Decision Making under Uncertainty: An Experimental Study in Market Settings

Federico Echenique   Taisuke Imai   Kota Saito [*]

December 6, 2019

Abstract

We design and implement a novel experimental test of subjective expected utility theory and its generalizations. Our experiments are implemented in the laboratory with a student population and pushed out through a large-scale panel to a general sample of the U.S. population. We find that a majority of subjects' choices are consistent with the maximization of *some* utility function, but not with subjective expected utility theory. The theory is tested by gauging how subjects respond to price changes. A majority of subjects respond to price changes in the direction predicted by the theory, but not to a degree that makes them fully consistent with subjective expected utility. Surprisingly, maxmin expected utility adds no explanatory power to subjective expected utility.

Our findings remain the same regardless of whether we look at laboratory data or the panel survey, even though the two subject populations are very different. The degree of violations of subjective expected utility theory is not affected by age nor cognitive ability, but it is correlated with financial literacy.

1 Introduction

We present an empirical investigation of the most widely used theories of decision under uncertainty, including subjective expected utility and maxmin expected utility. We consider economic environments, where an agent has to choose a portfolio of state-dependent payoffs, given state prices and a budget. Such environments are ubiquitous in economic theory, where agents choose a portfolio of Arrow-Debreu securities in complete markets. In our study, we record subjects' choices in a laboratory setting, and in a large-scale field panel. In consequence, we obtain results for very different populations, ranging from undergraduate students to older retirees, acting in economic environments that resemble real-world financial decisions. Our data allow us to see if subjects' demographic characteristics, such as age, income, and education, as well as cognitive ability and financial literacy, are related to how well they comply with the theories. We can also relate the results of our experiment to traditional measures of ambiguity aversion. Finally, our experiments speak to the external validity of laboratory studies, since we can compare behaviors in the lab to the behaviors of a sample of the general U.S. population.

Subjective expected utility theory (SEU; Savage, 1954) is the standard model of decision making under uncertainty (that is, where states of the world are uncertain, and no objective probabilities are known). The theory postulates an agent that has a subjective probabilistic belief over states of the world, and who maximizes the expected utility with respect to this belief. The starting point of our analysis is a methodological innovation: a nonparametric test for SEU using data on market choices (Echenique and Saito, 2015). Our experiments were conducted with the purpose of recreating the economic settings that are commonly assumed in economic theory: the choice of a financial state-contingent portfolio under uncertainty. They were also designed to use the new nonparametric tests to gauge the empirical performance of SEU.

While SEU is the dominant theory of choice under uncertainty, it is well known to face empirical challenges. In an influential paper, Ellsberg (1961) suggested that many agents would not conform to SEU. The phenomenon he uncovered, known as the "Ellsberg paradox," suggests that agents may seek to avoid betting on uncertain events in ways that cannot be represented with a subjective probability. Such avoidance of uncertain bets is termed ambiguity aversion. The Ellsberg paradox is based on a thought experiment, however, using bets on drawing a particular colored ball from an urn. The Ellsberg paradox therefore assumes an abstract, artificial, choice environment. One of our contributions is to empirically assess SEU in an economic environment that resembles the real-world financial markets where economists routinely assume that SEU guides agents' choices.

To account for the Ellsberg paradox, researchers have developed generalizations of SEU. Gilboa and Schmeidler (1989) suggest that an agent in Ellsberg's example may have too little information to form a unique subjective belief, and hence entertains multiple subjective probabilities. Being ambiguity averse, the agent maximizes the minimal expected utility over all subjective probabilities she entertains. The resulting theory is called maxmin expected utility (MEU). On the other hand, Machina and Schmeidler (1992) postulate that agents may have a unique subjective probability, but not necessarily decide according to the expected utility with respect to the probability. Such agents are called probabilistically sophisticated.[1]

Our understanding of ambiguity aversion is incomplete. It has been identified in different contexts, and in different subject populations (Trautmann and van de Kuilen, 2015); but the literature has relied almost exclusively on the paradigm introduced by Ellsberg (1961), where agents are offered bets on the color of balls drawn from urns whose composition is not fully specified. The simple binary choice structure of Ellsberg makes it easy to identify violations of SEU through violations of the so-called "sure-thing principle" (postulates P2 and P4 of Savage, 1954). But the artificial nature of the experiment may question the external validity of its findings. Despite its difficulty, designing choice environment that are more "natural," while providing clean identification, is an important task in the empirical literature on ambiguity aversion (Baillon et al., 2018b). In our paper, we investigate deviations from SEU and MEU in economic environments, combining a novel experimental paradigm and measurement techniques that are inspired by recent work on revealed preference theory. We are also able to partially test for probabilistic sophistication.

Echenique and Saito (2015) provide a necessary and sufficient condition for an agent's behavior in the market to be consistent with (risk-averse) SEU. Chambers et al. (2016) provide a similar condition for MEU when there are two states of the world. Echenique et al. (2018) characterize "approximate" SEU by allowing for errors and thus relaxing the empirical content of the model. These revealed-preference characterizations provide nonparametric tests for SEU and MEU, as well as a measure quantifying "how much" a dataset deviates from these theories. While the cited studies focus mostly on establishing theoretical revealed-preference conditions, the main motivation of the current paper is to bring the theoretical machinery to actual choices people make in the face of uncertainty. Our empirical approach is *nonparametric* in the sense that we do not impose any specific functional form, such as CRRA or CARA. We do assume that agents are risk averse or risk neutral (they have a concave von-Neumann-Morgenstern utility).

The theoretical revealed-preference results assume data on an agent's behavior in the market: meaning a collection of purchases of Arrow-Debreu securities at different budget constraints.

---

[1] Machina and Schmeidler (1992) were motivated by paradoxes of choice under risk, not uncertainty.

This setting naturally translates into our experimental design, which follows the spirit of portfolio choice tasks introduced by Loomes (1991) and Choi et al. (2007), and later used in many other studies (e.g., Ahn et al., 2014; Carvalho and Silverman, 2019; Choi et al., 2014; Hey and Pace, 2014). Subjects in our experiments are asked to allocate "tokens" into two accounts. Each account has an associated exchange rate which converts tokens into actual monetary rewards. These exchange rates define a budget set. Two accounts correspond to two mutually exclusive events, and subjects are told that they receive payment based on the chosen allocation, and on the realized event. Importantly, subjects are provided no information regarding the probabilities of these events. We generate uncertainty from two different sources. The first source is the classical Ellsberg-style "urns and balls." The second source comes from simulated stock prices.

As said, we ran our experiments in the laboratory where we used undergraduate students as subjects, and on a large-scale panel where we recruited representative of the U.S. population. See Section 2.2 for details.

1.1 Overview of Results

Our main findings are that: 1) subjects are consistent with utility maximization and probabilistic sophistication, but not SEU[2]; 2) MEU has no added explanatory power to SEU; 3) demand responds to price in the direction predicted by SEU, but not enough to make the data consistent with SEU; 4) subjects in the lab and in the panel display the same patterns; and 5) correlations with demographics exist but are limited.

The main purpose of our study was to nonparametrically test theories of decision under uncertainty. We find that most subjects are utility maximizers (they satisfy the Generalized Axiom of Revealed Preference), and satisfy Epstein's (2000) necessary condition for probabilistic sophistication. But the news is not good for more restrictive theories. In our experiments, across lab and panel, the vast majority of subjects do not conform to SEU. This finding would be in line with the message of the Ellsberg paradox, except that pass rates for MEU are just as low as for SEU. In fact, in all of our sample, only one subject's choice is consistent with MEU but not SEU.

Observe that Epstein's (2000) probabilistic sophistication axiom, which is largely satisfied by subjects in our experiments, is related to Machina and Schmeidler's (1992) "Strong Comparative Probability"; a stronger version of Savage's P4 axiom. Strong Comparative Probability is an essential axiom for probabilistic sophistication. In fact, Epstein (2000) proves that a violation of his axiom leads to a violation of utility maximization assuming Strong Comparative Probability.

---

[2]Since we test a necessary condition for probabilistic sophistication, we can only say that subjects are *not inconsistent* with probabilistic sophistication. See Section 3.

One might conjecture that the theories could be reconciled with the data if one allows for small mistakes, but our measures of the distance of the data to being rationalizable do not suggest so. A more forgiving test is to check if price changes are negatively correlated with quantity changes: we refer to this property as "downward-sloping demand," and it bears a close connection to SEU (see Echenique et al. (2018) for details). The vast majority of subjects exhibit the downward-sloping demand property, at least to some degree (meaning that the correlation between price and quantity changes is negative), but not to the extent needed to make them fully consistent with SEU. The downward-sloping demand property is strongly correlated with our measure of the distance between the data and SEU, so there is a precise sense in which the degree of compliance with downward-sloping demand can be tied to the violations of SEU.[3]

Our panel experiment allows us to compare the distance to SEU between subjects with different sociodemographic characteristics. We find that distance to SEU is correlated with financial literacy, with more financially literate subjects being closer to SEU than less literate subjects; and gender differences, with males being closer to SEU than females. A notable finding is the absence of a significant correlation with factors that have been shown to matter for related theories of choice (Choi et al., 2014; Echenique et al., 2018). In particular, older subjects, subjects with lower educational backgrounds, and subjects with lower cognitive ability, do not necessarily exhibit lower degrees of compliance with SEU.

One final implication of our results is worth discussing. Our experiments included a version of the standard Ellsberg question. The distance to SEU, or the degree of compliance with downward-sloping demand, are not related to the answers to the Ellsberg question, but the variability of uncertainty in our market experiment is. The experiments included a treatment on the variability of the uncertain environment, specifically the variability in the sample paths of the stock price whose outcomes subjects were betting on. Subjects who were exposed to more variable uncertainty seem less ambiguity averse (in the sense of Ellsberg) than subjects who were exposed to less variable uncertainty.

1.2   Related Literature

Starting with an influential thought experiment by Ellsberg (1961), many studies have tested SEU and related models of decision making under uncertainty using data from laboratory experiments. Trautmann and van de Kuilen (2015) provide an overview of this large but still growing empirical literature. Typical experiments involve "urns and colored-balls" following Ellsberg's (1961)

---

[3]Again, Echenique et al. (2018) contains a detailed discussion.

original thought experiment, and individual's attitude towards ambiguity is inferred by looking at valuations or beliefs elicited through a series of binary choices (e.g., Abdellaoui et al., 2011; Baillon and Bleichrodt, 2015; Chew et al., 2017; Epstein and Halevy, 2019; Halevy, 2007). Other studies try to parametrically estimate the models under consideration (e.g., Ahn et al., 2014; Dimmock et al., 2015; Hey et al., 2010; Hey and Pace, 2014). Unlike these studies, our approach is nonparametric, imposing no assumptions on functional form other than risk-aversion. While the use of artificially generated ambiguity as in Ellsberg-style urns and balls has attractive features that make the interpretation of choice behavior, and experimental implementation, simple, it has been argued that researchers should not rely too much on a paradigm that uses an artificial source of ambiguity. Instead, one should study more "natural" sources of ambiguity.[4] In response to these concerns, several studies use non-artificial sources of ambiguity such as stock market indices and temperature (Abdellaoui et al., 2011; Baillon and Bleichrodt, 2015; Baillon et al., 2018a). Baillon et al. (2018b) introduce a method that elicits ambiguity attitudes for natural events while controlling for unobservable subjective likelihoods.

It is also important to note that there are several studies that try to understand the relationship between sociodemographic characteristics, ambiguity attitudes, and real-world behavior (especially financial).[5] This is a subset of a growing empirical literature that seeks to understand the common foundation of a wide class of (behavioral) preferences and to relate cross-/within-country heterogeneity and cultural or sociodemographic characteristics (e.g., Bianchi and Tallon, 2019; Bonsang and Dohmen, 2015; Dimmock et al., 2015, 2016a,b; Dohmen et al., 2018; Falk et al., 2018; Huffman et al., forthcoming; Sunde and Dohmen, 2016; Tymula et al., 2013).

Finally, the analysis of our data uses theoretical tools developed and discussed in Chambers et al. (2016), Echenique and Saito (2015), and Echenique et al. (2018). They require coupling SEU and MEU with risk-aversion. The methods in Polisson et al. (2017) avoid the assumption of risk-aversion, but are computationally hard to implement in the case of SEU (their paper contains an application to objective EU, for which their method is efficient). Polisson et. al. also develop a test for first-order stochastic dominance in models with known (objective) probabilities. Their test could be seen as a first step towards an understanding of probabilistic sophistication.

---

[4]For example, Camerer and Weber (1992) note that: "Experimental studies that do not directly test a specific theory should contribute to a broader understanding of betting on natural events in a wider variety of conditions where information is missing. There are diminishing returns to studying urns!" (p. 361). Similarly, Gilboa (2009) writes: "David Schmeidler often says, 'Real life is not about balls and urns.' Indeed, important decisions involve war and peace, recessions and booms, diseases and cures" (p. 136).

[5]Trautmann and van de Kuilen (2015) note the importance of this direction: "Interestingly, the empirical literature has so far provided little evidence linking individual attitudes toward ambiguity to behavior outside the lab. Are those agents who show the strongest degree of ambiguity aversion in some decision task also the ones who are most likely to avoid ambiguous investments?" (p. 89).

TABLE 1: Order of the tasks.

Platform	Treatment	Task 1	Task 2	Task 3	Task 4
Laboratory	Large volatility	Market-stock	Market-Ellsberg	Standard Ellsberg	Survey
	Small volatility	Market-stock	Market-Ellsberg	Standard Ellsberg	Survey
Panel	Large volatility	Market-stock	Standard Ellsberg	—	—
	Small volatility	Market-stock	Standard Ellsberg	—	—

2 Experimental Design

We conducted experiments at the Experimental Social Science Laboratory (ESSL) at the University of California, Irvine (hereafter *the lab*), and on the Understanding America Study (UAS) panel, a longitudinal survey platform (hereafter *the panel*).[6] The general structure of tasks in the lab and in the panel were the same. We shall first in Section 2.1 describe the basic tasks, which were common to the lab and the panel experiments. Then in Section 2.2 we turn to the features that were unique to each. Further details and instructions are presented in online appendices D and E.

2.1 Tasks

We first describe the two basic tasks used in our experiments: the market task (also referred to as the allocation task), and the Ellsberg two-urn choice task. The market task has two versions, depending on the source of uncertainty. The exact set of tasks differed somewhat depending on the platform: the lab or the panel. Table 1 summarizes the lab and the panel experiments.

Market task. The market task is meant to represent the most basic economic problem of choice under uncertainty. An agent chooses among Arrow-Debreu commodities, given state prices and a budget. Experimental implementations of such portfolio choice problems were introduced by Loomes (1991) and Choi et al. (2007), and later used in Ahn et al. (2014), Choi et al. (2014), and Hey and Pace (2014), among others.

Uncertainty is represented through a *state space* $\Omega = \{\omega_1, \omega_2, \omega_3\}$. For each choice problem there are two relevant *events*, denoted by E_s, $s = 1, 2$. Events are sets of states, which are lumped together in ways that will be clear below. The events E_1 and E_2 are mutually exclusive. Subjects are endowed with 100 (divisible) tokens in each round. An event-contingent payoff may be

---

[6]Our experiment was approved by the Institutional Review Board of California Institute of Technology (#15-0478). It was then reviewed and approved by the director of ESSL and the board of UAS. The module number of our UAS survey is 116.

purchased at a price, which experimentally is captured through an "exchange value." Exchange values, denoted z_s, $s = 1, 2$, relate tokens allocated to an event, and monetary outcomes. Given a pair of exchange values (z_1, z_2), subjects are asked to decide on the allocation of tokens, (a_1, a_2), between the two events. A subjects who decides on an allocation (a_1, a_2) earns $x_s = a_s \times z_s$ if event E_s occurs. The sets of exchange values (z_1, z_2) used in the experiments are presented in Table D.1 in online appendix.

An allocation (a_1, a_2) of tokens is equivalent to buying a x_s units of an Arrow-Debreu security that pays \$1 per unit if event E_s holds, from a budget set satisfying $p_1 x_1 + p_2 x_2 = I$, where prices and income (p_1, p_2, I) are determined by the token exchange values (z_1, z_2) in the round.[7]

Our design deviates from the other studies mentioned above by introducing a novel event structure. There are three underlying states of the world: ω_i, $i = 1, 2, 3$, and we introduce two *types* of questions. In Type 1 questions, event 1 is $E_1^1 = \{\omega_1\}$ and event 2 is $E_2^1 = \{\omega_2, \omega_3\}$. In Type 2 questions, event 1 is $E_1^2 = \{\omega_1, \omega_2\}$ and event 2 is $E_2^2 = \{\omega_3\}$. See Figure 1 for an illustration. This event structure requires SEU decision makers to behave consistently not only within each type of questions but also across two types of questions.[8]

The design allows us to examine one aspect of SEU rationality, monotonicity of choice. The monotonicity follows from the fact that SEU rational agent should consider event E_1^2 is more likely than event E_1^1 and, hence, the agent should allocate more tokens on event E_1^2 than on event E_1^1 if the prices are the same. We term this property *event monotonicity*. A detailed discussion follows later in the paper.

Subjects in our experiments make decisions through a computer interface. The *allocation table* on the computer screen contains all the information subjects need to make their decisions in each question; see right panels in Figure 1. The table displays exchange values (z_1, z_2) for the current question, their current allocation of tokens (a_1, a_2), and implied monetary value of each account, referred to as the "account value," $(a_1 \times z_1, a_2 \times z_2)$. Subjects can allocate tokens between two events using a slider at the bottom of the screen; every change in allocation is instantaneously reflected in the allocation table.[9]

An important feature of our design is that we implement the task under two different sources

---

[7]We set $p_1 = 1$ (normalization) and $p_2 = z_1/z_2$. Then, the income is given by $I = 100 \times z_1$.

[8]Hey and Pace's (2014) design is the closest to ours. In their experiment, uncertainty was generated by the colors of balls in the Bingo Blower and subjects were asked to make 76 allocation decisions in two different types. In the first type of problems, subjects were asked to allocate between two of the colors. In the second type, they were asked to allocate between one of the colors and the other two. Note that the motivation of Hey and Pace (2014) is a parametric estimation of leading models of ambiguity aversion.

[9]Tokens are divisible (the slider moves in the increment of 0.01). This ensures that the point on the budget line which equalizes the payouts in the two events (i.e., on the 45-degree line) is technically feasible.
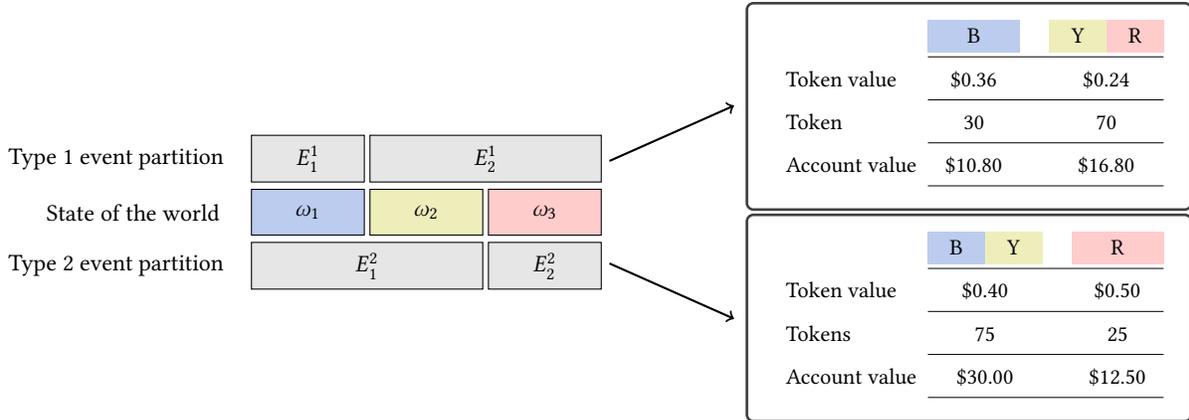
FIGURE 1: (Left) Event structure in two types of questions. (Right) Illustration of the allocation table for a type 1 question (top) and a type 2 question (bottom).

of uncertainty. Subjects face two versions of the market task, as we change the source of uncertainty. In the first version, called "market-Ellsberg", uncertainty is generated with an Ellsberg urn. In the second version, termed "market-stock", uncertainty is generated through a stochastic process that resembles the uncertain price of a financial asset, or a market index. The market-Ellsberg version follows Ellsberg (1961), and the empirical literature on ambiguity aversion (Trautmann and van de Kuilen, 2015). Subjects are presented with a bag containing 30 red, yellow, and blue chips, but they are not told anything about the composition of the bag. The three states of the world are then defined by the color of a chip drawn from the bag: state 1 (ω_1) corresponds to drawing a blue chip, state 2 (ω_2) corresponds to drawing a yellow chip, and state 3 (ω_3) corresponds to drawing a red chip.

In the market-stock task, uncertainty is generated through the realization of simulated stock prices. Subjects are presented with a history of stock prices, as in Figure 2.[10] The chart shows the evolution of a stock price for 300 periods; the next 200 periods are unknown, and left blank. Subject are told that prices are determined through a model used in financial economics to approximate real world stock prices. They are told that the chart represents the realized stock price up to period 300, and that the remaining periods will be determined according to the same model from financial economics. Let the price at period 300 be the "starting value" and the price at period 500 be the "target value." We define three states, given some threshold $R \in (0, 1)$: $\omega_1 = (R, +\infty)$,

---

[10]We used a Geometric Brownian Motion to simulate 100 stock price paths that share the common starting price and the time horizon. After visually inspecting the pattern of each price path, we handpicked 28 paths and then asked workers on Amazon Mechanical Turk what they believed the future price of each path would be. The elicited belief distributions were then averaged across subjects. Some price paths, especially those with clear upward or downward trends, tend to be associated with skewed distributions. Others have more symmetric distributions. We thus selected two relatively "neutral" ones from the latter set for the main experiment. See online appendix D.2.
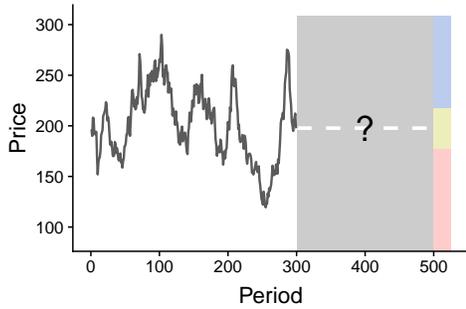
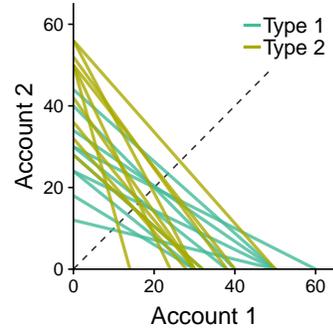FIGURE 2: Source of uncertainty in the market-stock task.

FIGURE 3: Set of 20 budgets.

in which the target value rises by more than $100R\%$ compared to the starting value (see the blue region in the figure), $\omega_2 = [-R, R]$, in which the price varies by at most $100R\%$ between the starting value and the target value (the yellow region in the figure), and $\omega_3 = [-1, -R)$, in which the target value falls by more than $100R\%$ compared to the starting value (the red region in Figure 2).

We chose token exchange values (z_1, z_2) for each question to increase the power of our tests. After running several choice simulations to calculate the power of our tests, we select 20 budgets (10 for type 1, 10 for type 2) shown in Figure 3 (and Table D.1 in online appendix). Note that event 1 is "more likely" in type 2 decision problems since $\{\omega_1\} = E_1^1 \subseteq E_1^2 = \{\omega_1, \omega_2\}$. In constructing budget sets, we made assets in account 1 more "relatively expensive" than assets in account 2 in type 2 questions. This is reflected in the steeper slope of the budget lines presented in Figure 3.

Several remarks about our experimental design are in order. First, we use the movement of stock prices as a source of uncertainty, not balls and urns. We are not the first to use financial information as the source of uncertainty (see Abdellaoui et al., 2011; Baillon and Bleichrodt, 2015; Baillon et al., 2018a), but it is still rare in the experimental literature. Second, subjects were allowed to make fractional allocations of tokens between accounts. Our fractional allocation design sought to mimic choices from a continuous budget line, as in the theoretical models we try to test. Third, we asked two types of allocation decisions. This makes our task demanding for subjects, but it creates a powerful environment for our revealed preference analysis, and allows for natural within-subject comparisons.

Ellsberg two-urn choice task. In addition to the market task described above, we presented our subjects with a standard two-urn version of Ellsberg's (1961) choice question. The purpose of including this standard task is to compare the behavior of subjects in the different designs. By this comparison, we can investigate how traditional evaluations of ambiguity aversion relate to market choices, and see if the market setting affects subjects' attitude toward uncertainty.

Subjects confront two bags: bag A and bag B, each of which contains 20 chips. They receive the following information: Bag A contains 10 orange chips and 10 green chips. Bag B contains 20 chips, each of which is either orange or green. The number of chips of each color in bag B is unknown to them, so there can be anywhere from 0 to 20 orange chips, and anywhere from 0 to 20 green chips, as long as the total number of orange and green chips sums to 20.

Subjects were offered choices between bets on the color of the chip that would be drawn at the end of the experiment. Before choosing between bets, subjects were first asked to choose a fixed color (orange or green; called "Your Color") for which they would be paid if they chose certain bets. They were then asked three questions.[11]

The first question asks to choose between a bet that pays $X+b if the color of the ball drawn from bag A is "Your Color" (and nothing otherwise), and a bet that pays $X if the color of a ball drawn from bag B is "Your Color" (and nothing otherwise). Similarly, the second question asks to choose between a bet that pays $X if the color of the ball drawn from bag A is "Your Color," and a bet that pays $X if the color of a ball drawn from bag B is "Your Color". Finally, the third question asks to choose between a bet that pays $X if the color of the ball drawn from bag A is "Your Color" and a bet that pays $X+b if the color of a ball drawn from bag B is "Your Color". The payoff X and the bonus b depended on the platform: $(X, b) = (10, 0.5)$ in our lab study and $(X, b) = (100, 5)$ in the panel. In our lab experiments, the content of bag B had already been determined at the beginning of the experiment by an assistant. The timing is important to ensure that there is no incentive to hedge (Baillon et al., 2015; Epstein and Halevy, 2019; Saito, 2015). The subjects were allowed to inspect the content of each bag after completing the experiment.

Post-experiment survey. In the lab experiment, subjects were asked to fill out a short survey asking for their age, gender, major in college, the three-item cognitive reflection test (CRT; Frederick, 2005), and strategies they employed in the allocation tasks if any (see online appendix D.3). In the panel study, subjects answered a standard questionnaire that the Understanding America Study (UAS) asks of all its panelist households.

2.2   Implementation

Interface. We prepared an experimental interface that runs on a web browser. In the panel study, our interface was embedded in the survey page of the UAS. Therefore, subjects in both the lab and panel experiments interacted with the exact same interface.

---

[11]We adopted the three-question setting akin to Epstein and Halevy (2019), as a way of identifying strict ambiguity preferences. The typical Ellsberg-style experiment would ask only one question, namely the second one.
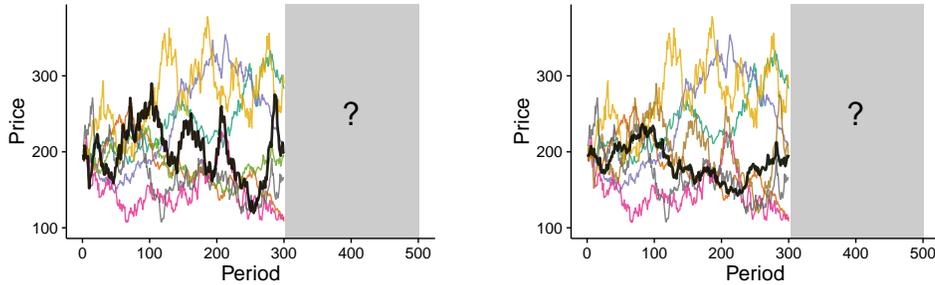
FIGURE 4: Context of market information: large volatility (left) and small volatility (right).

Recruiting and sampling. Subjects for our lab study were recruited from a database of undergraduate students enrolled in the University of California at Irvine. The recruiting methodology for the UAS survey is described in detail in the survey website.[12] Within the UAS sample, we drew a stratified random sub-sample with the aim of obtaining a representative sample of subjects in different age cohorts. In particular, we recruited subjects in three age groups: from 20 to 39, from 40 to 59, and 60 and above, randomly from the pool of survey participants. The purpose of stratifying the sample was to be able to assess the relation between age and pass rates for our revealed-preference tests.

Treatments. In the market-stock task, we prepared two simulated paths of stock prices with different degree of volatility, so that one path seems relatively more volatile than the other, while keeping the general trend in prices as similar as possible between the two paths. Since the perception of volatility is only relative, we embed each path in the common market "context" as shown in Figure 4. Here, the bold black lines indicate the stock under consideration, and the other lines in the background are the same in the two treatments.

Our treatment variation is the perceived volatility of simulated stock prices (we call the two treatments Large and Small). The subjects were randomly assigned to either a large volatility condition (left panel in Figure 4), or a small volatility condition (right panel).[13] The instructions for the market-stock task included one of the two charts of Figure 4, depending on the treatment (see online appendix E).

Order of the tasks. Subjects in the lab study performed three tasks in the following order: market-stock, market-Ellsberg, and standard Ellsberg. Subjects in the Panel study performed two

---

[12]https://uasdata.usc.edu/index.php.

[13]In the lab study, random assignment to one of the two treatments was done at the session level, meaning that all subjects in the same session were shown the same price path.

tasks, market-stock and standard Ellsberg, but due to time constraints we did not implement market-Ellsberg in the panel. Table 1, which has a summary of the structure of the experiments and treatments, lists the order in which the tasks were completed.

Incentives. In the lab study, we used the standard incentive structure of paying-one-choice-at-random. Subjects received a sealed envelope when they entered the laboratory room. The envelope contained a piece of paper, on which two numbers were written. The first number indicated the task number, and the second number indicated the question number in that task. Both numbers were randomly selected beforehand. At the end of the experiment, subjects brought the envelope to the experimenter's computer station. If the selected task was the market task with stock price information, the simulated "future" price path was presented on the screen. If, on the other hand, the selected task involved the Ellsberg urn, the subject was asked to pick one chip from the relevant bag. All subjects received a $7 showup fee.

In the panel study, four subjects were randomly selected to receive the bonus payment based on their choices in the experiments. Unlike the lab study, the bonus payment for these subjects was determined by a randomization implemented by the computer program, but payments were of a much larger scale. All subjects received a participation fee of $10 by completing the entire survey.

3 Results

This section presents results from the lab and the panel. For each dataset, we first discuss the basic patterns of subjects' choices, and then proceed to present our revealed-preference tests.

3.1 Results from the Lab

We conducted seven experimental sessions at the Experimental Social Science Laboratory (ESSL) of the University of California, Irvine. A total of 127 subjects (62 in treatment Small and 65 in treatment Large; mean age = 20.16, SD = 1.58; 35% male) participated in the study.[14] Each

---

[14]Three additional subjects participated in the study, but we excluded their data from the analysis. One subject accidentally participated in two sessions (thus, the data from the second appearance was excluded). Two subjects spent a significantly longer time on each decision than anyone else. We distributed the instructions for each task of the experiment just before they were to perform that task, meaning that each subject would have to wait until all the other subjects in the session completed the task. We had to "nudge" these two subjects that were extremely slow to make decisions more quickly, and hence eliminated their choices from our data.

session lasted about an hour, and subjects earned on average \$21.3 (including a \$7 showup fee, SD = 9.21).

Choices in the market tasks. Subjects faced budgets in random order, with one exception. The exception is related to event monotonicity, as discussed earlier. We fixed two consecutive questions: questions #5 and #6, that had the same budget set, but with different event structures. These were the only questions that were not presented in random order. Our purpose was to check that subjects had a basic understanding of the task. The 5th question was presented as a type 1 question while the 6th question was presented as a type 2 question (recall the terminology from Section 2). Since the event upon which the first account pays off is a larger set in question #6 than in question #5 ($\{\omega_1\} = E_1^1 \subseteq E_1^2 = \{\omega_1, \omega_2\}$ by construction), while prices and budget remain the same, subjects should allocate more to the first account in question #6 than in question #5.

More than 70% of subjects satisfy event monotonicity, and this number increases to 90% if we allow for a small margin of error of five tokens. Moreover, choices are clustered around the allocation which equalizes payout from the two accounts, which can be interpreted as the subjects' ambiguity aversion. See Figure C.2 in the online appendix.

As Echenique et al. (2018) discuss in depth, the empirical content of expected utility is captured in part by a negative relation between state prices and allocations: a property that can be thought of as "downward-sloping demand." We thus look at how subjects' choices responded to price variability between budgets; in particular, we focus on the relation between price ratios, $\log(p_2/p_1)$, and allocation ratios, $\log(x_2/x_1)$, pooling choices from all subjects. Figure 5 shows a negative relation between these two quantities, confirming the downward-sloping demand property at the aggregate level. It holds for both types of questions and in both tasks.

We also quantify the degree of compliance with the downward-sloping demand property at the individual level by calculating correlation ρ^{dsd} between $\log(p_2/p_1)$ and $\log(x_2/x_1)$.[15] A significant majority of the subjects made choices that responded to prices negatively ($\rho^{\text{dsd}} < 0$; Figure C.3 in the online appendix).

Revealed-preference tests. Did the subjects in our experiment make choices that are consistent with basic economic models of utility maximization, including the standard subjective expected utility (SEU) theory? In order to answer this question, we implement nonparametric, revealed-preference tests on each individual subject's choice data. These tests include: GARP,

---

[15]Let ρ_t be the (Spearman's) correlation coefficient in type t questions. We obtain the "average" correlation coefficient, ρ^{dsd}, by Fisher's z-transformation $\rho^{\text{dsd}} = \tanh\left(\sum_{t=1}^{2} \tanh^{-1}(\rho_t)/2\right)$.
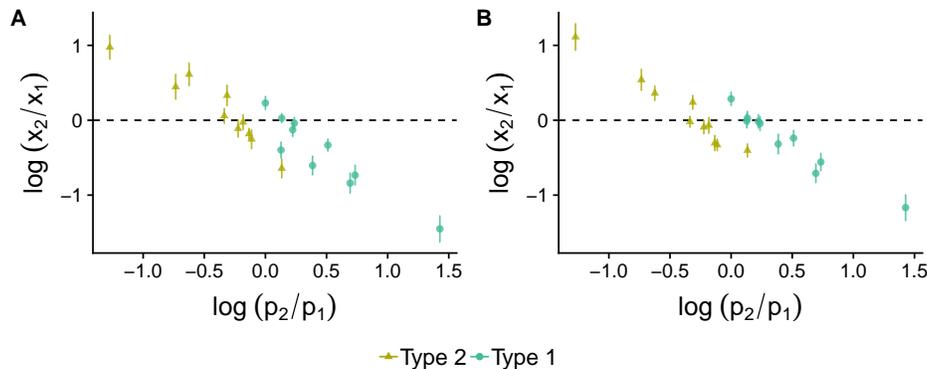
FIGURE 5: Downward-sloping demand in the market-stock task (A) and the market-Ellsberg task (B). Each dot represents mean $\log(x_2/x_1)$ at each $\log(p_2/p_1)$ and bars indicate standard error of means.

TABLE 2: Pass rates.

	GARP	PS	SEU	MEU
Market-stock	0.5827	0.4803	0.0000	0.0000
Market-Ellsberg	0.6693	0.6220	0.0157	0.0157

Note: Since Epstein's (2000) condition is only necessary for probabilistic sophistication, the numbers reported here capture the fraction of the subjects who are *not inconsistent* with probabilistic sophistication.

probabilistic sophistication (hereafter PS; Machina and Schmeidler, 1992), SEU (based on and extended from Echenique and Saito, 2015), and MEU (based on Chambers et al., 2016).[16]

Recall that, depending on how we partition the state space, we have two types of decision problems. For GARP and PS, we first test each type of problem separately and then combine the results. We say that a subject's data satisfies GARP if it passes the GARP test for both types. Similarly, we say that a subject's data is not inconsistent with PS if it is not inconsistent with PS in the sense of Epstein's (2000) condition for both types, and satisfies event monotonicity. For SEU and MEU, we implement the test directly on the data combining two types of problems. It is not obvious that this can be done. That the two types of problems can be combined, effectively testing the three-state design using bets on two events at a time, is one contribution of our paper: see online appendices A and B for details.

Table 2 presents the *pass rate* of each test. That is, the fraction of subjects (out of 127) who

---

[16]We can test whether a given dataset is consistent with SEU or MEU by solving the linear programming problem implied by the axiom characterizing the model. We say that a dataset passes the test if there is a solution to the problem. The construction of such linear programming problems closely follows the argument in the proofs of Theorems that appeared in Echenique and Saito (2015) and Chambers et al. (2016). The key step is to assemble a system of log-linearized Afriat inequalities.

passed each test. We find that a majority of subjects satisfy GARP, meaning that their choices are consistent with the maximization of *some* utility function. On the contrary, subjects clearly did not make choices that are consistent with SEU. The SEU pass rates are below 0.1, and not a single agent passed the SEU test in the market-stock task.[17]

Perhaps surprisingly, allowing for multiple priors via MEU does not change the result. Pass rates for MEU are the same as for SEU, implying that *MEU does not capture violations of SEU in our experiment.* These findings are consistent with data from the experiment in Hey and Pace (2014): see Chambers et al. (2016), which performs the same kind of analysis as we do in the present paper for Hey and Pace's (2014) data.

Finally, we look at PS to investigate whether observed behavior is (in)consistent with preferences being based on probabilities, using the necessary condition proposed by Epstein (2000) and checking event monotonicity in questions #5 and #6. We find that 48% to 62% of subjects are not inconsistent with PS.

Distance measures. The Critical Cost Efficiency Index (CCEI; Afriat, 1972) is a measure of the degree of compliance with GARP. It is heavily used in the recent experimental literature to gauge how close subjects are to being rational economic agents (e.g., Choi et al., 2014). In our lab data, the average CCEI is above 0.98, which implies that on average budget lines needed to be shifted down by about two percent to eliminate a subject's GARP violations (Table 3). The CCEI scores reported in Table 3 are substantially higher than those reported in Choi et al. (2014), but close to the CCEI scores in Choi et al. (2007). This would seem to indicate a higher level of compliance with utility maximizing behavior than in the experiment by Choi et al. (2014), and about the same as the experiment by Choi et al. (2007). Note, however, that there are several substantial differences in the settings and the designs between the two aforementioned studies and ours. We had two types of events (other studies typically have one fixed event structure), each type involved 10 budgets (i.e., total 20 budgets) while the aforementioned studies had 25 and 50 budgets respectively. Most importantly, objective probabilities were not provided in our study.

The pass rates for SEU are very small, but it is possible that small mistakes could account for a subjects' violation of SEU. We turn to a measure of the severity of violations of SEU. Table 3 reports e_* (*minimal e*), a measure of the degree of deviation from SEU proposed by Echenique et al. (2018). The number e_* is a perturbation to the model that allows SEU to accommodate the data: It

---

[17]Along similar lines, Echenique et al. (2018) find that only five out of more than 3,000 participants in three online surveys (Carvalho et al., 2016; Carvalho and Silverman, 2019; Choi et al., 2014) make choices that are consistent with *objective* expected utility theory.

TABLE 3: Distance measures.

| Task | CCEI | | | e_* (SEU) | | | e_* (MEU) | | |
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Market-stock	0.9805	1.0000	0.0450	0.8508	0.8238	0.3749	0.8448	0.8148	0.3784
Market-Ellsberg	0.9892	1.0000	0.0317	0.7631	0.6931	0.3834	0.7605	0.6931	0.3839
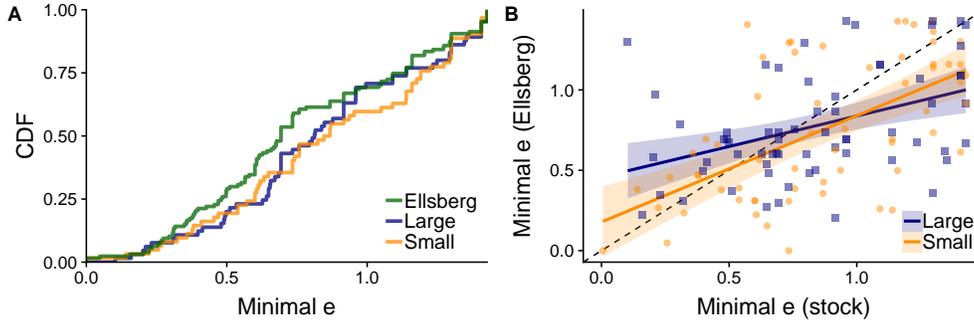
FIGURE 6: Comparing e_* across tasks.

can be interpreted as the size of a utility or belief "error" that can rationalize the observed choices. Thus, the number e_* is zero when data are consistent with SEU, meaning that no perturbation is needed to rationalize the data by means of SEU, but takes a positive value if data violate SEU. The larger is e_*, the larger is the size of the perturbation needed to rationalize data by means of a perturbed version of SEU. See Echenique et al. (2018) for the theoretical background and online appendix B.2 for implementation details.

We find that e_* in the market-stock task is significantly higher than in the market-Ellsberg task (paired-sample t-test; $t(126) = 2.635$, $p = 0.009$). See also Figure 6A. This finding suggests that subjects made choices that were closer to SEU when the source of information was an Ellsberg urn than when the source was a stock price, but the result has to be qualified because the order of the two market tasks was not counterbalanced.

In the two market tasks, subjects faced the same set of 20 budgets in random order, with the exception of two budgets for which the order was fixed (see below). The choices made by about three-quarters of the subjects are positively correlated between the two tasks (Figure C.1 in the online appendix), and 36% of those subjects exhibit statistically significant positive correlation (one-sided, $p < 0.05$). This correlation is reflected in the degree of violation of SEU—Figure 6B shows that e_* from two tasks are highly correlated (Spearman's correlation coefficient: $r = 0.406$ for treatment Large, $r = 0.583$ for treatment Small).

We also find that e_* and the downward-sloping demand property (specifically, the aggregate correlation coefficient between price and quantity, as described above) are closely related; see Figure C.5 in the online appendix. The subjects' e_* tend to be large when their choices do not respond to price changes, indicating larger deviations from SEU. This is particularly true when the subjects are choosing allocations that are close to the 45-degree line in order to hedge against uncertainty. On the contrary, CCEI can be (close to) one even when choices are not responding to price changes (Figure C.6). The connection between e_* and downward sloping demand (and the disconnection with CCEI) is natural. Compliance with SEU requires a certain kind of downward-sloping demand property. Our subjects largely display quantity-price responses that go in the direction predicted by SEU, but not to the degree that the theory demands.

Table 3 also shows that the data is not much closer to MEU than to SEU. The MEU model has little added explanatory power beyond SEU. In other words, the way in which subjects' choices deviate from SEU is not captured by the MEU model. In MEU, agents' beliefs can depend on choices, as in the perturbation of the SEU model behind our calculaiton of e_*. However, in MEU, the dependency is specific: beliefs are chosen so as to minimize expected utility. Our finding suggests that subjects' beliefs may depend on choices, but not determined pessimistically. Therefore, the MEU model cannot explain the subjects' choices better than SEU; the needed size of perturbation for MEU is not much lower than that of SEU.

We do not observe gender differences on e_*. We do, however, observe an effect of cognitive ability as measured with the three-item Cognitive Reflection Test (CRT; Frederick, 2005). Subjects who answered all three questions correctly exhibit lower e_* than those who answered none of them correctly. This effect is statistically significant only in the market-stock task (Figure C.7, Table C.2).

Ambiguity attitude. Finally, we look at the relation between behavior in the market tasks and subjects' attitudes toward ambiguity, measured with a standard Ellsberg-paradox design. As explained in Section 2.1, we asked three questions regarding choices between an ambiguous bet and a risky bet to identify subjects' attitude toward ambiguity. Figure 7 shows the frequency with which subjects preferred to bet on the risky urn, for each question.

In the first question, the risky bet pays an additional $0.5 in case of winning. This bonus made almost all (95.3%) subjects choose the risky bet. The third question has instead a bonus for choosing the ambiguous bet, which then pays an additional $0.5 in case of winning. A little more than half of the subjects (61.5% in the Large treatment, 53.2% in the Small treatment) preferred the risky bet, but the difference from 50% (i.e., indifference at the aggregate level) is not significantly
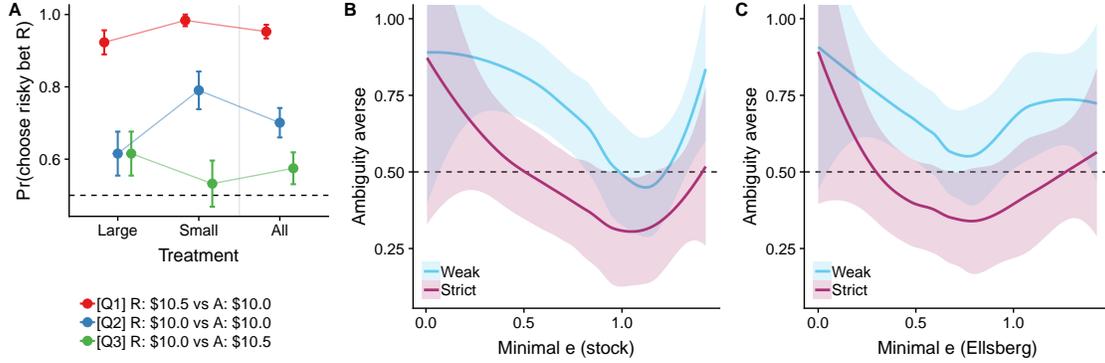
FIGURE 7: (A) Probability of choosing a risky bet in each question in the standard-Ellsberg task in the lab data. Bars indicate standard errors of means. (BC) LOESS curves relating e_* and ambiguity attitude.

large (z-test for proportion; $p = 0.063$ in the Large treatment and $p = 0.612$ in the Small treatment). In the second question, which pays the equal winning prize in the two bets (as in many other Ellsberg-style studies), subjects in the Small treatment chose the risky bet more frequently than those in the Large treatment (61.5% in the Large treatment and 73.0% in the Small treatment; two-sample z-test for proportion, $p = 0.031$).

We classify subjects as *weakly ambiguity averse* if they chose the risky bet, both in the first and in the second question (68.5% of the subjects). Similarly, we classify subjects as *strictly ambiguity averse* if they chose the risky bet in all three questions (44.1% of the subjects). In order to connect the deviation from SEU captured by e_* and a measure of ambiguity attitude standard in the literature, we nonparametrically estimate how the probability of being classified as ambiguity averse depends on e_*. Figure 7BC suggest a weak but quadratic relationship between these two. Ambiguity aversion is the leading explanation for violations of SEU, so our finding may seem counter-intuitive. One might instead expect a monotonic relation between e_* and ambiguity aversion. It is, however, important to emphasize that e_* captures *any* deviation from SEU. Not only those that could be traced to ambiguity aversion.

3.2 Results from the Panel

A total of 764 subjects (mean age = 50.26, SD = 15.45; 50.4% male) completed the study. The median survey length was 29.1 minutes. In addition to $10 baseline payment for completing the survey, four randomly selected subjects received additional payment from one of the choices they made during the survey (average $137.56).

We tried to get subjects to do our experiment on a desktop or laptop computer, but many of them took it with their mobile devices—such as smartphones or tablets. These devices have
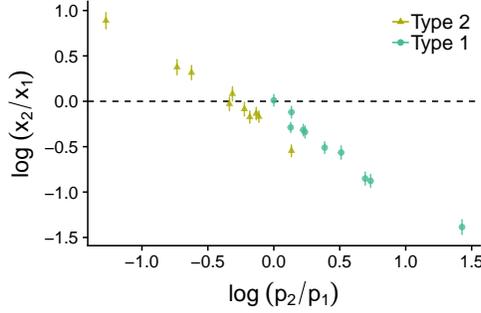
FIGURE 8: Downward-sloping demand. Each dot represents mean $\log(x_2/x_1)$ at each $\log(p_2/p_1)$ and bars indicate standard error of means.

screens that are smaller than desktop/laptop computers, which makes it quite difficult to understand our experiments, and perform the tasks we request them to complete. We thus analyze the data following three inclusion criteria, (i) desktop/laptop computer only (66%), (ii) desktop/laptop computer and tablet (76%), and (iii) all devices combined. We treat the first as the "core" sample. Table 4 provides summary statistics of individual sociodemographic characteristics across the three inclusion criteria. We present the entire sample as well as the core sample (those who used desktop/laptop computers), and the excluded sample (those who did not use desktop/laptop computers). It is evident that the type of device used is correlated with some of the demographic variables (age: $\chi^2(2) = 17.79$, $p < 0.001$; education level: $\chi^2(3) = 53.70$, $p < 0.001$; income level: $\chi^2(4) = 43.97$, $p < 0.001$). The sub-samples of subjects exhibited markedly different patterns of behavior as well. Throughout the rest of the paper, we analyze data from the core sample.[18]

The set of 20 budgets used in the market task is the 10-times scaled-up version of the one used in the lab (Figure 3). This keeps the relative prices the same between two studies, making the distance measure e_* comparable between data from the lab and the panel.[19]

We start by checking event monotonicity, along the lines of our discussion for the lab experiment. Our subjects' choices on questions #5 and #6 are informative about how attentive they are when they perform the tasks in our experiment. We find that about 60% of subjects satisfy event monotonicity, and that this number jumps to 78% if we allow for a margin of error of five tokens (see Figure C.8). There is no treatment difference. Our subjects also made choices that are, to some extent, responding to underlying price changes (Figure 8).

---

[18]Results from the same set of analyses on the entire subjects, or comparison across sub-samples, are available upon request.

[19]The distance is measured in units of relative price.

TABLE 4: Sociodemographic information.

Variable	All	Device Desktop/laptop	Other
Gender			
Male	0.504	0.529	0.456
Age group			
20-39	0.319	0.279	0.395
40-59	0.353	0.345	0.369
60-	0.327	0.375	0.236
Education level			
Less than high school	0.258	0.190	0.388
Some college	0.219	0.200	0.255
Assoc./professional degree	0.187	0.200	0.163
College or post-graduate	0.336	0.410	0.194
Household annual income			
-$25k	0.211	0.148	0.331
$25k-$50k	0.258	0.246	0.281
$50k-$75k	0.202	0.230	0.148
$75k-$150k	0.262	0.297	0.194
$150k-	0.068	0.080	0.046
Occupation type			
Full-time	0.497	0.509	0.475
Part-time	0.102	0.100	0.106
Not working	0.401	0.319	0.418
Marital status			
Married/live with partner	0.690	0.713	0.646
# of obs. in the sample	764	501	263

Revealed-preference tests, distance measures, and ambiguity attitude. The pass rates for GARP, SEU, and MEU presented in Table 5 are very similar to those of our lab data. We find high GARP pass rates, but very low rates for SEU and MEU. Importantly, MEU again does not have more explanatory power than SEU: there is no room for additional rationalizations by allowing for multiple priors. Only one non-SEU subject is rationalized by MEU. High compliance

TABLE 5: Pass rates.

Treatment	N	GARP	PS	SEU	MEU
Large variance	245	0.4492	0.3945	0.0122	0.0122
Small variance	256	0.4367	0.3959	0.0156	0.0195
Combined	501	0.4431	0.3952	0.0140	0.0160

Note: Since Epstein's (2000) condition is only necessary for probabilistic sophistication, the numbers reported here capture the fraction of the subjects who are *not inconsistent* with probabilistic sophistication.

TABLE 6: Distance measures.

Treatment	N	CCEI			e_* (SEU)			e_* (MEU)		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Large variance	245	0.9720	0.9950	0.0509	0.8974	0.9309	0.3740	0.8927	0.9163	0.3765
Small variance	256	0.9688	0.9958	0.0552	0.8868	0.9062	0.3820	0.8852	0.9062	0.3829
Combined	501	0.9704	0.9954	0.0531	0.8920	0.9163	0.3778	0.8888	0.9163	0.3794

with GARP pushes the average CCEI score above 0.97 (Table 6). The average e_* of 0.907 is not statistically different from the average 0.878 in the lab study (two-sample t-test, $t(626) = 0.772$, $p = 0.441$). As in our lab study, we find that e_* and how well choices respond to prices are positively associated (Figure C.10). Subjects who violated event monotonicity (monotonicity in questions #5 and #6) for more than five-token margin have significantly higher e_* on average (mean 0.999 vs. 0.881, two-sample t-test, $t(499) = 2.925$, $p < 0.01$), but the difference is not statistically significant when we do not allow for this margin (mean 0.928 vs. 0.894, two-sample t-test, $t(499) = 0.988$, $p = 0.324$). Among the subjects who satisfied (exact) event monotonicity, the larger the difference between tokens allocated in two questions becomes, the higher e_* becomes (Spearman's correlation coefficient $r = 0.127$, $p = 0.024$). So there is some evidence that the degree of violation of monotonicity in questions #5 and #6 is related to the magnitude of deviation from SEU.

The pattern of choices in the standard-Ellsberg task is also similar to what we observed in the lab data, but the overall frequency with which the risky bet is chosen is smaller. In particular, only 70% of subjects (regardless of treatment) chose the risky bet in the first question, in which the risky bet pays a $5 more than the ambiguous bet in case of winning (note that almost everybody chose the risky bet in the lab, albeit with a reward magnitude that is 1/10th of what we used
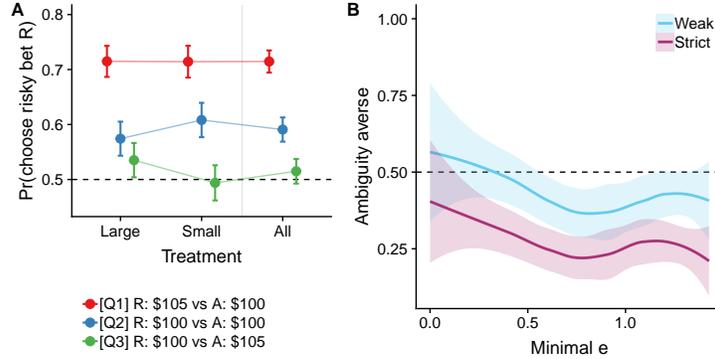
FIGURE 9: (A) Probability of choosing a risky bet in each question in the standard-Ellsberg task in the panel data. (B) LOESS curves relating e_* and ambiguity attitude.

in the panel). There are thus 44% (26%) of subjects who are weakly (strictly) ambiguity averse (Figure 9). These numbers are lower than in the lab data. Now, using this classification, we look at the relationship between ambiguity aversion and e_*. Unlike Figure 7BC, using lab data, Figure 9A exhibits a decreasing relation between the two (there is a slight indication of reflection around $e_* \approx 0.8$, but it is not as strong as Figure 7BC). Combining these two observations, we can see that subjects with small e_* (close to SEU) are not necessarily less ambiguity averse in the standard Ellsberg task.

Sociodemographic correlation.   One of the great advantages of using the UAS survey is that registered researchers can access datasets from past surveys, and use subject responses in related surveys and experiments. In particular, we use basic demographic information collected through the survey, as well as measures of cognitive ability, financial literacy, and other behavioral data from relevant experiments.[20]

We estimated a linear model

$$y_i = \boldsymbol{X}_i \boldsymbol{\beta} + \varepsilon_i,$$

where the dependent variable y_i is subject i's value of e_* or downward-sloping demand measured by correlation ρ^{dsd} between $\log(p_2/p_1)$ and $\log(x_2/x_1)$, and $\boldsymbol{X}_i$ is a vector of sociodemographic characteristics. These explanatory variables include: age group (omitted category is "20-39 years old"), above-median financial literacy (measured in UAS modules #1 and #6; omitted category is "below-median score"), cognitive ability measured with CRT (omitted category is "score is 0"),

---

[20]The cognitive ability measure is taken from survey module #1. Two financial literacy measures are taken from modules #1 and #6, which asked both the basic and sophisticated financial literacy questions in Lusardi and Mitchell (2017). One caveat to this approach is the time lag between previous the surveys and ours. For example, the first survey module #1 was administered in May 2014.

education level (omitted category is "high school or less"), annual income group (omitted category is "less than \$25,000"), gender, and employment status. The model is estimated by OLS with robust standard errors. We also estimate logistic regressions where the dependent variable y_i is event monotonicity (= 1 if monotonicity is violated with a margin of five tokens) and ambiguity attitude in the sense of standard Ellsberg (= 1 if choices indicate weak ambiguity aversion).

Regression results are presented in the first two columns of Table 7. First, there is no effect of age on e_*. Cognitive ability measured with CRT is negatively associated but the effect is not strong. The financial literacy variable measured in UAS module #6 is negatively correlated with e_* (i.e., subjects with higher financial literacy are closer to SEU). Subjects in higher income brackets have larger e_* (i.e., further away from SEU), compared to those in the lowest bracket in our sample. Educational background has an effect in the expected direction, but only in the category "associate or professional degree," not in "college or post-graduate degree."[21] Demographic characteristics do not capture variation in the compliance with the downward-sloping demand property (column 2), but a similar effect of income is observed. Other two measures, violation event monotonicity and ambiguity attitude in the sense of Ellsberg, also exhibit non-significant association with demographic characteristics (except that high CRT score subjects tend to be ambiguity averse compared to low CRT score counterpart).

3.3   Comparing the Lab and the Panel

Finally, we compare the distribution of e_* in the lab and panel data. We can make this comparison because the same set of prices was used in the two experiments. Budgets were very different, but e_* is about relative prices and not about budgets (in contrast with CCEI; see Echenique et al. (2018) for details). It is evident from Figure 10 that there is no difference in distributions of e_*. As a basic check to compare that subjects' decisions are at least different than what random choices would offer, we compared the observed distributions to what purely random choices would give rise to: the two distributions are significantly different from the distribution of e_* when simulated subjects make uniformly random choices.

---

[21]In contrast to these observations, Echenique et al. (2018) find that older subjects have larger e_* for OEU (i.e., further away from OEU, not SEU) than younger subjects; a robust finding in the sense that it holds across data from three different panel surveys (Choi et al., 2014; Carvalho et al., 2016; Carvalho and Silverman, 2019).

TABLE 7: Relation between demographic characteristics and measures for several aspects of behavior in the experiment.

	OLS		logistic	
	(1)	(2)	(3)	(4)
Dependent variable	e_*	ρ^{dsd}	Violate mon.	Weak AA
Treatment: Large	0.016	0.055	0.003	0.182
	(0.034)	(0.036)	(0.233)	(0.198)
Age: 40-59	−0.012	−0.008	−0.109	−0.134
	(0.045)	(0.049)	(0.316)	(0.263)
Age: 60+	0.026	−0.041	0.365	−0.249
	(0.048)	(0.052)	(0.319)	(0.288)
Financial literacy (UAS #1): High	0.034	0.033	−0.291	0.307
	(0.043)	(0.043)	(0.263)	(0.250)
Financial literacy (UAS #6): High	−0.106**	−0.054	0.067	0.204
	(0.041)	(0.041)	(0.268)	(0.247)
CRT score (UAS #1): 1	−0.013	−0.031	−0.362	0.436
	(0.040)	(0.041)	(0.264)	(0.232)
CRT score (UAS #1): 2+	−0.059	−0.029	−0.609	0.711*
	(0.051)	(0.053)	(0.362)	(0.286)
Education: Some college	0.046	−0.000	0.142	−0.070
	(0.053)	(0.059)	(0.342)	(0.331)
Education: Assoc. or pro. degree	−0.107*	−0.095	−0.167	−0.026
	(0.054)	(0.059)	(0.374)	(0.324)
Education: College or postgrad	−0.015	−0.050	−0.478	0.574
	(0.050)	(0.055)	(0.346)	(0.299)
Income: 25,000-49,999	0.109	0.096	−0.055	0.470
	(0.059)	(0.059)	(0.368)	(0.335)
Income: 50,000-74,999	0.184**	0.122*	0.635	0.071
	(0.058)	(0.059)	(0.374)	(0.353)
Income: 75,000-149,999	0.155**	0.142*	−0.112	0.226
	(0.060)	(0.061)	(0.414)	(0.349)
Income: 150,000+	0.124	0.187	−0.087	0.675
	(0.085)	(0.101)	(0.614)	(0.484)
Male	−0.062	−0.019	−0.130	0.277
	(0.036)	(0.036)	(0.247)	(0.210)
Working	0.024	−0.005	−0.311	−0.309
	(0.040)	(0.043)	(0.299)	(0.250)
Constant	0.838***	−0.302***	−0.779	−1.237**
	(0.070)	(0.075)	(0.431)	(0.430)
Observations	490	490	490	490
R^2	0.070	0.033		
Adjusted R^2	0.039	0.001		
Log likelihood			-239.470	-309.069

Note: Robust standard errors are presented in parentheses. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.
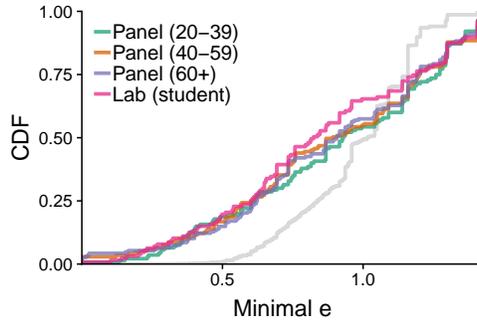
FIGURE 10: Comparing distributions of e_* from the panel study and the lab study.

4 Conclusion

Motivated by recent theoretical advances that provide revealed-preference characterizations of expected utility theory, we design and implement a novel experimental test of the theory. We find that subjects are mostly consistent with utility maximization, and respond to price changes in the expected direction: they satisfy the downward-sloping demand property, at least to some degree, but not enough to make their choices consistent with SEU. Our findings are the same, regardless of whether we look at lab or panel data. In fact, there is a striking similarity in how SEU is violated across the two studies. The subject populations are very different but look very similar in terms of the distribution of the degree of violation of SEU.

Motivated also by the literature on ambiguity aversion, we study the possibility that violations of SEU are due to ambiguity aversion, and look at whether maxmin expected utility (MEU) can explain the data. MEU adds no explanatory power to SEU: with a single exception, *all* subjects who fail to satisfy SEU also fail MEU. It is possible that other models of ambiguity aversion could do a better job of accounting for our experimental data. We are restricted to MEU because it is the only model for which there exist nonparametric tests of the kind that we use in our paper; it is also arguably the best known, and most widely applied, model in the ambiguity literature. The testable implications of other models of ambiguity-averse choice is an interesting direction for future research.

Finally, the results in our experiments are markedly unaffected by some of the demographic characteristics that other studies (on risky choice, not uncertain) have found significant. Older subjects do not seem to violate SEU to a larger degree than younger subjects. Neither do we see higher degrees of SEU violations in our broad sample of the U.S. population, compared to our laboratory experiment conducted on undergraduate students. There are modest effects of income and education. Financial literacy is correlated with subjects' distance to SEU.

There is no doubt that further studies are necessary to fully understand the behavior in environments that are more "natural" than traditional artificial Ellsberg-style settings. Our non-parametric revealed-preference tests and the empirical approach driven by these theories should hopefully be a useful tool to collect more evidence in this direction.

References

ABDELLAOUI, M., A. BAILLON, L. PLACIDO, AND P. P. WAKKER (2011): "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation," *American Economic Review*, 101, 695–723.

AFRIAT, S. N. (1972): "Efficiency Estimation of Production Functions," *International Economic Review*, 13, 568–598.

AHN, D. S., S. CHOI, D. GALE, AND S. KARIV (2014): "Estimating Ambiguity Aversion in a Portfolio Choice Experiment," *Quantitative Economics*, 5, 195–223.

BAILLON, A. AND H. BLEICHRODT (2015): "Testing Ambiguity Models through the Measurement of Probabilities for Gains and Losses," *American Economic Journal: Microeconomics*, 7, 77–100.

BAILLON, A., H. BLEICHRODT, U. KESKIN, O. L'HARIDON, AND C. LI (2018a): "The Effect of Learning on Ambiguity Attitudes: An Experiment Using Initial Public Offerings on a Stock Market," *Management Science*, 64, 2181–2198.

BAILLON, A., Y. HALEVY, AND C. LI (2015): "Experimental Elicitation of Ambiguity Attitude Using the Random Incentive Systems," Unpublished manuscript.

BAILLON, A., Z. HUANG, A. SELIM, AND P. P. WAKKER (2018b): "Measuring Ambiguity Attitudes for All (Natural) Events," *Econometrica*, 86, 1839–1858.

BIANCHI, M. AND J.-M. TALLON (2019): "Ambiguity Preferences and Portfolio Choices: Evidence from the Field," *Management Science*, 65, 1486–1501.

BONSANG, E. AND T. DOHMEN (2015): "Risk Attitude and Cognitive Aging," *Journal of Economic Behavior & Organization*, 112, 112–126.

CAMERER, C. AND M. WEBER (1992): "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty*, 5, 325–370.

CARVALHO, L. AND D. SILVERMAN (2019): "Complexity and Sophistication," NBER Working Paper No. 26036.

CARVALHO, L. S., S. MEIER, AND S. W. WANG (2016): "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday," *American Economic Review*, 106, 260–84.

CHAMBERS, C., F. ECHENIQUE, AND K. SAITO (2016): "Testing Theories of Financial Decision Making," *Proceedings of the National Academy of Sciences*, 113, 4003–4008.

CHEW, S. H., B. MIAO, AND S. ZHONG (2017): "Partial Ambiguity," *Econometrica*, 85, 1239–1260.

CHOI, S., R. FISMAN, D. GALE, AND S. KARIV (2007): "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American Economic Review*, 97, 1921–1938.

CHOI, S., S. KARIV, W. MÜLLER, AND D. SILVERMAN (2014): "Who is (More) Rational?" *American Economic Review*, 104, 1518–1550.

DIMMOCK, S. G., R. KOUWENBERG, O. S. MITCHELL, AND K. PEIJNENBURG (2015): "Estimating Ambiguity Preferences and Perceptions in Multiple Prior Models: Evidence from the Field," *Journal of Risk and Uncertainty*, 51, 219–244.

——— (2016a): "Ambiguity Aversion and Household Portfolio Choice Puzzles: Empirical Evidence," *Journal of Financial Economics*, 119, 559–577.

DIMMOCK, S. G., R. KOUWENBERG, AND P. P. WAKKER (2016b): "Ambiguity Attitudes in a Large Representative Sample," *Management Science*, 62, 1363–1380.

DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2018): "On the Relationship Between Cognitive Ability and Risk Preference," CRC TRR 190 Discussion Paper No. 76.

ECHENIQUE, F., T. IMAI, AND K. SAITO (2018): "Approximate Expected Utility Rationalization," Caltech Social Science Working Paper #1441.

ECHENIQUE, F. AND K. SAITO (2015): "Savage in the Market," *Econometrica*, 83, 1467–1495.

ELLSBERG, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643–669.

EPSTEIN, L. AND Y. HALEVY (2019): "Ambiguous Correlation," *Review of Economic Studies*, 86, 668–693.

EPSTEIN, L. G. (2000): "Are Probabilities Used in Markets?" *Journal of Economic Theory*, 91, 86–90.

FALK, A., A. BECKER, T. DOHMEN, B. ENKE, D. HUFFMAN, AND U. SUNDE (2018): "Global Evidence on Economic Preferences," *Quarterly Journal of Economics*, 133, 1645–1692.

FREDERICK, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 19, 25–42.

GILBOA, I. (2009): *Theory of Decision under Uncertainty*, Cambridge: Cambridge University Press.

GILBOA, I. AND D. SCHMEIDLER (1989): "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

HALEVY, Y. (2007): "Ellsberg Revisited: An Experimental Study," *Econometrica*, 75, 503–536.

HEY, J. D., G. LOTITO, AND A. MAFFIOLETTI (2010): "The Descriptive and Predictive Adequacy of Theories of Decision Making under Uncertainty/Ambiguity," *Journal of Risk and Uncertainty*, 41, 81–111.

HEY, J. D. AND N. PACE (2014): "The Explanatory and Predictive Power of Non Two-Stage-Probability Theories of Decision Making Under Ambiguity," *Journal of Risk and Uncertainty*, 49, 1–29.

HUFFMAN, D., R. MAURER, AND O. S. MITCHELL (forthcoming): "Time Discounting and Economic Decision-making in the Older Population," *Journal of the Economics of Ageing*.

LOOMES, G. (1991): "Evidence of a New Violation of the Independence Axiom," *Journal of Risk and Uncertainty*, 4, 91–108.

LUSARDI, A. AND O. S. MITCHELL (2017): "How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy and Retirement Readiness," *Quarterly Journal of Finance*, 7, 1750008.

MACHINA, M. J. AND D. SCHMEIDLER (1992): "A More Robust Definition of Subjective Probability," *Econometrica*, 60, 745–780.

POLISSON, M., J. QUAH, AND L. RENOU (2017): "Revealed Preferences over Risk and Uncertainty," University of St. Andrews School of Economics and Finance Discussion Paper No. 1706.

SAITO, K. (2015): "Preference for Flexibility and Randomization under Uncertainty," *American Economic Review*, 105, 1246–1271.

SAVAGE, L. J. (1954): *The Foundations of Statistics*, New York: Wiley.

SUNDE, U. AND T. DOHMEN (2016): "Aging and Preferences," *Journal of the Economics of Ageing*, 7, 64–68.

TRAUTMANN, S. T. AND G. VAN DE KUILEN (2015): "Ambiguity Attitudes," in *The Wiley Blackwell Handbook of Judgment and Decision Making*, ed. by G. Keren and G. Wu, Oxford, UK: John Wiley & Sons, 89–116.

TYMULA, A., L. A. R. BELMAKER, L. RUDERMAN, P. W. GLIMCHER, AND I. LEVY (2013): "Like Cognitive Function, Decision Making across the Life Span Shows Profound Age-Related Changes," *Proceedings of the National Academy of Sciences*, 110, 17143–17148.