Research Dialogue | Issue no. 153
June 2019

TIAA Institute

Decision making under uncertainty: An experimental study in market settings

Federico Echenique, California Institute of Technology

Taisuke Imai, Ludwig Maximilian University of Munich

Kota Saito, California Institute of Technology

Abstract

We design and implement a novel experimental test of subjective expected utility theory and its generalizations. Our experiments are implemented in the lab, and pushed out through a large-scale panel to a general sample of the U.S. population. We find that subjects respond to price changes in the expected direction, but not enough to make their choices consistent with the theory. Surprisingly, maxmin expected utility adds no explanatory power to subjective expected utility. Our findings are the same, regardless of whether we look at laboratory data or a large panel survey, even though the subject populations are very different. The degree of violations of subjective expected utility theory is not affected by age, but is correlated with financial literacy and income. The effects of education level and gender are weak and not independent from the effects of other demographic variables.

The authors wish to thank Aurélien Baillon, Yoram Halevy, Pietro Ortoleva, Maricano Siniscalchi, and Charles Sprenger for helpful discussions. The authors also thank Noriko Imai for developing the software for the experiment, Yimeng Li for research assistance, John Duffy, Michael McBride, and Jason Ralston for supporting our experiments at ESSL, and Tania Gutsche and Bart Orriens at CESR (University of Southern California) for setting up and implementing the survey on the Understanding America Study. This research is supported by Grant SES1558757 from the National Science Foundation and the TIAA Institute and Wharton School's Pension Research Council/Boettner Center. In addition, Echenique thanks the NSF for its support through the grant CNS-1518941, and Imai acknowledges financial support by the Deutsche Forschungsgemeinschaft through CRC TRR 190. This project received funding from the TIAA Institute and Wharton School's Pension Research Council/Boettner Center. The content is solely the responsibility of the authors and does not necessarily represent official views of the TIAA Institute or Wharton School's Pension Research Council/Boettner Center. The project described in this paper relies on data from surveys administered by the Understanding America Study, which is maintained by the Center for Economic and Social Research (CESR) at the University of Southern California. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of USC or UAS.

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 the presence of complete markets. In our study, we record human subjects' choices in a laboratory setting, and in a large-scale field panel. As a consequence, we obtain results for very different populations, ranging from undergraduate students to older retirees. Our data allows 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, as well as how well the agents conform to "objective" expected utility. Our experiments also speak to the external validity of laboratory studies, since we can compare the outcomes in the lab to the outcomes of a sample from the general U.S. population.

Subjective expected utility theory (SEU; Savage, 1954) is the standard model of decision making in the presence of 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 belief over the states of the world, and who maximizes their expected utility with respect to this probability. 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, and with an eye 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 wish to avoid bets on uncertain events, in ways that cannot be represented with a single subjective probability. This avoidance is termed ambiguity aversion. The Ellsberg paradox is based on a thought experiment, using the choice of bets based on drawing a ball from urns. One of our contributions is to empirically assess SEU in an economic setting that closely resembles the real-world environments 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.

While ambiguity aversion has been identified in many different contexts, and in different subject populations (Trautmann and van de Kuilen, 2015), our understanding of this phenomenon is still incomplete. For one thing, researchers have relied almost exclusively on the paradigm introduced by Ellsberg (1961), in which agents are offered bets on the color of a ball 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 relaxing conditions in 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 SEU. The measures establish the degree of a dataset's violation of SEU. While these studies focus mostly on establishing revealed preference conditions, the main motivation of the current paper is to bring these theoretical machineries 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 and CARA. We do assume that agents are risk averse or risk neutral (they have a concave von-Neumann-Morgenstern utility).[1]

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. This setting naturally translates into our proposed experimental design, which follows the spirit of portfolio choice task introduced by Loomes (1991) and Choi et al. (2007), and later used in many other studies (e.g., Ahn et al., 2014; Choi et al., 2014; Hey and Pace, 2014; Carvalho and Silverman, 2017). Subjects in our experiment 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 for a given decision problem. Two accounts correspond to two mutually exclusive events, and subjects are told that they receive payment based on the allocation and the realized event. Importantly, subjects are provided no information regarding the objective probabilities of these events. We generate uncertainty using two different sources. The first source is the classical Ellsberg-style "bags and balls." The second source comes from simulated stock prices. We ran our experiments in a lab setting, using undergraduate students: we refer to this data as "the lab." We also ran our experiments on a large scale panel, representative of the US population. The second data set is called "the panel" for short. See Section 2.2 for details.

1.1 Overview of results

The main purpose of our study was to test theories of decision under uncertainty. The news is not good for the 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 the pass rates for MEU are just as low as for SEU. In fact, in all of our samples, only one subject's choice is consistent with MEU but not SEU. One positive finding is that subjects seem to be utility maximizers (they satisfy Generalized Axiom of Revealed Preference), and do not violate Epstein's (2000) necessary condition for probabilistic sophistication.

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 is related 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 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.

[1] See Polisson et al. (2015) for revealed preference tests that do not assume risk aversion.

Our panel experiment allows us to compare the distance to SEU between subjects with different sociodemographic characteristics. The most notable finding is that demographics that would seem to matter for SEU, and that do matter for related theories of choice (Choi et al., 2014; Echenique et al., 2018), do not show up as significant factors in compliance with SEU. In particular, older subjects do not necessarily exhibit lower degrees of compliance with SEU. Financial literacy, on the other hand, exhibits a significant relation with the degree of compliance with SEU (subjects with higher literacy score are closer to SEU). Although, this significant relation is only evident in one of the two financial literacy measures. There are weak effects of gender (male subjects are closer to SEU than female subjects) and cognitive ability (subjects with higher cognitive ability score are closer to SEU) but these effects are not evident after controlling for other demographic characteristics such as education level and income.

A subset of our subjects previously participated in a separate experiment (on the same survey platform) that tested for objective expected utility (OEU; meaning that the probability of each state was known). Using this subsample of subjects, we do find robust effects of age, education, cognitive ability and financial literacy. Our results suggest that SEU and OEU may be unrelated phenomena. Situations where agents are provided with objective probabilities may be viewed, and reasoned about, by agents in substantially different ways than situations with uncertainty.

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 seems less ambiguity averse 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) 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., Halevy, 2007; Abdellaoui et al., 2011; Baillon and Bleichordt, 2015; Chew et al., 2017; Epstein and Halevy, forthcoming). Other studies try to parametrically estimate the models under consideration (e.g., Hey et al., 2010; Ahn et al., 2014; Hey and Pace, 2014; Dimmock et al., 2015). 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. For example, Camerer and Weber (1992) notes that (p. 361):

"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!"

Similarly, Gilboa (2009) writes (p. 136):

"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."

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 Bleichordt, 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).[2] 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., Bonsang and Dohmen, 2015; Sunde and Dohmen, 2016; Dohmen et al., 2018; Falk et al., 2018; Huffman et al., forthcoming).

Dimmock et al. (2015) use an Ellsberg-type choice data from the American Life Panel (ALP) survey to estimate parameters of α-maxmin model (Ghirardato et al., 2004). The authors then relate estimated ambiguity aversion parameter and perceived level of ambiguity with individual characteristics including age, gender, employment, income, and find that (i) male participants perceived higher level of ambiguity and are more ambiguity averse, (ii) age is negatively correlated with ambiguity aversion but does not influence ambiguity perception, and (iii) participants with college education are more ambiguity averse than less educated.

Dimmock et al. (2016a) study, using the ALP survey data, the relationship between ambiguity aversion measured in an experimental task and stock market participation and portfolio choice in the real world. The authors find, among others, that ambiguity aversion is negatively related with stock market participation, portfolio allocation to equity, and foreign stock ownership (home-bias). Bianchi and Tallon (forthcoming) find similar results using French survey data. Dimmock et al. (2016b), however, find weak correlation between ambiguity attitudes and demographic variables in the Longitudinal Internet Studies for the Social Sciences (LISS) panel.

Using a simple binary lottery choice task, involving either known or unknown probabilities, Tymula et al. (2013) study the effect of aging on decision making under risk and uncertainty. They find that healthy elders (between 65 and 90 years old) exhibit larger frequency of inconsistent choices (measured by first-order stochastic dominance or switching) than younger subjects.

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, longitudinal survey platform (hereafter “the panel”).[3] 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.

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.

[2] 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).

[3] 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.

Table 1. Order of the tasks

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

Market task. The market task is meant to represent the most basic economic problem of choice under uncertainty, where an agent chooses among Arrow-Debreu commodities with given state prices and 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 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 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 subject 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 A.1 in Appendix.

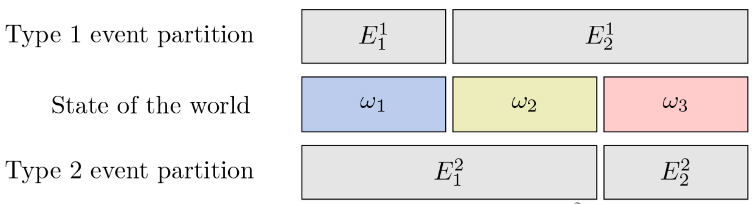
An allocation (a_1,a_2) of tokens is equivalent to buying a x_s units of an Arrow-Debreu security that pays $1 per unit if event E_s holds, from a budget set satisfying $p_1 x_2+ p_2 x_2=I$, where prices and income (p_1,p_2, I) are determined by the token exchange values (z_1,z_2) in the round.[4] A graphical presentation of budget sets used in the experiment is provided in Figure 4 below.

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 question but also across two types of questions. It allows us to examine one aspect of SEU rationality, monotonicity of choice.[5] 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 include a more detailed discussion later in the paper.

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

[5] 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 parametric estimation of leading models of ambiguity aversion.

Figure 1. Event structure in two types of questions



Subjects in our experiment 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 Figure 2. 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.[6]

The allocation table also makes it clear which type of question is presented. For example, the left table in Figure 2 indicates that the question is of type 1, since Y (ω_2) and R (ω_3) are grouped together. Similarly, the right table indicates that the question is of type 2, since B (ω_1) and Y (ω_2) are grouped together.

Figure 2. Illustration of the allocation table for type 1 questions (left) and type 2 questions (right)

	B	Y R
Token value	\$0.36	\$0.24
Token	30	70
Account value	\$10.80	\$16.80

	B Y	R
Token value	\$0.40	\$0.50
Tokens	75	25
Account value	\$30.00	\$12.50

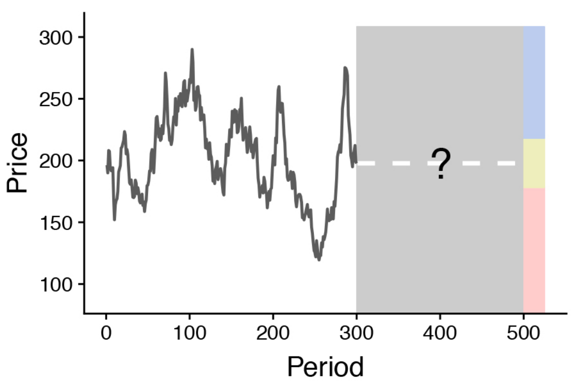
An important feature of our design is that we implement the task under two different sources of uncertainty. Subjects face two versions of the market task, as we change the source of uncertainty. In the first version, called market-Ellsberg, uncertainty is generated with an Ellsberg urn. In the second 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 red chip (R in Figure 2), state 2 (ω_2) corresponds to drawing a yellow chip (Y), and state 3 (ω_3) corresponds to drawing a blue chip (B).

[6] 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.

In the market-stock version of our task, uncertainty is generated through the realization of simulated stock prices.[7] Subjects are presented with a history of stock prices, as in Figure 3.[8] 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 $R \in (0,1)$: $\omega_1 = (R,+\infty)$, in which the target value rises by more than 100 x R% compared to the starting value (the blue region in Figure 3; B in Figure 2), $\omega_2 = [-R,R]$, in which the price varies by at most 100 x R% between the starting value and the target value (the yellow region in Figure 3; Y in Figure 2), and $\omega_3 = (-1,-R)$, in which the target value falls by more than 100 x R% compared to the starting value (the red region in Figure 3; R in Figure 2).

Figure 3. Source of uncertainty in the market-stock task

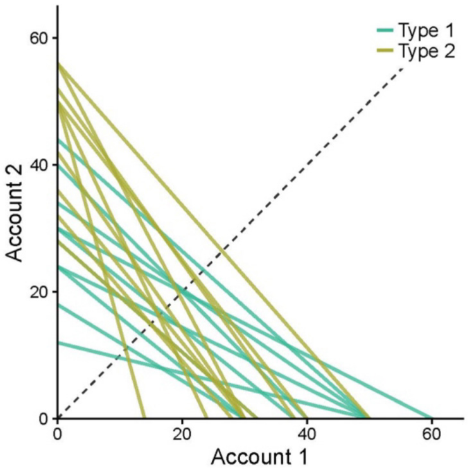


We chose token exchange values (z_1,z_2) for each question to increase the power of our tests. After running several choice simulations to calculate the power of our tests, we select 20 budgets (10 for type 1, 10 for type 2) shown in Figure 4 (and Table A.1 in Appendix).

[7] 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's Mechanical Turk (MTurk) 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 trend, 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.

[8] This chart was presented (above the allocation table) to all subjects in one of the treatments (see Section 2.2 for description of treatment variation). Note that the chart was kept constant throughout the 20 questions in the market-stock task. Only the allocation table varied between questions.

Figure 4. Set of 20 budgets used in the allocation task



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), but it is 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 two market tasks 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 (Figure 5): Bag A contains 10 orange chips and 10 green chips. Bag B contains 20 chips. In bag B, each chip 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.

Figure 5. Ellsberg urns

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.[9] 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 (X,b) = (100,5) and in the panel. In our lab experiment, 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, forthcoming; Saito, 2015). The subjects were allowed to inspect the content of each bag after completing the experiment.

Post-experiment survey. In our lab experiments, subjects were asked to fill out a post-experiment 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. 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 experiments saw the identical interface.

Pilot study. We ran two sessions of pilot experiment using Amazon's Mechanical Turk (MTurk). The main purpose of these pilot sessions was to calibrate our experimental design: (i) we checked whether subjects responded to prices in the set of budget we selected (Figure 4), and (ii) we implemented a belief elicitation task to examine how simulated stock prices generated uncertainty. The second part was the basis for our choice of price paths presented in the market-stock task. We collected the main data using the UAS platform instead of MTurk because the UAS has a nationally representative pool of subjects and also brings a higher-quality data (MTurk has been struggling with declining data quality due to the presence of "bot" workers, which are automated programs mimicking actual human behavior).

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.[10] 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. As we describe above, subjects in the experiment saw only one price path (like the one in Figure 3) on the screen. This feature makes it difficult to vary the volatility of price path between treatments since the perception of volatility is only relative. In order to effectively induce treatment variation, we embed each path in the common market "context" as shown in Figure 6. Here, the bold

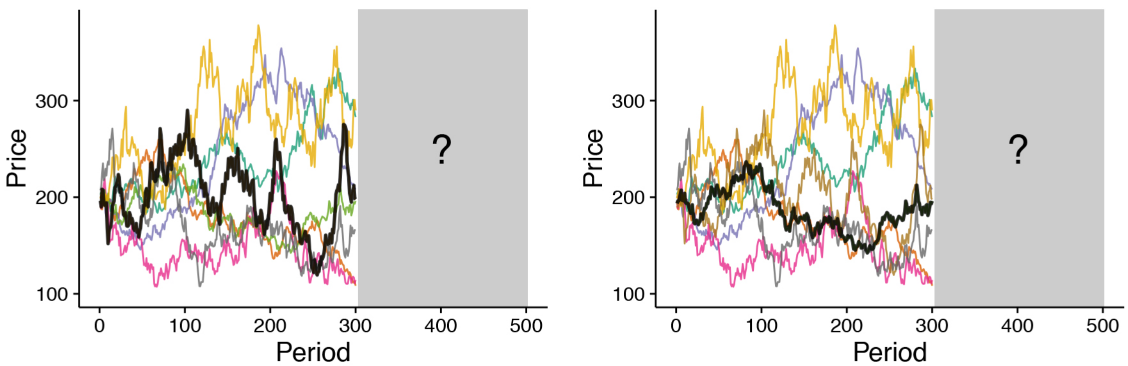
9 We adopted the three-question setting from Epstein and Halevy (forthcoming), as a way of identifying strict ambiguity preferences. The typical Ellsberg-style experiment would ask only one question, namely the second one.

10 https://uasdata.usc.edu/index.php

black lines indicate the stock under consideration in each treatment, and the other lines in the background (also simulated using a Geometric Brownian Motion) are the same in the two treatments.

Our main treatment variation is the perceived volatility of simulated stock prices. The subjects were randomly assigned to either a large volatility condition (left panel in Figure 6), or a small volatility condition (right panel). The instructions (available upon request) for the market-stock task included one of the two charts of Figure 6, depending on the treatment.[11]

Figure 6. Context of market information: large volatility (left) and small volatility (right)



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

[11] There are studies using market condition priming like ours. Cohn et al. (2015), for example, show subjects a chart of hypothetical stock prices which is either with an increasing trend ("boom") or with a decreasing trend ("bust"), to study countercyclical risk aversion.

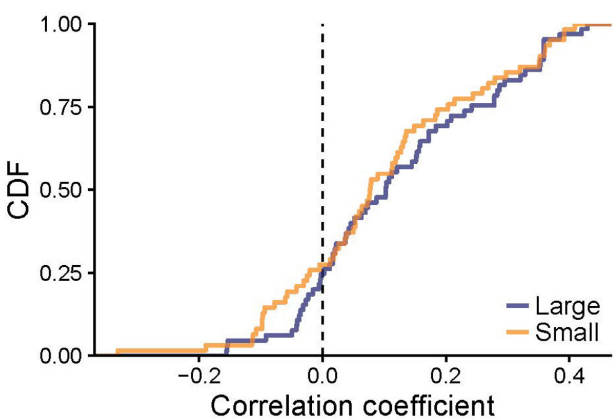
3.1 Results from the lab

We conducted seven experimental sessions at the Experimental Social Science Laboratory (ESSL) of the University of California, Irvine. A total of 127 subjects (age mean = 20.16, SD = 1.58; 35% male; 62 in treatment Small and 65 in treatment Large) participated in the lab portion of the study.[12] Each session lasted about an hour, and subjects earned on average $21.3 (including a $7 show-up fee, SD = 9.21). We used software specifically coded for the purpose of running our experiment.

As we described in Section 2, subjects in the lab performed three tasks in the following order: the market task with stock price information (market-stock), the market task with Ellsberg information (market-Ellsberg), and the standard Ellsberg two-urn choice (standard-Ellsberg). In the market-stock task, we prepared two price paths with different levels of variability (we call the two treatments Large and Small). Subjects were randomly assigned to one of the two treatments at the session level (meaning that all subjects in the same session were shown the same price path).

Choices in the market tasks. In the two market tasks, subjects face 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 7); and 36% of those subjects exhibited significantly positive correlation (one-sided, $p < 0.05$).

Figure 7. Empirical CDF of within-subject correlation (Spearman) between allocations in the market- stock task and the market-Ellsberg task

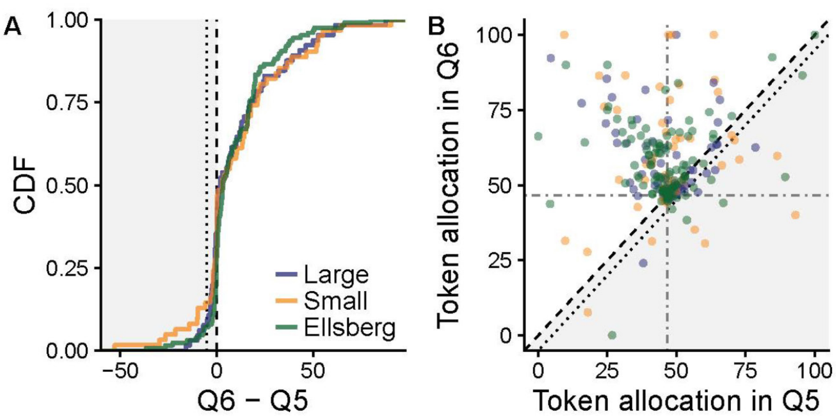


The two distributions are not significantly different (two-sample Kolmogorov-Smirnov test, $p = 0.57$).

[12] 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 significantly longer time for each decision than anyone else. They had to be eliminated from our data, because they were delaying the experiment for the rest of the subjects in their session. We distributed the instructions for each task of the experiment just before they were to perform that task, so each subject would have to wait until all the other subjects in the session completed the task. We had to “nudge” the two subjects that were extremely slow, and hence eliminated their choices from our data.

We prepared two consecutive questions, questions 5 and 6, that had the same budget, but with different event structures. These were the only questions that were not presented in random order, and we included them 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. By construction of the events ($E_1^1 = \{\omega_1\}$ and $E_1^2 = \{\omega_1, \omega_2\}$, we expect that subjects would allocate more tokens to the first account in 6th than in 5th. Since the event upon which the first account pays off is a larger set in question 6 than in question 5, while prices and budget remains the same, subjects should allocate more to the first account in question 6 than in question 5: we term this property monotonicity of allocation with respect to event structure. Figure 8 confirms this hypothesis. More than 70% of our lab subjects satisfied monotonicity, and this number increased to 90% if we allow for a margin of error of five tokens (Figure 8A). Choices are clustered around 46.67 tokens, which equalize payout from two accounts (dot-dashed lines in Figure 8B). This can be interpreted as subjects displaying ambiguity aversion.[13]

Figure 8. Monotonicity of allocations with respect to event structures



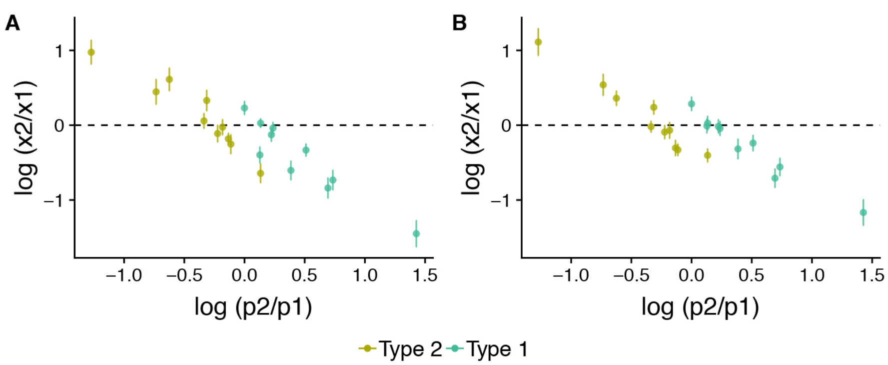
(A) Empirical CDFs of token allocation difference. The dotted line represents a 5-token margin. No two pairs of distributions is significantly different (two-sample Kolmogorov-Smirnov test). (B) Token allocations in two questions. The dot-dashed lines at 46.67 indicate the number of tokens which equalizes payouts in two events.

[13] A formal test for ambiguity aversion is discussed in the form of a test for max-min expected utility.

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."[14] We thus look at how the subjects' choices responded to price variability between budgets; in particular we focus on the relationship between price ratios, $\log(p_2/p_1)$, and allocation ratios, $\log(x_2/x_1)$, aggregating choices from all subjects. Figure 9 shows a negative relation between these two quantities, confirming that the "downward sloping demand" property at the aggregate level. It holds for both types of questions and in both tasks (LOESS curves have negative slope).

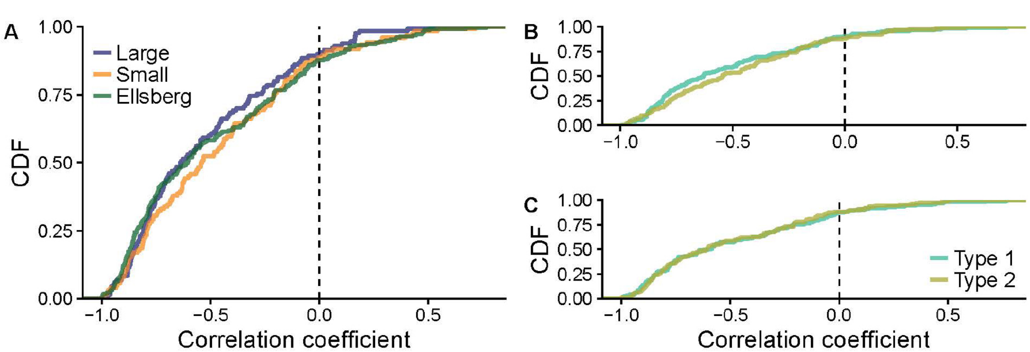
We also quantify the downward-sloping demand property at the individual level by calculating Pearson's correlation coefficient between $\log(p_2/p_1)$ and $\log(x_2/x_1)$. Let r_i^t be the correlation coefficient in type t questions from subject i. We then obtain the "average" correlation coefficient, r_i by Fisher's z-transformation $r_i = \tanh(\sum_{t=1}^{2} \tanh^{-1}(r_i^t)/2)$. Figure 10 shows that a significant majority of the subjects made choices that responded to prices negatively.

Figure 9. Downward sloping demand in the market-stock task (A) and the market-Ellsberg task (B)



Bars represent standard errors of means.

Figure 10. Downward sloping demand at the individual level



(A) Comparison across tasks. (B) market- stock. (C) Market-Ellsberg.

[14] See also Friedman et al. (2018), which uses the similar idea to recover coefficient of risk aversion.

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-based tests on each individual subject's choice data. These tests include: GARP, probabilistic sophistication (hereafter PS), SEU (based on and extended from Echenique and Saito, 2015), and MEU (based on Chambers et al., 2016). Since we have two types of problems, each of which is associated with a partition of the state space, we test GARP, PS, and MEU on each type of problem separately. For SEU, we also implement the test on the data combining two types of problems (it is not obvious that this can be done). Table 2 presents the *pass rate* of each test, i.e., the fraction of subjects (out of 127) who passed each test.

Table 2. Pass rates (%)

	GARP		SEU			MEU			PS	
Task	Type 1	Type 2	Type 1	Type 2	Joint	Type 1	Type 2	Joint	Type 1	Type 2
Market-stock	76.4	68.5	4.7	1.6	0.0	4.7	1.6	0.0	73.2	81.1
Market-Ellsberg	82.7	56.5	7.9	3.2	1.6	7.9	3.2	1.6	81.1	83.5

Note: Since Epstein's (2000) condition is only necessary for probabilistic sophistication (PS), the numbers reported here capture the upper bound of the fraction of the subjects who are consistent with probabilistic sophistication. Type 1 and Type 2 refer to the two types of question subjects faced in the experiment.

We find that a majority of the subjects satisfied GARP, meaning that these subjects' choices are consistent with maximization of *some* utility function, for problems of each type. The pass rates for the type 1 problem are higher, both in the market-stock and the market-Ellsberg tasks, but the differences are only marginally significant at the 10% level (McNemar's test; p=0.0956 for type 1 and p=0.0679 for type 2).

On the contrary, subjects clearly do not make choices that are "as if" they were maximizing a subjective expected utility: SEU pass rates are all below 10%, and not a single agent passed the SEU test in the market-stock task, when we implement the test on the whole dataset combining type 1 and type 2 problems.[15]

Allowing for multiple priors via MEU does not change the result— pass rates are the same between SEU and MEU, implying that MEU does not capture violation of SEU in our experiment. These findings are consistent with the low pass rates reported in Chambers et al. (2016), which uses choice data from Hey and Pace (2014).

Finally, we look at probabilistic sophistication (Machina and Schmeidler, 1992) to investigate whether observed behavior is (in)consistent with preferences being based on probabilities. Using a necessary condition proposed by Epstein (2000), we find that at most 73% to 83% of subjects are consistent with probabilistic sophistication.

[15] Similarly, 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, 2017; Choi et al., 2014) make choices that are consistent with objective expected utility theory.

Table 3. Distance measures

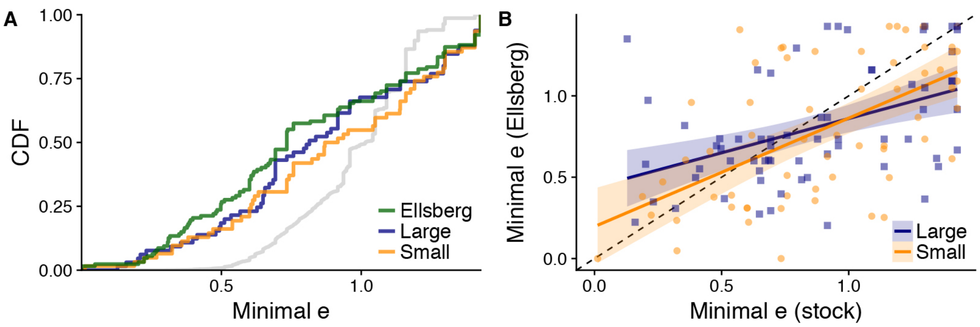
Task	Stat.	CCEI		e_*		
		Type 1	Type 2	Type 1	Type 2	Joint
Market-stock	Mean	0.9895	0.9868	0.6382	0.6381	0.8782
	Median	1.0000	1.0000	0.6004	0.6221	0.8675
	SD	0.0369	0.0382	0.4231	0.3883	0.3772
Market-Ellsberg	Mean	0.9925	0.9960	0.5964	0.5967	0.7985
	Median	1.0000	1.0000	0.6004	0.5390	0.7340
	SD	0.0299	0.0133	0.4126	0.3965	0.3943

Distance measures. The Critical Cost Efficiency Index (CCEI; Afriat, 1972; Varian, 1990) 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 2% 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 2014 experiment, and about the same as the 2007 experiment. 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, and 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 theory proposed by Echenique et al. (2018). The number e_* comes from a perturbation to the model that allows SEU to accommodate the data: It can be interpreted as the size of a utility perturbation that can rationalize the observed choices. Thus, the number e_* is zero if a choice data is consistent with SEU, meaning that no perturbation is needed to rationalize the data by means of SEU, but takes a positive value if it violates SEU. The larger is e_*, the larger is the size of the perturbation needed to rationalize the data by means of a perturbed version of SEU. See Echenique et al. (2018) for details.

One basic finding from our experiments is that the joint e_* (i.e., calculated from the data combining both types of questions) in the market-stock task is significantly higher than in the market-Ellsberg task (paired-sample t-test; $t(126)=2.3686$, $p=0.0194$). See also Figure 11A. Note, however, that this result does not necessarily mean that the subjects made choices that were closer to SEU when the source of information was an Ellsberg urn, since the order of the two market tasks was not counterbalanced.

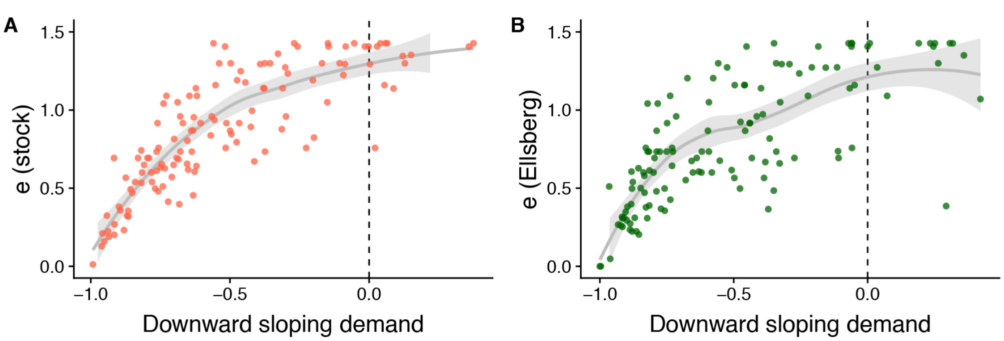
Figure 11. Comparing e_* across tasks



(A) Empirical CDFs. The grey line indicates the distribution generated by random choices. (B) The relationship between e_* from the market-stock and the market-Ellsberg task. Each dot represents a subject.

As we have seen in Figure 7, subjects' choices in the two market tasks are correlated. This correlation is reflected in the degree of violation of SEU—Figure 11B shows that e_* from two tasks are highly correlated (Pearson's correlation coefficient: r=0.4476 for treatment Large, r=0.5821 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 12. The subjects' e_* tend to be large when their choices do not respond to price changes, indicating larger deviation 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.

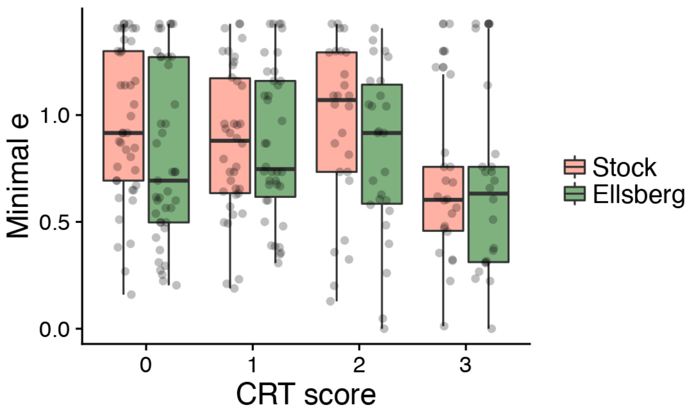
Figure 12. Relation between the degree of conformity to downward sloping demand and e in the market-stock task (A) and in the market-Ellsberg task (B)



Note: correlation coefficient r is first calculated for each type of problem and then aggregated by Fisher's z-transformation ($r_i = \tanh(\sum_{t=1}^{2} \tanh^{-1}(r_i^t)/2)$). Gray lines represent LOESS curves together with 95% confidence bands.

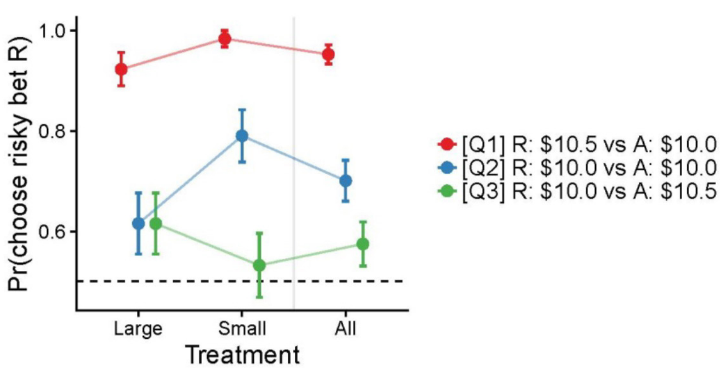
We do not observe gender differences on e_*. We do, however, observe an effect of cognitive ability as measured with three-item Cognitive Reflection Test (CRT; Frederick, 2005). The subjects who answered all three questions correctly exhibit lower e_* than those who answered none of them correctly. This effect is statistically significant only in the market-stock task (Figure 13).

Figure 13. Cognitive Reflection Test score and e_*



Ambiguity attitude. Finally, we look at the relationship between behavior in the market tasks and the 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 14 shows the frequency with which subjects preferred to bet on the risky urn, for each question.

Figure 14. Probability of choosing a risky bet in each question in the standard-Ellsberg in the lab data

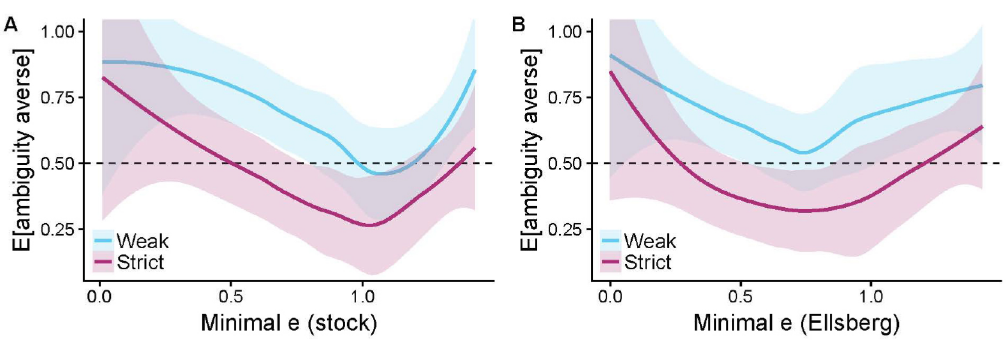


Bars represent standard errors of means.

In the first question, the risky bet pays an additional \$0.5 in case of winning. This bonus makes 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 large (z-test for proportion; p=0.0628 in the Large treatment and p=0.6115 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 significantly 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=0.0314).

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 e_* by 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 15 suggests a weak but quadratic relationship between these two. It may seem counterintuitive since one may expect positive relation between ambiguity aversion and e_*, given that ambiguity aversion is the leading explanation for violations of SEU. It is important to note here that e_* captures any kind of deviation from SEU, and not only those that could be traced to ambiguity aversion.

Figure 15. LOESS curves relating e_* and ambiguity attitude in the lab data



The shaded regions represent 95% confidence bands.

3.2 Results from the panel

A total of 764 subjects (age mean = 50.26, SD = 15.45; 50.39% male) completed the study.[16] 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 a significant proportion of them took it with their mobile devices—such as smartphones or tablets. These devices usually have 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) computer-only (66%), (ii) computer-

[16] Ninety-nine more subjects "opened" the survey link but did not start taking it.

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 computers), and the excluded sample (those who did not use computers). It is evident that the type of device used is correlated with some of the demographic variables (age: χ^2 (2) = 17.79, p<0.001; education level: χ^2 (3)=53.70, p<0.001; income level: χ^2 (4)=43.97, p<0.001). The sub-samples of subjects exhibited markedly different patterns of behavior as well. Throughout the paper, we analyze data from the core sample.[17]

The set of 20 budgets used in the market task is the 10-times scaled-up version of the one used in the laboratory experiment (Figure 4; Table A.1). This keeps the relative prices the same between our two studies, making the distance measure comparable between data from the lab and the panel.

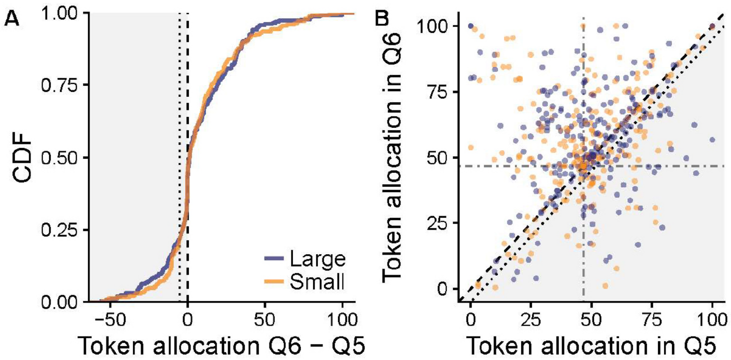
We start by checking the monotonicity of allocations with respect to event structure, along the lines of our discussion for the lab experiment. Our subjects' choices on questions 5 and 6 are informative about how attentive they are when they perform the tasks in our experiment. We find that about 60% of subjects satisfy monotonicity, and that this number jumps to 78% if we allow for a margin of error of five tokens (see Figure 16). There are no treatment differences. Our subjects also made choices that are, to some extent, responding to underlying price changes (Figure 17).

Table 4. Sociodemographic information

Variable	All	Computer	Other devices
Gender			
Male	0.504	0.529	0.456
Age			
20-39	0.319	0.279	0.395
40-59	0.353	0.345	0.369
60 or older	0.327	0.375	0.236
Education			
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			
Less than $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 or more	0.068	0.080	0.046
Occupation			
Full-time	0.497	0.509	0.475
Part-time	0.102	0.100	0.106
Not working	0.401	0.319	0.418
# of obs. in the sample	764	501	263

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

Figure 16. Monotonicity of allocations with respect to event structure



(A) Empirical CDFs of token allocation difference. The dotted line represents a 5-token margin. No two pairs of distributions are significantly different (two-sample Kolmogorov-Smirnov test). (B) Token allocations in two questions. The dot-dashed lines at 46.67 indicate the number of tokens which equalizes payouts in two events.

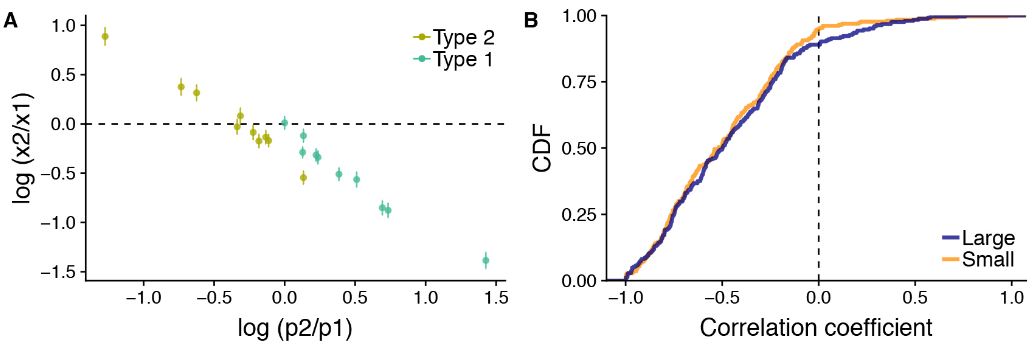
Table 5: Pass rates (%)

		GARP		SEU			MEU			PS*	
Treatment	*N*	Type 1	Type 2	Type 1	Type 2	Joint	Type 1	Type 2	Joint	Type 1	Type 2
Large variance	245	66.5	59.2	6.53	4.90	1.22	6.53	4.90	1.22	67.8	80.4
Small variance	256	65.2	60.2	4.30	4.30	1.56	4.30	4.30	1.95	66.8	78.1
Combined	501	65.9	59.7	5.39	4.59	1.40	5.39	4.59	1.60	67.3	7.92

Note: Since Epstein's (2000) condition is only necessary for probabilistic sophistication (PS), the numbers reported here capture the upper bound of the fraction of the subjects who are consistent with probabilistic sophistication. Type 1 and Type 2 refer to the two types of question subjects faced in the experiment.

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 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 with GARP pushes the average CCEI score above 0.982 (Table 6). The average e_* of 0.907 is not statistically different from the average 0.8782 in the lab study (two-sample t-test, $t(626)=0.7719$, $p=0.4405$). As in our lab study, we find that e_* for the market-stock task and how well choices respond to prices are positively associated (Figure 18). Subjects who violated monotonicity of choices (between 5th and 6th question) for more than a five-token margin have significantly higher e_* on average (mean 0.9991 vs. 0.8814, two-sample t-test, $t(499)=2.9253$, $p<0.01$), but the difference is not significant when we do not allow for this margin (mean 0.9283 vs. 0.8942, two-sample t-test, $t(499)=0.9877$, $p=0.3238$). Among the subjects who satisfied (exact) monotonicity, the larger the difference between tokens allocated in two questions becomes, the higher e_* becomes (Pearson's correlation coefficient $r=0.1248$, $p=0.0273$). 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.

Figure 17. Downward sloping demand



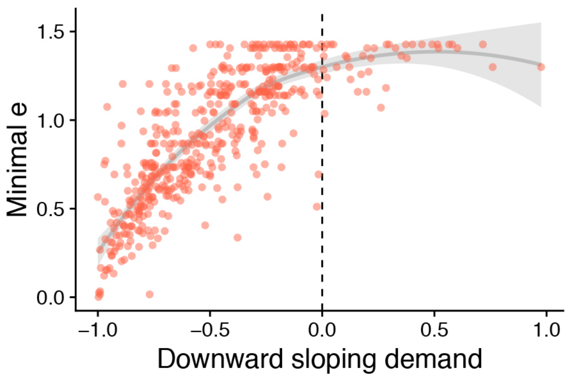
(A) Relationship between prices and quantities at the aggregate level. Bars represent standard errors of means.
(B) Empirical CDFs of Pearson's correlation coefficient between $\log(p_2/p_1)$ and $\log(x_2/x_1)$.

Table 6. Distance measures

	CCEI		e_*		
Stat.	**Type 1**	**Type 2**	**Type 1**	**Type 2**	**Joint**
Mean	0.9823	0.9824	0.6793	0.6920	0.9070
Median	1.0000	1.0000	0.6931	0.6376	0.9255
Std. dev.	0.0439	0.0404	0.4143	0.3844	0.3745

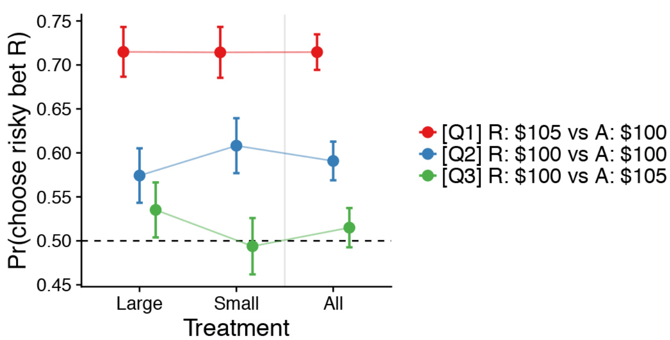
The pattern of choices in the standard-Ellsberg task is also similar to what we observed in the lab data, but the overall frequency of choosing the risky bet is reduced. In particular, only 70% of the subjects (regardless of the treatment) chose the risky bet in the first question, in which the risky 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 in the panel). There are thus 44% (26%) of subjects who are weakly (strictly) ambiguity averse (Figure 19). 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 15 which looks at the lab data, Figure 20 exhibits a decreasing relation between the two (there is a slight indication of reflection around $e_* = 0.8$, but it is not as strong as Figure 15). Combining these two observations, we can see that subjects with small e_* (close to SEU) does not necessarily mean that they are less ambiguity averse.

Figure 18. Relation between the degree of conformity to downward sloping demand and e_* in the market-Stock task for the Panel



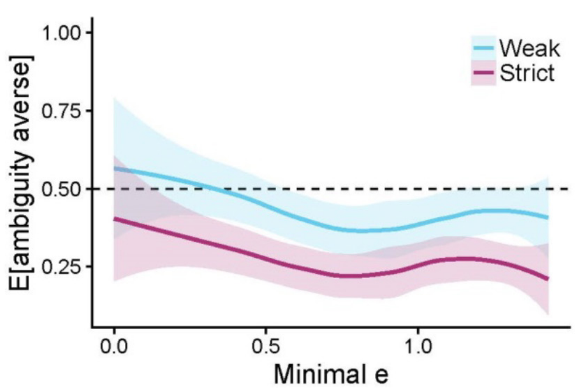
The shaded region represents 95% confidence bands.

Figure 19. Probability of choosing a risky bet in each question in the standard-Ellsberg task in the Panel data



Bars represent standard errors of means.

Figure 20. LOESS curves relating e_* and ambiguity attitude in the Panel data

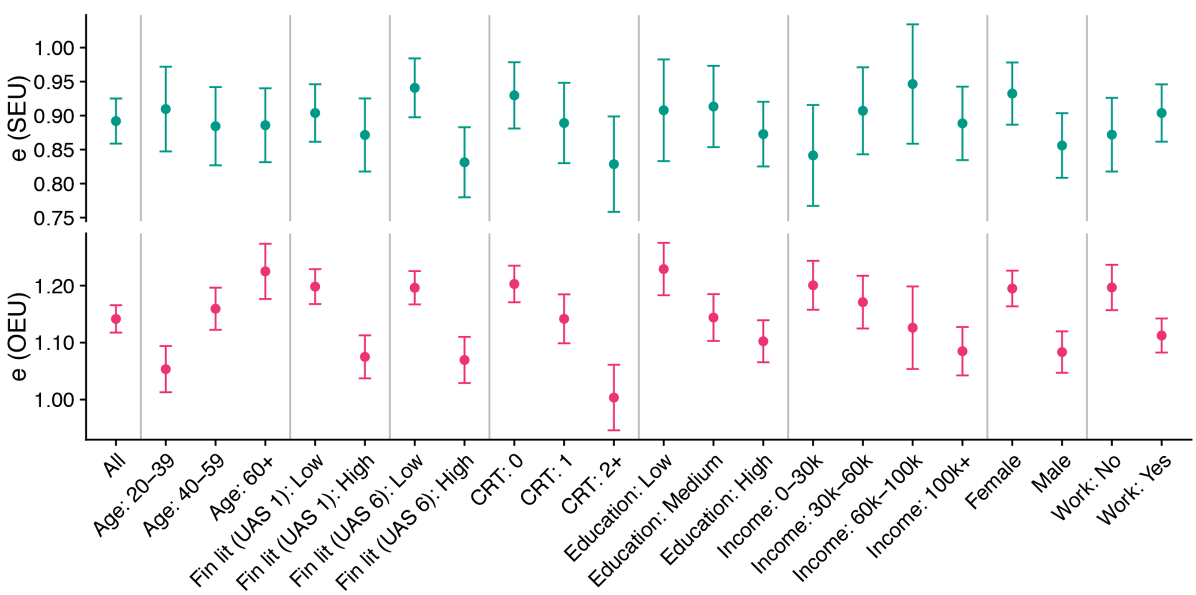


The shaded region represents 95% confidence bands.

Sociodemographic correlation. One of the great advantages of using the UAS survey is that registered researchers can access datasets from past surveys (unless they are under embargo), and use subject responses (through unique participant identifiers) 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.[18] Echenique et al. (2018) re-analyze this data and calculate e_* for objective expected utility (OEU).

Figure 21, top panel, shows average e_* for each category of sociodemographic characteristics. The first notable finding is that age and are not significantly correlated (comparison of age groups "20-39" and "60 or older"; $t(326)=0.5674$, $p=0.5708$), which is in a stark contrast with previous findings as we discuss below. Second, we find that subjects with higher financial literacy score (measured in UAS module #6) have significantly smaller e_* ($t(491)=-3.2277$, $p=0.0013$), meaning that these subjects are closer to SEU. Although financial literacy scores measured in two UAS survey modules #1 and #6 are based on a similar set of questions and are highly correlated ($r=0.6533$, $p<0.001$), high/low financial literacy based on module #1 is not strongly (nor significantly) associated with e_* ($t(497)=-0.9417$, $p=0.3468$). Similarly, cognitive ability (as measured by the score on the Cognitive Reflection Test) has a significantly negative relation with the distance to SEU (comparison of "score 0" and "score 2 or higher"; $t(332)=-2.3895$, $p=0.01743$). Finally, we observe a gender effect. Male subjects made choices that are significantly closer to SEU compared to those made by female subjects ($t(499)=-2.2678$, $p=0.0238$). Education and employment status do not exhibit a marked effect on e_*.

Figure 21. Average e_* for each demographic category



Bars represent 95% confidence interval.

[18] The cognitive ability measure is taken from UAS survey module #1, which includes the 5-item version of Sinayev and Peters (2015). Our two financial literacy measures are taken from UAS modules #1 and #6, which asked both the basic and the sophisticated financial literacy questions in Lusardi and Mitchell (2009). One caveat of using these supplementary data, of course, is the time lag between previous surveys and ours. For example, the first survey module UAS #1 was administered in May 2014.

In order to understand the relationship between several measures of decision making in our experimental environment and sociodemographic characteristics, we estimate a linear model

$$y_i = X_i \beta + \varepsilon_i ,$$

where y_i is the measure of decision making quality, such as e_*, for subject i, and 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"), education level (omitted category is "high school graduate 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.

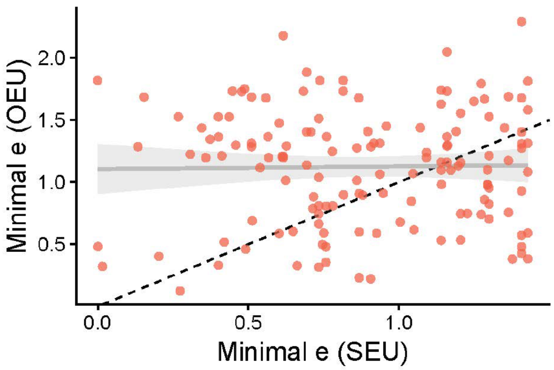
Regression results are presented in the first two columns of Table 7. First, it confirms our observation above, that there is no significant effect of age on e_*. The financial literacy variable measured in UAS module #6 (but not in UAS module #1) is significantly negatively correlated with e_* (i.e., subjects with higher financial literacy are closer to SEU), and its effect remains even after we control for education, income, and cognitive ability (column 2). Subjects in higher income brackets have significantly 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 the effect is only significant in the category "associate or professional degree," not in "college or post-graduate degree."

This set of results is in stark contrast with the findings reported for OEU in Echenique et al. (2018)—older subjects have significantly 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, 2017). The three OEU panels exhibit the same pattern. Since the survey of Carvalho and Silverman (2017) was administered on the same panel as ours, the UAS, we calculate average e_* using the set of demographic variables as above and also run the same set of regressions. In their data, we observe that e_* for OEU are significantly correlated with age, financial literacy, cognitive ability, and gender (Figure 21, bottom panel; columns 3 and 4 in Table 7). These results indicate that compliance with SEU and OEU may be unrelated.

3.3 Subjective and objective expected utility

We compared the distance to SEU based on our experiment, with the distance from OEU calculated using the dataset from an experiment reported in Carvalho and Silverman (2017). The main distinction between the two is that we give subjects no information regarding underlying probabilities, while in Carvalho and Silverman (2017) subjects (also on the same UAS panel) knew that two states were equally likely to happen. The overlap between our sample and Carvalho and Silverman's (2017) is small (143 subjects), but substantive enough that a comparison of for SEU and for OEU is feasible. We find no correlation between these two measures (Figure 22).

Figure 22. Relationship between e_* for SEU (this study) and e_* for OEU (from UAS #5; Carvalho and Silverman, 2017)



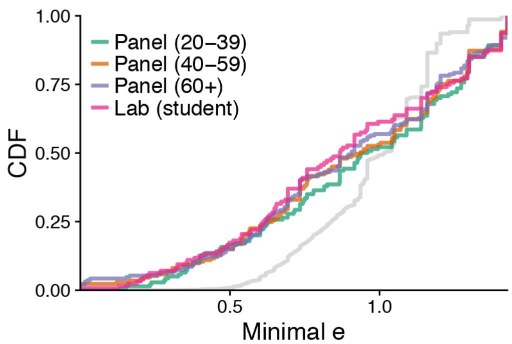
Correlation between these two measures are not significantly different from zero (p=0.6979). The grey line represents a linear fit along with 95% confidence band.

The obvious conclusion one would draw from our results is that SEU and OEU are fundamentally different phenomena, so that one agent may be close to OEU while being far from SEU. The counter-argument would be that there are additional distinctions between the SEU and OEU studies. The environment and computer interface are different. Past studies, including Carvalho and Silverman (2017), where our OEU data is from, use the paradigm introduced by Choi et al. (2007), where subjects are presented with a graphical illustration of a budget line and are asked to choose a point on the line. Our design, on the other hand, involves a financial market context, requires explicit allocation of tokens, given exchange rates between tokens and a monetary reward (which implicitly determine the budget line), somewhat similar to Convex Time Budget protocol of Andreoni and Sprenger (2012). One may thus argue that our experiment is more complex than previous studies, and that this added complexity is the main driver for many of our findings, including that OEU and SEU seem unrelated. But if complexity were the main driver of our results, we would expect an age-effect (or a more pronounced effect of cognitive ability). Or, we would expect differences between the lab results, where subjects were monitored to pay attention to the task, and the panel, where we had no control over how the subjects performed the task. Finally, we would expect that subjects who have trouble with the complexity of our study to perform poorly in both OEU and SEU terms, while subjects who can handle the complexity would do well on both. We do not see any of these effects.

3.4 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 23 that there is no significant difference in distributions. 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.

Figure 23. Comparing distributions of e_* from the panel study and the lab study



The grey line indicates the distribution generated by random choices.

Conclusion

Motivated by recent theoretical advances providing revealed-preference characterizations of expected utility theory, we design and implement a novel experimental test of the theory. We find that subjects 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 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 exists 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 significantly 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. In one of the two measures, financial literacy is significantly correlated with subjects’ distance to SEU. Together with the finding that the distances to OEU and SEU seem to be largely unrelated, our results suggest that behavior in the presence of uncertainty is fundamentally different from risk.

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.

Table 7: Relationship between demographic characteristics and e_*

	e_* (SEU)		e_* (OEU)	
	(1)	(2)	(3)	(4)
Treatment: Large	0.023	0.016		
	(0.034)	(0.034)		
Age: 40-59	-0.023	-0.012	0.129***	0.130***
	(0.044)	(0.045)	(0.027)	(0.028)
Age: 60+	0.023	0.026	0.215***	0.208***
	(0.048)	(0.048)	(0.034)	(0.035)
Fin. lit. (UAS #1): High	0.052	0.034	-0.056	-0.058
	(0.041)	(0.043)	(0.030)	(0.031)
Fin. lit. (UAS #6): High	-0.117**	-0.106**	-0.074*	-0.070*
	(0.042)	(0.041)	(0.030)	(0.031)
CRT score (UAS #1): 1	-0.021	-0.013	-0.038	-0.040
	(0.040)	(0.040)	(0.028)	(0.028)
CRT score (UAS #1): 2+	-0.052	-0.059	-0.122**	-0.122**
	(0.050)	(0.051)	(0.037)	(0.038)
Education: Some college		0.046		-0.040
		(0.053)		(0.035)
Education: Assoc. or professional degree		-0.107*		-0.062
		(0.054)		(0.038)
Education: College or postgraduate		-0.015		-0.021
		(0.050)		(0.037)
Income: 25,000-49,999		0.109		0.057
		(0.059)		(0.035)
Income: 50,000-74,999		0.184**		0.033
		(0.058)		(0.040)
Income: 75,000-149,999		0.155**		0.007
		(0.060)		(0.039)
Income: 150,000+		0.124		0.041
		(0.085)		(0.058)
Male	-0.052	-0.062	-0.082**	-0.084**
	(0.036)	(0.036)	(0.025)	(0.025)
Working	0.053	0.024	-0.006	-0.014
	(0.040)	(0.040)	(0.027)	(0.029)
Constant	0.923***	0.838***	1.173***	1.183***
	(0.051)	(0.070)	(0.031)	(0.038)
Observations	490	490	1,377	1,367
R	0.036	0.070	0.073	0.077
Adjusted R^2	0.018	0.039	0.068	0.066

Robust standard errors are presented in parentheses.*p<0.05; **p<0.01; ***p<0.001.

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.

Andreoni, J. and C. Sprenger (2012): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102, 3333–3356.

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

Baillon, A., H. Bleichordt, 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 (forthcoming): "Ambiguity Preferences and Portfolio Choices: Evidence from the Field," *Management Science*.

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 (2017): "Complexity and Sophistication," Unpublished manuscript.

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.

Cohn, A., J. Engelmann, E. Fehr, and M. A. Maréchal (2015): "Evidence for Countercyclical Risk Aversion: An Experiment with Financial Professionals," *American Economic Review*, 105, 860–85.

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.

Friedman, D., S. Habib, D. James, and S. Crockett (2018): “Varieties of Risk Elicitation,” Unpublished manuscript.

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004): “Differentiating Ambiguity and Ambiguity Attitude,” *Journal of Economic Theory*, 118, 133–173.

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 (2009): “How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy and Retirement Readiness,” NBER Working Paper No. 15350.

Machina, M. J. and D. Schmeidler (1992): “A More Robust Definition of Subjective Probability,” *Econometrica*, 60, 745–780.

Polisson, M., J. K.-H. ah, and L. Renou (2015): “Revealed Preference over Risk and Uncertainty,” Unpublished manuscript.

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.

Sinayev, A. and E. Peters (2015): “Cognitive Reflection vs. Calculation in Decision Making,” *Frontiers in Psychology*, 6, 532.

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.

Varian, H. R. (1990): “Goodness-of-Fit in Optimizing Models,” *Journal of Econometrics*, 46, 125-140.

Appendix

Table A.1. The set of 20 budgets

#	Type	Order	Lab		Panel	
			Account 1 (z_1)	Account 2 (z_2)	Account 1 (z_1)	Account 2 (z_2)
1	1	random	0.30	0.18	3.0	1.8
2	1	random	0.30	0.24	3.0	2.4
3	1	random	0.38	0.30	3.8	3.0
4	1	random	0.40	0.40	4.0	4.0
5	1	random	0.50	0.12	5.0	1.2
6	1	random	0.50	0.24	5.0	2.4
7	1	random	0.50	0.34	5.0	3.4
8	1	random	0.50	0.44	5.0	4.4
9	1	random	0.60	0.30	6.0	3.0
10	1	fixed (5th)	0.32	0.28	3.2	2.8
11	2	random	0.14	0.50	1.4	5.0
12	2	random	0.24	0.50	2.4	5.0
13	2	random	0.28	0.32	2.8	3.2
14	2	random	0.30	0.36	3.0	3.6
15	2	random	0.30	0.42	3.0	4.2
16	2	random	0.30	0.56	3.0	5.6
17	2	random	0.38	0.52	3.8	5.2
18	2	random	0.40	0.50	4.0	5.0
19	2	random	0.50	0.56	5.0	5.6
20	2	fixed (6th)	0.32	0.28	3.2	2.8

About the authors

Federico Echenique is the Allen and Lenabelle Davis Professor of Economics at the Division of the Humanities and Social Sciences at Caltech. He graduated from the Universidad de la República in Uruguay, and obtained a Ph.D. in Economics from the University of California at Berkeley. Echenique works on game theory, two-sided markets, and revealed preference theory. He has published over sixty journal articles, and is active in the economics and computer science community. Echenique has served on the editorial boards of, among other journals, the American Economic Review, Econometrica, and the Journal of Economic Theory.

Taisuke Imai is an Assistant Professor of Economics at the Department of Economics, Seminar of Economic Theory, at the Ludwig Maximilian University of Munich. He graduated from the University of Tokyo and obtained a Ph.D. in Social Science from Caltech. Imai works on behavioral/experimental economics, microeconomic theory, and reproducibility of scientific research.

Kota Saito is Professor of Economics at the Division of the Humanities and Social Sciences at Caltech. He obtained a Ph.D. in Economics from Northwestern University. Saito works on decision theory and experimental economics.