552.

Payne, John W., James R. Bettman, and Eric J. Johnson, *The Adaptive Decision Maker*, New York: Cambridge University Press, 1993.

Rayner, Keith, “Eye Movements in Reading and Information Processing: 20 Years of Research,” *Psychological Bulletin*, 1998, *124* (3), 372–422.

Reisen, N., U. Hoffrage, and F. W. Mast, “Identifying decision strategies in a consumer choice situation,” *Judgment and Decision Making*, 2008, *3* (8), 641–658.

Rosen, L. D. and P. Rosenkoetter, “Eye Fixation Analysis of Choice and Judgment with Mulitattribute Stimulit,” *Memory & Cognition*, 1976, *4* (6), 747–752.

Schkade, David A. and John W. Payne, “How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation,” *Journal of Environmental Economics and Management*, 1994, *26*, 88–109.

Simon, Herbert A., *Reason in Human Affairs*, Stanford, CA: Stanford University Press, 1983.

Stanovich, Keith E. and Richard F. West, “Individual Differences in Reasoning: Implications for the Rationality Debate?,” *Behavioral and Brain Sciences*, 2000, *23*, 645–726.

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen, “Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam,” *American Economic Review*, 2010, *100* (1), 557–571.

Toplak, Maggie E., Richard F. West, and Keith E. Stanovich, “The Cognitive Reflection Test as a Predictor of Performance on Heuristics-and-Biases Tasks,” *Memory & Cognition*, 2011, *39* (7), 1275.

Tversky, Amos and Daniel Kahneman, “Advances in Prospect Theory: Cumulative Representation of Uncertainty,” *Journal of Risk and Uncertainty*, 1992, *5*, 297–323.

Weller, Joshua A., Nathan Dieckmann, Martin Tusler, CK Mertz, and Ellen Peters, “Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach,” *Unpublished manuscript*, 2011.

Wilcox, N. T., “Lottery Choice - Incentives, Complexity and Decision time,” *Economic Journal*, 1993, *103* (421), 1397–1417.

Willemsen, Martijn C. and Eric J. Johnson, "Visiting the Decision Factory: Observing Cognition with MouselabWEB and Other Information Acquisitions Methods," in Michael Schulte-Mecklenbeck, Anton Kuhberger, and Rob Ranyard, eds., *A Handbook of Process Tracing Methods for Decision Making*, Taylor & Francis, 2010, pp. 21–42.